Reviewing Exchange Traded Funds

Market dimensional impacts on profitability

Johan Burck, jbuep07@student.lnu.se, 0709963706
2015-05-20
I’d like to thank my mentor Magnus Willesson for his feedback and guidance and Poja Mofakhari’s initial contributions to this paper.
ABSTRACT

Course: 4FE11E / 4FE03E Examensarbete Civilekonomprogrammet

Title: Reviewing Exchange Traded Funds: Market dimensional impacts on profitability

Author: Johan Burck

Background: Ever since Sharpe, Treynor and Jensen advanced the methods of fund performance evaluation in the 60’s it has been a popular field of study in academia. As the intricacies of fund performance was untangled it became clear that paying for active management doesn't yield higher cost adjusted returns. An Index investment strategy is the most sensible approach and it's the associated cost which separate index vehicles. Exchange traded funds have risen as a competitor to the conventional index mutual fund but the research evaluating these is very scarce. The research conducted comparing the costs of the two vehicles do not take into account implicit transaction costs that in turn depend on specific market microstructure designs and could affect the cost relationship.

The problem: Do liquidity and market structural disparities between markets affect the cost relationship between exchange traded funds and index mutual funds, through the implicit transaction cost?

Objective of the research: The objective of this paper is to examine whether structural differences between markets affect implicit transaction costs to the extent that the cost relationship between index funds and exchange traded funds differ from earlier findings.

Method: The need to generalize the findings prompted a quantitative approach to the research. Comparative examination will be done on the microstructure and liquidity of two different markets. The transaction costs will then be measured with statistical means and incorporated in a cost comparison model.

Result and conclusion: There are architectural and liquidity differences between the two sample markets allowing for systematic differences in transaction cost, which were found but were not a significant contributor to the tracking error cost of the index mutual funds. The Swedish ETF do not get more profitable as the investment sum increases. A finding which contradicts earlier findings and is likely a consequence of the Swedish tax-laws for capital gains as well as the higher levels of management fees for ETFs. ETFs might still be a worthwhile investment since they possess unique qualitative benefits.

---

1 Less tangible costs associated with funds such as the bid-ask spread. These are not as noticeable as explicit fees and commissions.
# CONTENTS

Preface ................................................................................................................. 1

Abstract ............................................................................................................... 2

Problem background .......................................................................................... 6

**Mutual Funds** ................................................................................................. 8
  History of the Mutual Fund ............................................................................. 8
  The workings of a mutual fund ..................................................................... 8
  Benefits of mutual funds .............................................................................. 9
  Passive management .................................................................................... 10
  Active management .................................................................................... 11
  Active compared to passive management ............................................... 11
  Fees and expenses ..................................................................................... 12

**Exchange Traded Funds** ............................................................................. 13
  Origin and Growth ..................................................................................... 13
  Workings of an ETF ................................................................................... 13
  Comparative studies ................................................................................... 15

Problem discussion ......................................................................................... 17

Statement of the problem ................................................................................. 18

Objective of the research ................................................................................ 18

Methodology ..................................................................................................... 19
  Scientific Approach ..................................................................................... 19
  Research Design ......................................................................................... 20
  Contrasting markets .................................................................................... 20
  Choice of Theory ......................................................................................... 21

Theory ............................................................................................................... 22

**Mutual Fund Performance** .......................................................................... 22
  Early Performance Measure ....................................................................... 22
  Later Studies ............................................................................................... 26
  Costs ........................................................................................................... 27
  Tracking Error ............................................................................................. 28

**ETF Performance** ....................................................................................... 30
Results and analysis .................................................................................................................. 65

Differences in the NYSE and OMXS Trading Systems ......................................................... 65
  Dealer Presence .................................................................................................................. 65
  Automation ....................................................................................................................... 65
  Continuous Trading ......................................................................................................... 66
  Protocols .......................................................................................................................... 66
  Post and Pre-trade information ....................................................................................... 67
  Expected Effects from Market Microstructure Differences ............................................. 68

Disparity in Liquidity: The Roll's Estimate .......................................................................... 69

Quantifying the Spread ....................................................................................................... 69

The Cost Comparison ........................................................................................................ 73
  Single Period Comparison .............................................................................................. 73
  Multi-Period Investment ................................................................................................ 81

Conclusions .......................................................................................................................... 83

References ............................................................................................................................. 85
PROBLEM BACKGROUND

In financial theory, investors maximize discounted expected returns. Investors desire these expected returns and is often presumed to oppose unwarranted risk-taking. Markowitz (1952) showed us that risk can be viewed as the variability of expected returns and ought to be minimized for any given level of return. This can be done through diversification. The age old idiom “putting all eggs in one basket” describes financial diversification in a nutshell. If not only being a pronounced rule within finance it has certainly been commonsensical amongst most people since time immemorial.

---

\[
\text{Diversification is both observed and sensible; a rule of behavior which does not imply the superiority of diversification must be rejected both as a hypothesis and as a maxim. (Harry Markowitz 1952)}
\]

---

A mutual fund is an investment company that enables investors to diversify their portfolio risk and get professional management by simply purchasing fund shares, without the having to bear the costs of creating an individual portfolio of assets. Rouwenhorst (2004) shows how, for small investors, mutual funds have become the primary investment over the past two decades. Sweden is no exception. Mutual funds have overtaken the traditional savings account as the most prevalent form of investment as a staggering 85% of swedes between the ages of 17 to 85 own shares. A large contributor to this number is the reform to the Swedish pension system (PPM) which places a part of citizens’ pension savings into mutual funds. While no less than 73% of swedes save through the PPM system – 71% still invest in funds separate to their PPM savings. (http://www.aktiespararna.se) The idea behind mutual funds is to accept cash deposits from investors and then issuing shares of their fund. The value of the shares is based on the fund’s Net Asset Value (NAV). They were first created in the Netherlands during the eighteenth century in the wake of a financial crisis as an affordable way for smaller investors to diversify risk but have also been recognized to offer other benefits such as being highly liquid and introducing professional asset management for investors, promising above average market returns.

For a long time mutual funds where actively managed to generate maximum returns at the lowest possible level of risk. By investing in accordance to an expressed strategy or policy the active fund manager will try time the market or select superior securities so that the fund yields surplus returns compared to the general market. The alternative is the passively managed fund or index fund that instead of trying to beat the market aims to replicate it through a benchmark index. The latter form of investing is adherent to the notion of efficient markets where it is assumed that investors nor fund managers cannot systematically beat market performance due to prices already reflecting all relevant information.

Evaluating fund performance has captured a whole lot of interest from the academia and took a quantum leap in the late 60’s when Treynor, Sharpe and Jensen advanced the field by introducing risk adjusted measures. These three gentlemen along with the majority of later studies conclude that the risk adjusted performance of mutual funds in general is poorer than what investors could achieve by a buy-and-hold strategy. There has been little evidence suggesting that active fund managers neither have superior security selection capabilities nor being able to out-time the
general market. Additionally Index funds tend to have lower turnover than the actively managed ones which lowers their expense costs. Frino & Gallagher (2001) argues that the allocation of capital to active funds from investors appears to make little economic sense in review of the literature on the subject.

---

*When we buy an actively managed fund, we are like gamblers in Vegas. We know it is likely to be a losing proposition, yet somehow we feel we are getting our money’s worth.* (The Wall Street Journal Feb 27 2001)

---

Index fund performance has gotten little review from academia despite their superiority over actively managed funds and rapid growth. Performance comparisons evaluate their index tracking ability and are essentially their ability to manage the composition of a portfolio of securities to replicate a given index. Since indexes are measured as ‘paper’ portfolios they are not exposed to any of the market frictions that the index tracking vehicles have to overcome, some tracking error – discrepancy in the performance of the fund portfolio and the underlying benchmark – is unavoidable according to Beasley et al. (2003). What has also become clear by Kostovetsky (2003) in comparing the performance of mutual funds is that turnover costs, expense ratios and transaction costs more often than not is key components separating funds from each other. If it is not economically defendable to pay the higher fees of the actively managed funds then these small activity cost differences will be the determining factor that compares index funds.
Mutual Funds

History of the Mutual Fund

Rouwenhorst (2004) states that a couple of core advances in finance during the eighteenth century allowed for the emergence of mutual funds. Securitization takes the cashflow of illiquid claims – such as plantation loans – as collateral for securities that can be traded on financial markets. Stock substitution is the process in which existing securities are repackaged to make them easier to trade – either in smaller denominations or at a lower cost than the underlying bond, stock etc.

After the credit crisis of 1772-1773 Abraham van Ketwich – a Dutch merchant and broker – formed the Eendragt Maak Magt-trust with the goal to provide the means to diversify for small time investors that otherwise couldn’t afford to do so. Risk minimization through diversification was done by having the trust invest in a wide array of different securities with uncorrelated cash flow. By doing this the investor need not to worry about any single given government’s or bank’s risk of financial distress since it would only constitute a small part of the entire trust’s portfolio. The investors were promised 4% annual dividends granted by paying subscription to the trust which were used by the trust manager to buy bonds and invest in plantation loans in the West Indies. However the Eendragt Maak Magt-trust was not without its flaws as outstanding shares were heterogeneous in the respect of cash flow rights due to the redemption-process (the return of the investors investment) leaving some shares better off than others.

It would take close to a century before investment trusts started to appear outside of the Netherlands. In 1868 the London based Foreign and Colonial Government Trust aimed to – much like its Dutch predecessor – invest in a varied collection of government bonds to provide investors with limited means the opportunity to limit the risk level of their savings through diversification. By 1875 eighteen trusts had been created in London.

In 1924 the Massachusetts Investors Trust became the first U.S. mutual fund with the ability to continuously issue (sell) and redeem (buy) shares at a price that was comparative to the value of the underlying portfolio of securities. This was in other words a fund with open-end capitalization compared to the customary model of former trusts or funds that had closed-end capitalization where only a fixed number of shares were issued at their conception.

The Workings of a Mutual Fund

Elton et al. (2007) explains how a mutual fund is an investment company that enables investors to invest indirectly in securities and currencies by purchasing its shares. By pooling money from investors the fund or trustee invests in stocks, bonds, money-market instruments and other securities according to a stated policy and objective. There’s two kinds of mutual funds: open-end and close-end funds.

When an investor purchases (or sells) shares in an open-end fund it is done from (to) the fund itself or through a broker for the fund and not on the secondary market. Additional shares are created when bought and they are taken out of circulation when sold (redeemed). The fund may in fact have to sell off part of its investment to fulfill large redemptions.

The price of open end funds are not listed like common stock since they don't trade on the open market. Instead the fund is reprised at the end of each trading day in accordance to the net value of its holdings or net asset value (NAV). In addition to paying the NAV per share the investor often pay
fees, such as sales loads. This open-end form of capitalizing funds is the most common model for mutual fund organization today.

Close-end mutual funds are instead created by an Initial Public Offering (IPO) to raise money to be invested and are then traded on the open market. Only a set number of initial shares are released for purchase – making the price of the shares subject to the forces of supply and demand where the supply is inelastic. So even if the fundamental value of a close-end fund share is equal to the NAV – it can still be traded at a discount or premium. In essence it's like owning a share in any corporation which assets are stocks and bonds.

**Benefits of Mutual Funds**

**Economies of scale**

One of the primary reasons that the trust gained in popularity throughout the eighteenth and nineteenth century is the economies of scale that comes with buying stocks and bonds in big bulks. Buying and selling securities comes with associated transaction fees in the forms of commission charges whether you're a private investor or a manager for a large fund. Even though some of the expenses increase with investment size, not all of them do and buying a few stocks at a time will often turn out to be a costly endeavor. By buying fund shares the small time investor could lessen some of the costs of transacting since these costs are borne by the fund which were buying stocks in big bulks. (http://www.investopedia.com (A))

**Diversification**

The financial crisis that hit Western Europe 1772-1773 would hit capitalists with large portions of their invested wealth in the region very hard. The early Eendragt Maak Magt-trust arose as a means for people to spread their investments to cash flow sources that was detached to such a scenario in the future.

<table>
<thead>
<tr>
<th>Market Condition</th>
<th>Asset 1</th>
<th>Asset 2</th>
<th>Portfolio with 60% in Asset 1 and 40% in Asset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bull market</td>
<td>$1.16</td>
<td>$1.01</td>
<td>$1.10</td>
</tr>
<tr>
<td>Stagnant</td>
<td>$1.10</td>
<td>$1.10</td>
<td>$1.10</td>
</tr>
<tr>
<td>Bear market</td>
<td>$1.04</td>
<td>$1.19</td>
<td>$1.10</td>
</tr>
</tbody>
</table>

Diversification is a risk management tool. The risk of a combination of assets is very different from a simple average of the risk of the individual assets. Portfolio theory tells us that the variance in returns of a combination of two assets may even be smaller than the variance of either of those assets themselves. It all has to do with imperfect correlation between the fluctuations in price of different assets as the market condition shifts. Table 1, much like the one used by Eltonet al. (2007 pp49), shows this effect by combining two assets which prices react differently to changes in the market.

By investing in a mutual fund the investor reaps the benefit of immediate diversification and asset allocation without having to pay the money associated with creating an individual portfolio.
Divisibility
Smaller investors may find it troublesome to buy securities for smaller sums at a time since some of them are sold in round lots – a group of 100 shares or any number evenly divided by 100. Mutual funds can often be purchased in smaller values at a time making it more attractive to small investors and opens up the opportunity for periodic investments to reduce the risk of mistiming the market or dollar cost averaging.

Liquidity
While divisibility provides respite for investors to time their purchases according to financial rationale rather than restrictions on the market in the form of round lots – mutual funds are generally easy to sell in a quick succession without affecting its price. While not necessarily true for close end funds which are traded like common stock on the open market – open end funds are redeemed at NAV at the end of the day by the fund.

Professional management
The investing strategy of a mutual fund follows a publicized policy according to Fama (1972). In the case of an actively managed mutual fund the fund manager will try to generate the highest possible return at a certain level of risk for the investors. This will be done by utilizing professional insight to purchase undervalued securities and/or make timely rearrangements of the portfolio to benefit from market swings.

Index tracking
While the strategy of investment in the eighteen century Netherland trusts where at the mercy of the fund managers’ ruling. They were actively managed funds where the portfolio composition is created in accordance to predictions of future market conditions. The alternative passively managed fund or index fund – aim not to beat the market and create excess returns for its investors. Instead these funds provide a vehicle for investors wanting to match the rate of return of their investment to a target index. (http://www.investopedia.com(B), (C))

PASSIVE MANAGEMENT
Passively managed funds have grown rapidly in popularity and totaled 865 billion dollars or roughly 9% of the total assets managed by investment companies. These funds are often components of a passive investment strategy which purpose is to benefit from long term price developments rather than short term fluctuations and entails a very limited amount of ongoing buying and selling. Proponents of tying ones investment to an index would argue that markets are efficient in the sense that all relevant information is already reflected in the prices of the assets. Therefore one cannot consistently beat the market after costs and taxes are taken into account.

Beasley et al. 2003 indicates that by investing in an index tracking fund the investor gets a well-diversified index portfolio at a low cost compared to going through the process of owning a the market portfolio – a mirror image if the index in weight – through individual securities. Funds that hold the market portfolio are said to have a full replication strategy. Other fund managers may not aim for a perfect copy of the market portfolio. Instead alternative portfolio constructs are calculated that closely mimics the index. Often the alternative method to holding the market portfolio entails a smaller amount of stocks in an attempt to lower the fund’s overall transaction costs according to Elton et al. (2007)
ACTIVE MANAGEMENT

An actively managed fund is a fund trying to beat the performance of the market/index by not holding each stock in proportion it represents in that index. Fama (1972) shows us that by not holding the market portfolio the fund is making a bet on a future forecasted market condition that its portfolio composition will benefit from. The forecast skill of active managers can be divided into two subcomponents. Timely recompositioning of the portfolio to benefit from predicted market movements on a whole (macroforecasting) or market timing. Forecasts of price movements in individual stocks (microforecasting) also known as security analysis. Here the manager tries to identify stocks that are over or undervalued and rearrange the portfolio to benefit from this.

Essentially active management forecasting can be split into two different methods: technical analysis and fundamental analysis. These methods are however not mutually exclusive and often used as complements to each other. The latter method seeks to gauge the value of stocks — or any other security — based on information about the corporation and the environment it operates in. Financial statements — past and present — from the firm being evaluated and competitors as well as related reports on market conditions are all used to make predictions of the future cashflow.

Bryman & Bell (2005) suggests that technical analysis on the other hand is not for determining whether a security is over or undervalued. Instead a security is said to be overbought or oversold. This is all based on the view that historical price movements are relevant for future price movements — that by analyzing historic price data of a security, trends and distinct price formations can be identified and used to make predictions of the price in the future.

ACTIVE COMPARED TO PASSIVE MANAGEMENT

Active and passive management both have strong and weak points that has to be taken into account. As described above the actively managed fund will inevitably be exposed to the unsystematic risk — company or industry specific risk that could be diversified — by not holding the market portfolio — which is only exposed to the systematic risk or market risk. The flexible purchasing and selling policies of the active fund often incur higher transaction costs as the managers try to find winners and time the market. Also, according to Beasley et al. (2003), additional costs arise by providing for the manager team. Even if the active management of funds is value adding it still has to create enough additional value to overcome the following accompanying costs of this activity:

- Higher salaries to forecasters or management fees charged by active managers compared to passive managers.
- The cost of having a higher level diversifiable risk or unsystematic risk which requires compensation in the form of higher return to the investor.
- Active management requires higher turnover in contrast to the low turnover of the buy and hold strategy in an index fund. This creates higher transaction costs.
- Depending on the tax law environment the actively managed fund may induce the investor to pay capital gains taxes as the turnover of the active fund is relatively high.

Truly passively managed funds — those which constituent stocks and weight fully replicate the underlying benchmark, full replication — will however experience certain difficulties that actively managed funds does not necessarily have to deal with:
A true replication of an index might impose holdings of comparatively small quantities of stocks which can be difficult from a managerial point of view but also in the sense of high cost of acquiring them through possibly illiquid markets.

The market portfolio is not static. Underlying benchmarks such as stock indices often change in their make-up and weightings as a consequence of an ever changing market. The S&P 500, for instance, must continuously rebalance, replace which companies and how big of a footprint they should make as they grow shrink, merge etc. Redefinition of indices can happen several times each year.

(www.standardandpoors.com)

Buy and selling shares will indubitably incur costs upon the fund by the mere reality of market frictions – transaction costs. A full replication strategy of investing will define the composition of the portfolio but it will not help in setting a roof on the transaction costs.

**FEES AND EXPENSES**

While there definitely are benefits to investing in mutual funds – it does not come without paying the fees and expenses needed to cover for the associated costs of managing a fund such as: shareholder transaction cost, investment advisory fees and marketing and distribution expenses. (http://www.sec.gov (B) (C))

All funds have continuously occurring operating expenses which are typically paid for out of the fund assets. Some funds however charge the investor shareholder fees at purchase or redemption of fund shares. Now, even though the operating expenses are covered by the fund’s assets – the investor still bears the whole cost and is charged in the form of a reduced net asset value of his or her shares.

**Shareholder fees**

One of the plethora of added expenses when buying fund shares is the Sales charge (Load) on purchase or Front-end Load which is a way to compensate the fund’s broker. It's basically a predetermined fraction on the invested amount that is going to the broker firm instead of being invested by the fund. Conversely there is sometimes a Deferred Sales Charge (Load) or back-end load on redemption. A fee paid to the broker when selling shares and which size is often dependent on the amount of time the investor has hold on to the shares. In addition to these there are purchase and redemption fees which are much like previous described albeit the proceeds goes to the fund itself and not to the broker.

**Fund operating expenses**

A very important statistic to observe when investing in a fund or when comparing different funds is the expense ratio. This number contains the total annual fund operating expenses as a percentage of the average net assets over the same period of time. The fund’s operating expenses are fees paid to the fund’s investment adviser for investment portfolio management, the costs of marketing and selling fund shares, shareholder services, legal and accounting expenses, etc.
Exchange Traded Funds

ETFs were originally introduced in 1993 and have become an ever increasingly popular alternative to the conventional index fund. ETFs are very similar to the traditional closed end fund and shares the index mutual fund’s idea of providing investors with an option to invest in a well-diversified portfolio that is set up to track an underlying index. But the two instruments differ in how they are structured and managed. Unlike mutual funds ETFs do not invest in securities by purchasing them on the open market. Instead ETFs are created through an in kind process where a basket of securities mimicking an underlying index is exchanged for ETF shares.

Origin and Growth

Barring instruments introduced by major US brokerage firms in the 70s and 80s with comparable features the world’s first exchange traded fund (ETF) is considered to be the Standard & Poor’s 500 Depositary Receipt – SPDR; ticker SPY and popularly known as “Spider” – which began being traded on the American Stock Exchange (AMEX) in 1993. The SPDR is designed to track the S&P500 stock market index and were accompanied by several new ETFs – such Diamonds; ticker DIA, tracking the Dow Jones index – the following years. But it wasn't until the introduction of the popular Nasdaq-100 Index Tracking Stock – Cubes; with ticker QQQ, recently changed to QQQQ – in March 1999 that the market for ETFs underwent its effective boom.

The success of the Cubes increased the awareness for other ETFs and the total assets managed by these funds more than doubled in 2000. The increasing interest in ETFs only continued with an 27% increase in 2001, 23% in 2002, 48% in 2003, 50% in 2004 and lasting throughout 2013 with an 21% increase in total assets from feb 2013 to feb 2014. As of 2014 there are over 1.300 different ETFs in the US with about $1,7 trillion under management. While total net assets of mutual funds in the US in January 2014 dwarfs this number with its 14,8 trillion dollars one might want to consider that by the end of 2002 there were only 113 ETFs with a mere $102 billion in invested assets according to Deville (2008). While iShares has become very popular by extremely diversified offer among sectors/countries, Cubes (QQQQ) and Spiders (SPY) still remain in the top five of the largest and most traded ETFs with Cubes being the most actively traded listed equity in 2005.

Workings of an ETF

Exchange traded funds are a combination of open-end funds and close end-funds – a hybrid instrument – taking advantage of the creation and redemption process of the open-end fund while still offer the possibility for continuous stock market tradability like a close-end fund. Open end funds redeem shares by buying back units for cash but only at the end of the trading day when the fund’s Net Asset Value has been calculated. To be able to meet these redemption claims requires the fund to keep a portion of its assets in cash holdings. Clearly these cash holdings are not being invested in securities and are as such causing a cash drag effect by not generating returns. Since close-end funds do not need to meet the obligation of redemption claims they can avoid this cash drag effect and investors who want to part with their shares can trade it on the secondary market. However by the laws of basic supply and demand mechanics, as demand fluctuates the fund with its fixed supply will trade at a discount or premium with respect to its NAV.

ETFs to exhibits superior tax efficiency and low management fees compared to conventional mutual funds, enabled by the exclusive “in-kind” creation/redemption process. The basic idea behind ETFs
are much like and stems from being organized as commodity warehouse receipt – a futures contract guaranteeing the quantity and quality of a particular commodity in a storage facility – with the physicals delivered and stored, although only the receipts are being traded and holders of the receipt can take delivery. The stock market variety would have market makers and institutional investors deposit stock baskets underlying an index plus a cash amount to the fund trustee and get shares of the ETF in turn. These ETF shares can now be traded on the secondary market much like common stock or used to be redeemed for the stock basket comprising the index plus a cash amount at that point. For some ETFs creation is allowed in cash only but this comes with higher creation fees for the AP to make up the increased transaction costs borne by the trustee. The market for ETFs actually is formed by two separate ones; the primary market for the creation and redemption process between institutional investors and the fund trustee and the secondary market in which ETFs are traded like stock in smaller denominations than the primary market. At the end of each market day the mix of securities that closely approximates the holdings of your ETF, the Portfolio Composition File (PCF), is created. The PCF contains the information about the securities and share quantities that is required to create a creation unit and its purpose is to deliver this information to the institutional investor in order for a creation or redemption the next day.

However not everyone are entrusted to participate in the primary market. Only authorized participant (AP) is allowed to exchange a basket of securities for shares and only in blocks of specified minimal amounts called creation units – usually 50,000 shares. Some of the APs that can enter this creation/re redemption process with ETFs are leading investment banks and institutional investors like Handelsbanken, Carnegie and Deutsche Bank according to Gerber (2008). Investors on the secondary exchange market cannot redeem individual shares themselves. This can only be done by offering to the trust shares in redemption units – in blocks of creation units – which is used by the AP to deliver the ETF shares to the custodian, and in exchange, receive the basket of securities from that day's PCF with the value still based on the NAV. This process – where the creator/redeemer trades a basket of stocks constituting the underlying index plus a cash amount – is called the In-Kind Creation and Redemption process. The cash part of the deal ensures that the yield of the shares between creation and redemption is equal to the index, less fees. It is included in the Creation Unit share to reflect the NAV of the PCF.

Deville (2008) describes how market makers are – thanks to the in kind creation and redemption process – able to absorb liquidity shocks on the secondary market by either making new or redeeming old shares. This would not be possible with a close-end fund with a fixed amount of outstanding shares. Because there are two markets for ETFs there will naturally be a chance of two different prices to exist for the same asset. The first one is the NAV/PCF of the fund holdings that is calculated at the close of the trading day and the second one is the prevailing market price on the exchange, determined by the supply and demand for shares. These two prices are not identical or fixed in ratio due to some contractual mechanism inherent in the instrument and could in theory diverge from each other. In reality, however, if the ETF price and the stocks constituting the ETF basket were to depart enough from each other the Authorized Participants could earn arbitrage profits. This ensures that the differences are not too sizeable.

---

2 A broker or individual that accepts the risk of holding a certain quantity of shares of a particular security in order to facilitate trading in that security. Since market makers list and trade buy and sell orders with a gap they earn the difference between the two.
Another major benefit of the In-Kind Creation and Redemption process lies in the receipt and delivery of the fully index-replicating stock basket which allows the trustee to stay invested. They do not need to sell any stock to meet redemptions. Moreover these in-kind operations are not taxable in the United States, making ETFs instruments very tax efficient.

**COMPARATIVE STUDIES**

Agapova (2011) demonstrates that the differences between the two instruments allow for divergences in their performance and costs however few studies exists on ETFs due to limited data as a result of their short period of existence. The studies that have been conducted on ETF performance has concluded that there may be a small benchmark tracking difference compared to mutual funds mainly due to factors such as fund expenses and the age of the fund according to Rompotis (2011).

Examining ETFs and comparing them to index mutual funds leads us to a wide selection of different attributes to consider. One positive attribute with ETFs is that they can be purchased and sold during anytime of the day, whereas a share of a mutual fund only can be purchased or sold at the stock markets closing time. Dellva (2001) shows that this benefit can easily be turned against the investor due to transaction costs incurred by high frequency trades. Rosella and Pugliese (2006) signify that acquiring fund shares are usually free of charge unlike ETF shares that has to be bought from a brokerage firm with commission fees attached to the deal.

Poterba and Shoven (2002) study tells us that the attribute of ETFs being more accessible than index funds makes them more valuable to institutional investors, intraday traders and speculators. Add the ability of short selling and ETF shares being able to have stop-loss and the trading flexibility increases even more. Kostovetsky (2003) suggests that tax efficiency is another attribute that the redemption in kind process of ETFs reduces alongside their withholding of realized capital gains. Shares in ETFs can never be redeemed by the fund, except in large clusters and by or through an authorized participant, and then only for an in-kind basket of securities. Hence, unlike index funds which can choose other strategies than to fully replicate the market portfolio, ETFs must have a composition and weight that is comparable with the underlying index. ETF are bought and sold as shares on the secondary market, making demand and supply the determining factors for the price, rather than the calculated net asset value (NAV) which is the case for index funds according to Rosella and Pugliese (2006).

While some research has been conducted on ETF and index fund performance in the sense of their ability to track benchmark indices, surprisingly few studies the cost differences which has been shown to be key in comparing index tracking funds.

Apart from qualitative differences between ETFs and index mutual funds Dellva (2001) inspects their expense ratios and how they affect fund performance. He could conclude that ETFs that tracks tracking broadly diversified indexes have the lowest expense ratios and that there is price competition between the two investment vehicles that doesn't favour mutual funds. Dellva also conducted a cost comparison considering different sizes in commissions paid when buying and selling ETFs. Results indicated that the spread between bid and ask orders fluctuates depending on the size of the transaction as well as the market liquidity of the ETF.

Dellva combines the commission and expense ratios into a single cost comparison model taking into account different lengths in holding period and amounts of total cash investments. The widely
considered to be advantageous expense ratios of ETFs are consumed by their high cost of being traded and the repeated costs of commissions and bid ask spreads will kill profitability for dollar-cost-averaging investors with small means.

Kostovetsky (2003) advances Dellva's cost comparison work by quantifying the cost difference in both a single and multiple-period model. The study outlines the non-tracking error costs or the non-implicit transaction costs to be management fees, shareholder transaction costs like broker's fees and bid-ask spreads and lastly taxation costs. Furthermore Kostovetsky stresses the fallacious notion of comparing fund performance to benchmark indices due to the existence of real work market frictions making it very difficult to do accurately.
**Problem Discussion**

The research that has been conducted comparing the activity cost differences between ETFs and index mutual funds is very scarce. While Kostovetsky’s multi-period cost model is the pinnacle on the subject, to say it comprehensively discloses the cost differences wouldn’t be accurate. Firstly, tracking ability is extraneous. While ignoring index tracking error of the instruments is fair in an activity cost based comparison it does pose some considerations when mutual funds are being compared to ETFs and not to other mutual funds. The unique architectural differences in how an ETF is created compared to a traditional fund make it so that the implicit transaction costs-part of what would be the tracking error for an index fund, is to be considered as an activity related cost; not tied to the performance of the ETF.

Consequently, leaving out tracking error would ceteris paribus skew the results since the implicit transaction costs stands on different sides of the fence between activity based costs and tracking error depending on which instruments is to be considered.

The manner of which this potential problem were tackled in the literature were the very setting in which Kostovetsky chose to conduct his study. It is a well-documented phenomenon, by authors such as Goyenko et al. (2009) and Madhavan (1992), that high levels of market liquidity reduces the impact of transaction costs and by focusing on some of the most traded ETFs on the US market Kostovetsky could certainly put forth an argument to omit these from his model. The implicit transaction costs were found to be minimal and could therefore be excluded completely from the model. This exercise of maneuvering around the inclusion of transaction costs from the model seems appropriate since the aim of the research were to better understand the cost relationship between some of the world’s most traded instruments and not to present an exhaustive model of all factors influencing their performance.

However, the dependency on this high liquidity condition might hamper attempts of extrapolating the results to less liquid markets since they cannot be assumed to exhibit such high liquidity and consequently their transaction costs might be important to account for. Can a trader on the OMXS put faith into the latest conclusions of the cost relationship between these financial vehicles when the research that has been done rely on circumstances that are foreign to the Swedish market? If the prevailing conclusions of the competitiveness of ETFs get skewed in less liquid exchanges then traders cannot be expected to be able to conduct rational decision-making when trading in regional markets.

And there are strong reasons to believe why liquidity ought to be different in regional exchanges. Liquidity research basically studies the ease of transacting on markets and the costs arise from it and while the definitions of the liquidity concept does vary somewhat; a centerpiece to this field of inquiry is the contribution of the amount of traders to the increase in liquidity. Smaller markets tend to enjoy less traffic in trades compared to the big financial centers. In fact, the congregation of traders put downward pressure on transaction costs which in turn draws the attention of other traders seeking a low market friction alternative.

In addition to liquidity having a significant impact on transaction costs, it has become more and more evident that differences in the microstructure of markets plays a big role in determining not only the level of liquidity, but also has a direct effect on transaction costs. It is the specific regulations, protocols and information dissemination practices of markets which market
microstructure theory state have impact on factors that are known to determine the magnitude of the costs of transacting. For example; the NYSE has partially preserved the traditional floor traded system of exchange while the Swedish OMXS is fully automated. This lends itself, according to Tse and Erenburgs (2003) study of Electronic Communication Networks (ECN), to the expectation of the NYSE, ceteris paribus, exhibiting elevated transaction costs. Other such differences prevails between the two markets. Madhavan (2002) suggest that it is through the differences in market architecture that liquid and low volatile markets are brought about. The former Chairman of the United States Securities and Exchange Commission (SEC), Arthur Levitt (2000) emphasizes in a testimony about strengthening the U.S. market system in 2000 the importance of the market structure influence on investor confidence, transparency and liquidity.

The body of market micro structure-theory behooves us to consider systematic and enduring differences in both liquidity and transaction costs between markets.

To answer the question of the market microstructural impact on the cost relationship between ETFs and index funds the study needs to take a two pronged approach to the problem.

The question of how liquidity and market microstructural differences affect transaction costs ought to be answered if we are to be able to anticipate certain tendencies in the cost relationship between markets. For this to happen, an understanding of the very nature of the relevant transaction cost, the parts which constitute it and the mechanics by which it functions and how they relate to market variables has to be understood.

Secondly, even if significant and persistent differences in transaction cost between markets were to be found, it still has to be investigated whether or not it is something that has any real bearing on the cost condition between the two instruments. Do the prevailing differences in market friction cause a real discrepancy from our current understanding ETFs relative profitability? Is it negligible? If so, how far is it from being significant for financial calculation? What kind of market differences are relevant for investors who buys and sells ETFs on various exchanges? Can they at least make qualitative forecasts based on observed market differences?

STATEMENT OF THE PROBLEM
Do liquidity and market structural disparities between markets affect the cost relationship between exchange traded funds and index mutual funds, through the implicit transaction cost?

OBJECTIVE OF THE RESEARCH
The objective of this paper is to examine whether market dimensional differences between exchanges affect implicit transaction costs to the extent that the cost relationship between index funds and exchange traded funds differ from earlier findings.
METHODOLOGY

SCIENTIFIC APPROACH

In methodology, quantitative and qualitative methods of inquiry are often distinguished. The quantitative approach basically covers methods that rely on explaining relationships or phenomenon with the help of statistical and mathematical measurements. The study usually starts off with formulating a hypothesis where the problem of the study is expressed and which the rest of the study is built upon. After that it’s sensible to select theory sources and methods that are going to be employed for the research. After that, all the data that is going to be analyzed is collected. It is of utter importance that the data being collected is reliable and isn’t erroneous. Data is reliable if it can be repeatedly collected with minimal variation. Lastly the data has to be worked on and analyzed in accordance with the stated method in such a way that it can either support or reject the hypothesis. The significance of the study is also to be highlighted as highlighted by Lundahl & Skärvad (1999).

The qualitative approach uses qualitative data and tries to achieve the goals of the study with qualitative analysis. While hard to exactly define, qualitative data can be said to be data which isn’t measurable or quantifiable in a numerical way. An example of this would be the story from someone expressing an experience of something.

This study is using numerical data which is quantifiable and is going to be utilized in both a statistical and mathematical manner to test the hypothesizes which would place it within the realm of quantitative research. However as the two markets structural differences are compared the study will touch upon the qualitative side since some structural differences aren’t black and white but up to interpretation.

Methodology also distinguishes between inductive and deductive inquiry. Induction basically relies on constructing theory and laws from observations the real world while deduction flips that on its head and makes prediction of the real world based on theory and laws. Induction of observations must happen for there to be theory from which to make deductive studies. Pålsson (2001) shows that one danger with the deductive approach would be the assumption that because the laws and theories spawned from inductive research held true at one or several points, it always holds true.

This study relies heavily on previous studies from the fund performance literature, transaction cost/liquidity- and market microstructure-theory. The cost relationship between ETFs and index mutual funds has been studied before by Kostovetsky (2003), Dellva (2001), and Agapova (2011) amongst others. A huge body of theoretical work has been done on liquidity, market microstructure and the bid-ask spread. To the extent this study stands on the shoulders of these giants it is a deductive approach, however it will be inductive in the sense that nobody has done this exact study before. Nobody has examined how the conditions of different trade systems affect the cost relationship between ETFs and index mutual funds. And as long as the virtue of skepticism and source criticism permeates the study, the dangers of deduction ought to be mitigated.

Since the paper aims to examine a sample that is large enough to generalize it to the population it has to aim for a large number of observations and as low as possible number of statistical drop-outs. Replicability is another key point for a deductive study as emphasized by Jacobsen (2002) and will be ensured though a structured working conduct where measures and the execution of the study
are detailed. Another pitfall of the deductive approach is the tunnel visioning that can happen when the researcher has a preposition on what is relevant literature and forgoes relevant information. To prevent this, literature in neighboring fields is superficially reviewed so that a better image of the theoretical framework is formed.

**RESEARCH DESIGN**

*CONTRASTING MARKETS*

In order to contrast a market which is relatively absent from frictions, like the NYSE, to an example which does not necessarily exhibit this attribute, I've chosen the Swedish OMXS. OMXS is a much smaller exchange than the global financial nexus that constitutes the NYSE because it ought to house significant lower volumes of trade and therefore lower liquidity. Apart from some market structural differences, the other main factor for selecting the OMXS for juxtaposition is the mere fact that the study is being done in Sweden, at a Swedish university. A circumstance that assuming lends itself to the focus on the regional market out of mere accessibility and ease.

To enable the study to define an answer to the devised problem it is important to identify how structural differences in trading systems affect transaction costs and how this difference in transaction costs affect the cost relationship between ETFs and index funds. Previous research from Madhavan (2002) amongst others has compiled the literature and explained that market structure factors are of key importance in influencing liquidity through functions such as the distribution of information on the marketplace, although some of these correlations have in other studies been weak or even contradictory. The importance of market microstructure theory for explaining the origin of liquidity, the bid-ask spread and market impact has been emphasized in both academic research, such as in Amihud et al. (2005) paper and by prominent economists such as the former SEC chairman Arthur Levitt (2000).

However the use of a market microstructure approach in explaining the differences in transaction cost would make it difficult to quantify and prone to subjective estimations since these markets seldom have distinct characteristics. Market microstructure theory will give a comprehensive overlook of factors in each market with proven connection to the bid-ask spread. The comparison of market architecture will give an indication of whether there is a reason to expect a significant and persistent difference in liquidity and the bid ask spread between the markets.

I will employ the body of theoretical work on market microstructure theory that Madhavan (2002) has gathered and presented to determining the structural differences between the two markets. A comparative research design in accordance with Bryman & Bell (2005) work will be employed. This form of investigation is suitable since what are compared are two different contrasting cases. A comparative approach does not require quantitative data which is important since it will be both quantifiable and qualitative differences in the markets which will be compared. Data from both cases will be gathered in a cross sectional manner upon which similarities and differences will be pointed out. A comparative approach requires that some understanding of some social behavior comes as a result. Liquidity and market friction is in essence investment behavior of large groups and if they are affected by differences in market microstructure, this research design would fulfill this criterion.

The market activity comparison of the two markets will be a cross sectional approach since it examines two different cases at one point in time with quantifiable data since the very nature of a
liquidity measure is to enable quantification. The same stands for the cost comparison between ETFs and Index funds. To ensure high reliability and validity is it important that well established measurements are employed. To ensure Replicability it will be important to describe all measurements, models and data in detail.

While market structural changes does affect transaction costs and liquidity levels, liquidity is its own monster that can behave autonomously and have a very direct impact on transaction costs according to authors such as Mcinish & Wood (1992), Tinic (1972), Stoll (1989) and Demsetz (1968).

I can conclude that liquidity is an elusive notion that cannot be captured by a single measurement. This prompts the effort to find the most appropriate measurements that can be used – either a single measurement or a set of measurement if they can complement each other. The usefulness of using a liquidity measurement to complement the market microstructure comparison is the quantitative nature of the result and its ability to compare the two different stock markets with a similar scale. There will be less room for misinterpreting the results. A liquidity measure will however only explain a confined aspect of the spread. It will give us a gradable measurement of the trading activity and, indirectly, the volatility part of the spread but not the asymmetric information part.

**Choice of Theory**

To accurately quantify the effects and significance of the bid ask spread a regression model will be employed. It is expected that the spread cost will have a significant impact on the tracking error costs and will differ in magnitude as the liquidity costs are expected to be heterogeneous. Frino & Gallagher’s (2002) model will aid in this endeavor as the significance and magnitude of each variable expected to impact tracking ability is estimated.

Lastly to determine how the bid-ask spread affect the cost relationship between ETFs and index funds this study will build upon the cost comparison model developed by Kostovetsky (2003). Kostovetsky could disregard the spread component for the funds on the U.S. market because according to his study, their size and significance as a cost component were negligible. But since the market structure differences on the OMXS are expected to alter the magnitude of the spread it will be included in the cost comparison analysis. This will be a departure from Kostovetsky’s assumption of perfect tracking ability since the spread do affect index funds’ ability to mirror the returns of the benchmark index. One major source of tracking error in index funds is the bid ask spread that managers have to pay when they invest new funds or readjust the composition of the fund. However, ETFs is exposed to the spread directly through the investor. Thus the cost comparison formula employed by Kostovetsky is to be tailored for the spreads unique impact on ETFs and index funds respectively.
THEORY

If I am to evaluate and test the cost relationship between two structurally different financial vehicles I have to build an understanding of how their performances can be measured. This paper is not breaking new grounds in this attempt but builds upon a rich history of earlier works on the sophistication of this craft. A brief review of the major groundwork done by earlier studies will be done in order to give proper context to and comprehension of the more recent research. As I wade through the fund performance literature it becomes clear how a couple of studies on the very fringe of this behemoth of academic inquiry will play key roles in shaping this paper.

In order to adequately examine whether there are reasons to suspect a systematic disparity in the implicit transaction cost between two markets, I first have understand its very nature. Unless an understanding of what constitutes, shapes and drives this implicit transaction cost exists, there will be little to no hope in trying to understand how it interacts with markets. The literature treat the bid-ask spread as the major and sole important contributor to the implicit transaction cost in relation to the market impact cost and will therefore receive special attention in this section.

Also, the literature on markets, their modi operandi and degree of activity, will be examined in an effort to provide further insight into the workings of the bid ask spread and how it interacts with markets. It will become evident that categorically separated fields of inquiry have big overlap in the explanation of the spread component but contribute separately to the study at hand. If the body of work constituting the collective knowledge of the spread helps me come to terms with its very nature and determining factors, then the literature on how different market characteristics relate to it will provide a useful tool for the solution of my problem.

MUTUAL FUND PERFORMANCE

EARLY PERFORMANCE MEASURE

In 1965 Jack Treynor highlighted the difficulties inherent in rating mutual fund management performance in an ever fluctuating market environment. By Treynors (1965) account, earlier, the standard practice would be to look at the average return over a period of time. However this method does not take into account:

- The disparity in performance incurred by market fluctuations due to differences in fund portfolio volatilities. Usually the effect of management on the rate of return is overwhelmed by fluctuations in the general market. Depending on whether it’s a bull or bear market some funds will outperform others due to higher or lower sensitivity – volatility – to market changes. Averaging return over a longer time period does not solve the problem since the market is always dominated by market trends.
- Average return comparisons do not factor in the investors aversion to risk. By simply averaging the return over some period of time as a tool for comparing fund performance masks the swings in NAV experienced during the time. Since the key function of these instruments is to provide affordable diversification it’s safe to assume that these fluctuations are of importance to the investor.
By plotting the percent rate of fund returns to the rate of return of the general market average, Treynor could determine that in this light, fund performance tends to be stationary over time. The slope of the *characteristics line* measures the volatility of the fund since the steeper the slope the bigger of a response in the fund performance is exhibited from a change in the market. And inversely a more moderate slope would suggest a fund with a less sensitive portfolio to market fluctuations.

Also since not all of the observations lies on the characteristics line, not all of the variations in fund performance are due to fluctuations in the general market. If the fund plotted would mimic the market to a T there would be no deviations from the line. If a fund shows great deviations from the line it suggests that the fund is not efficiently diversified and is experiencing fluctuations unique to the particular securities held by the fund – the fund is affected by risk that’s unrelated to the general market.

**FIGURE 1**

To battle the problem of risk aversion, a line is charted of any combination of a safe – less volatile –, low return portfolio and a risky high return portfolio where the investor is indifferent to any blend between the two as long as it’s on the line. To figure out which one portfolio between two choices that’s optimal they are initially marked on a chart. Since it’s assumed that the investor has the choice to seamlessly weight his personal investment between the mutual fund and a low risk money fixed claim – e.g. government bond – the *portfolio possibility line* can be drawn for each fund’s combination with the riskless asset. The best fund/risk free asset combination for the
investor will lie where the uppermost indifference curve touch a portfolio possibility line. From this graphical procedure Treynor creates the first risk adjusted measure for fund performance – the slope of the portfolio possibility line.

FIGURE 2

With the backdrop of modern portfolio theory, Sharpe (1966) would later explain that the performance of an efficient portfolio depends on the expected rate of return and the predicted variability or standard deviation of return.

Modern portfolio analysis theory concerns techniques for selecting and evaluating portfolios on the basis of predictions about change in the value of individual securities. From a risk and expected return centric viewpoint portfolio analysis aims to identify the preferred combination of these two elements. While the preferred combination strongly depends on specific investor risk preferences, portfolio analysis attempts to find efficient portfolios – the portfolio offering the greatest reward for a given degree of risk. Differences in mutual fund performance would from this standpoint depend: a conscious selection of a different degree of risk, a lack of understanding of how individual holdings affect the overall risk in the portfolio or inability to identify incorrectly priced securities.

However in the light of work done by Fama (1965) on how markets efficiently incorporate relevant information into the prices of assets, Sharpe detailed that since all investors are assumed to share the same predictions about the future performance of securities, all efficient portfolios will fall
along a straight line. Sharpe develops a measure of the ratio of the reward per unit of variability, the *Sharpe ratio*. An ode to Treynor is also done by substituting the standard deviation measure for variation in a second measure in the study done on 34 open end mutual funds during 1954-1963. The study showed that fund performance were somewhat correlated overtime and could therefore be predicted. The result could indicate that by superior portfolio management, some funds are able to consistently outperform others. But if the theory of how markets efficiently adjust to new information holds true the findings could be result of different expense ratios amongst funds. Further investigation in the matter led to the conclusion that good fund performances were associated with expense ratios, leaving the importance of management skill open for future study.

In his study, Sharpe compared a buy and hold strategy compared to investment in mutual funds. This was done by comparing the risk to variability ratio of the 30 stock constituting Dow-Jones index with that of 34 open end mutual funds during 1953-63. He found that the Dow Jones with its 16.3% return and 19.94% variability had a risk to variability ratio of 0.667 while it only was 0.633 for the sampled funds. Sharpe concluded that while the average fund manager selects a portfolio at least as good as the Dow-Jones index, the expense ratios made their performance vis-à-vis the investor, fall short from the index portfolio.

The development of the Capital Asset Pricing Model (CAPM) in the mid-1960s brought a tool for adjusting returns for risk. An important application of the CAPM, implemented by Jensen is the evaluation of the performance of managed portfolios.

Until Jensen's publication in 1968, portfolio evaluation was to rate a portfolio’s performance in relation to one and another, not to an absolute standard or benchmark. Jensen (1968) meant that to capture the predictive ability of portfolio managers one had to rework the formula used up till then since it only gave information about expected return on a portfolio given its level of riskiness / systematic risk.

If a security manager were to be able to predict future security prices better than the average market, he or she would be able to earn higher returns than suggested by the likes of Sharpe and Treynor. Jensen shows that the risk premium earned on any portfolio can be expressed as a linear function of its systematic risk, the realized returns on the market portfolio, the risk free rate and a random error term which has an expected value of zero.

In the case of fund a manager who consistently out selects the market the expected value of the random error term will not be equal to 0 since the fund earns more than its "normal" risk premium for its level of risk. By incorporating a non-zero constant – or not constraining the regression line to pass through the origin – and adding a new error term, Jensen could capture the fund manager's ability to forecast security prices in what would later be termed *Jensen's Alpha*.

Jensen's study on 115 mutual fund returns between 1945 and 1964 show that they on average were not able to predict the performance on individual securities well enough to outperform a buy-the-market-and-hold strategy. Funds on average earned 1.1% less than expected given their degree of systematic risk and could not generate enough returns to compensate for expenditure and fees.

---

3 $E_i = p + bσ_i$, where $p$ is the riskless interest rate and $b$ is the risk premium.

4 $R_{jt} - R_F = β_j[R_{Mt} - R_F] + e_{jt}$
**Later Studies**

Mutual fund performance research certainly didn't end after the major breakthroughs made in the mid-60s. Actively managed mutual funds have rapidly grown in popularity and as such a lot of studies examining fund managers’ selection and timing performance has followed. Chen & Stockum (1986) argued for a model re-specification from previous research since portfolio restructuring activities due fund managers trying to time the market would make ordinary least square estimations based on a fixed beta/systematic risk assumption produce biased estimates. Out of the 43 funds that were studied none had a statistically significant positive timing performance suggesting that none of the funds in the sample were able to time the market although some funds were able to select undervalued securities. A similar study performed by Lee & Rahman (1990) concluded that there are sign of superior forecasting prowess in a limited number of funds and that funds with no such skill should reevaluate their active management strategy for a passive, index tracking strategy.

The notion of efficient markets, spawned from the studies on price movements in the 60s and 70s, were relaxed in the 80s. New research showed that returns from securities were found to be positively correlated from period to period and various seasonal patterns were discovered – there appeared to be some degree of predictability in asset prices according to Fama (1991). Subsequently the literature on mutual fund performance, such as Ippolito (1989) and Goetzmann and Ibbotson (1994), displayed findings suggesting that manager performance produced enough added value to offset the their expenses and that past mutual fund returns were indication of future returns.

Malkiel published a paper in 1995 examining mutual fund returns during 1971 through 1991 with a unique dataset allowing for analysis of the survivorship bias – the tendency for poorly performing mutual funds to be eliminated by the trustee/fund company, leading to overestimation of past performance of mutual funds. Malkiel found that survivorship bias was considerably more significant than previous studies had implied, casting a shadow of doubt on discoveries of outperforming funds. All things considered Malkiel found that mutual funds tended to underperform the market, not only after management expenses were deducted, but also gross of all reported expenses except load fees. He concludes that the results do not provide any reason to abandon the efficient market hypothesis.

The findings of one of the more recent studies done by Frino & Gallagher (2001) reinforced the existing literature – active funds on average significantly underperform passive benchmarks. The index funds in the study earned higher risk adjusted returns after expenses than the large-cap active funds. The authors conclude that there doesn't seem to be any economic benefits from actively managed funds to the average investor.

In 1996 Gruber tries to solve the puzzle of why actively managed funds are growing in popularity when they are underperforming indices. Since buying and holding the market on one’s own is a costly adventure Gruber also examines the performance of index funds. Although a group of elite investors are able to find actively managed funds with good returns based on past performance, the findings show that mutual funds on average offered a negative risk adjusted return and that the average investor can get a better deal by buying index funds.
With simple arithmetic Professor and Nobel laureate William Sharpe (1991) asserted that:

- Before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar and
- After costs, the return on the average actively managed dollar will be less than the return on the passively managed dollar.

Starting off by classifying each investor in a given market as either:

- Passive – always holding every security from the market with each represented with the same weight as in the market
- Active – not being a passive investor in the sense that his or her portfolio will differ from that of the passive manager or market portfolio at some or all times.

Since the return of the general market over any time period will be the weighted average of the returns on the securities within the market, each passive manager will generate the market return, before costs. And because the market return must equal a weighted average of the returns generated by all active and passive investors, active investors must also earn the market return.

To prove the second assertion, Sharpe simply points out that the cost of active management will exceed the cost of passive management due to a higher frequency of trading and the greater amount of personnel employed. And since active and passive returns are the same before costs, it must follow that active management yields a lower return after costs than passive management.

Prior research in the field of mutual funds has indicated that investments in mutual funds have not yielded a higher risk adjusted return than a buy and hold strategy. This is not to say that the actively managed fund has to have a portfolio beta greater than zero – higher than the market risk level – but a positive alpha – management specific returns – is required to offset the increased fixed costs of increased management-requirements according to Beasley et al. (2003). There is only a select few fund managers who can beat the market and without previous knowledge of which these managers are it’s better to invest in index tracking vehicles. However Frino & Gallagher (2001) points out that empirical research evaluating these is remarkably rare. This is especially surprising considered the rampant growth of the number and amounts being invested in indexed assets.

**Costs**

As one sifts through the literature a key component materializes when comparing active funds and index funds: the cost of activity. Turnover costs, expense ratios and transaction costs all contribute to a 1% to 2% difference between funds which often if enough to make funds underperform compared to the market. The cited research indicates that active fund managers on average cannot beat the market but instead bring extra costs upon the investors. The specific relationship between costs and fund performance has served as a field of inquiry for a few researchers.

For example, the highly controversial law 12b-1, which authorizes funds to deduct a sum from net assets with the money to be paid to selling agents to cover sales and advertising, has been shown by Ferris & Chance (1987) to be a deadweight loss to the investor. It has been argued that the increase in total NAV from the enhanced incentives for brokers to sell the fund would have the economies of scale effect kick in, increasing the overall value creation from funds. Funds do however not show the decrease in expense ratios one would expect which makes the 12b-1 ruling nothing more than a
way for funds to preach “no-load” but practice something entirely different. Furthermore, Grinblatt & Titman (1994) argues that, there seem to be no correlation between a mutual fund performance and its expenses, management fees or loads that are being charged suggesting that there is no additional return premium for buying funds with higher costs.

Dellva & Olson (1998) shoulder the task of examining the relationship between front-end load charges, deferred sales charges, redemption feed and 12b-1 fees on total fund expenses and risk adjusted performance in their 1998 study. The result made it clear that funds that charged front-end loads had a lower risk adjusted return than those which didn’t. And since there were hints of a positive correlation between front-end loads and fund expenses the authors recommend investors to stay away from these fees. Additional increases in expense ratios and deadweight losses for the investor were the consequences of the other fees that were investigated.

**Tracking Error**

While the theory behind tracking an index is simple, the practical implications of real markets make index funds imperfect instruments for precise tracking. While being possible to be broken down into several sub components the biggest hurdle for an index tracking fund is that it is being compared to a market portfolio formed and maintained in a hypothetical market where securities are bought and sold at any time without cost – a paper portfolio. Beasley et al. (2003) emphasizes that asset indices are theoretical constructions of a basket of assets prices which can be weighted by for example market capitalization. In contrast to index funds – indices assumes a frictionless markets in the sense that re-weighting of the basket is calculated without considerations to any costs inherent to such an undertaking in the world of “imperfect markets”.

Frino & Gallagher (2001) asserts that while passive fund managers aim to fully replicate the returns of the benchmark portfolio/index, the mere fact of them operating in the real world with market frictions unavoidably causes error in their objective. A fund with a full replication strategy that exactly mimics the construct of the market portfolio cannot be guaranteed to have the same yield as the target. This is because the index is a mathematical calculation derived from a portfolio that’s not faced with the same market frictions that the index fund faces. This leaves index fund managers with the dual objective of firstly minimizing the above mentioned tracking error in the funds’ performance and secondly minimizing the costs that spring from tracking the index as closely as possible. The duality of passive fund management is that the more transacting the manager does to maintain the tracking of the target benchmark the greater the transaction costs will be, which in turn will expressed as an underperforming fund or in other words as tracking error. The index fund has to weigh the cons and benefits of a full replication strategy – which will cause tracking error – to a more active strategy with higher turnover – which leads to higher transaction costs.

The sources of the tracking error have been identified to the following:

- **Transaction costs** – The index portfolio is maintained in a hypothetical world were transactions are assumed to be instantaneous, costless and not restricted in quantity. The index fund experience transaction costs, such as the bid-ask spread – the discrepancy between buy and sell orders – as it has to rebalance the portfolio to match changes in the index, invest capital from new investors and at the inception of the portfolio.
• Fund cash flows – Edelen (1999) states that cash drag is a source of tracking error facing index funds. Due to the need to readily have cash available for redemption needs, some of the funds capital has to stay unvested. Another contributing factor to the cash drag phenomenon is new funds that stay latent before investment.

• Index recompositioning – Kostovetsky (2003) tells us how rebalancing costs arise due to alteration of the composition of the index as the index fund has to either sell or buy shares to mirror the portfolio structure but may not be able to make the complete the transaction in time. It has been showed that particularly strong effects on the tracking error of index funds stems from the quarterly dividends distribution, transaction costs and December tax loss selling.

• Corporate activity – some securities that are subject to corporate mergers or takeovers from companies outside the index may cause problems since the fund might receive the cash settlement at another date than it being removed from the index.

• Volatility – If the index fund hold and exactly weighted copy of the market portfolio, high volatility will bear no effect on its tracking ability. However when this is not the case, big market fluctuations will tend to bring out the difference in the fund portfolio and the market portfolio in the form of tracking error. The magnitude of tracking error will unsurprisingly be tied to how much asset prices change.

• Dividends – One major source of tracking error resides in the dividend policies of companies as index funds must wait for some time to receive the cash dividends for reinvestment.

• Liquidity – The liquidity of the securities in an index will have an impact on transaction costs for the funds with a full replication approach since they are expected to invest a proportion of their cash in less liquid assets. According to Frino & Gallagher (2001) illiquid stocks impose greater costs when fund managers have to trade them since it’s linked with greater transaction costs. Edelen (1999) brings up the issue of in-and-out trading or short term traders capitalizing on the price rigidity of index funds and how this also can be a meaningful influence on index fund tracking error. The in and out trading cost is an occurrence that arises as an indirect cost from providing market liquidity.

It has been shown that the tracking error of funds is seasonal in nature. December tax-loss selling and quarterly dividends distribution have strong impact on the way the tracking error manifests in funds. The above described factors all prevent index funds from perfectly duplicating the performance of an underlying index. However, index funds are not the only passive investment vehicle alternative to the active managed funds. Over the years Exchange Traded Funds have become an ever increasing alternative to traditional index mutual funds with their low fees and intraday trading capabilities.
ETF PERFORMANCE

The market for ETFs is growing rapidly with ETFs increasingly taking market shares from traditional index mutual funds. Agapova (2011) studies the substitutability for ETFs over index funds. She concludes that ETFs and index mutual funds are substitutes, but not perfect substitutes. Though, ETFs has added a level of competitiveness to the market and offers new features previously unavailable in mutual funds. This has led to lower prices for investors and added new service features which include (i) higher liquidity (ii) more tax efficiency (iii) lower management fees for long-term investors.

Deville (2008) argues that in the financial literature ETFs, with their exceptionally low management fees and expense ratios, are portrayed as a low-cost substitute to mutual funds. And as these instruments gained in popularity the mutual fund industry responded with lowering their fees drastically over a long period of time resulting in what some would call a “fee war”. Sequentially the expense ratios for some of the most popular ETFs were lowered as the expense ratio measure continues to be marketed as a competitive edge for ETFs over mutual funds. However, Dellva, Kostovetsky amongst others has shown that simply comparing expense ratios is too of a crude way of measuring the performance since it does not take into account the trading costs linked with ETF trading nor the possible disparity in tracking ability of the two instruments.

TRACKING RELATED DIFFERENCES

Due to the very nature of the creation / redemption process by which an ETF is created no liquidity costs arises. By providing a sufficient number of shares that matches an underlying index in weight and value, a market dealer can obtain ETF shares; ergo there’s no need for a purchase of the underlying assets on the market and thus no impact on the market price of these will be inflicted. The creation / redemption process also restraints the ETF price to stray too far from the NAV since an arbitrage opportunity would arise by the fact that investors would be able to create (redeem) new shares and sell (buy) them if the price of the ETF is too high (low). However, a cash drag effect occurs due to the frequent changes in price of the underlying index and the ETFs. The difference in value between the round number of ETF shares and the securities provided must be held as a cash component. The cash component of the ETF value will not be invested and thus contribute to the tracking error of the ETF. And just like mutual funds ETFs also experiences the effects of dividend policies and cost and timing mismatch as the index composition changes, on their tracking ability according to Kostovetsky (2003).

Apart from a comparison of cumulative daily returns Aber et al. (2009) checked the tracking ability of ETFs and index funds by doing a mean variance analysis of the return. This was done to capture the ability of the funds to replicate both the return and the risk of the benchmark index. This is done by defining the tracking error by the standard deviation of the difference between the tracking vehicle’s and benchmark’s returns over time. The conclusion from this analysis was a 2 to 3 basis points better performance of mutual funds compared to the ETFs on average. This was due to the ETF manager’s reluctance to make adjustment to the portfolio in advance to the official moment of the index adjustment.

Rompotis (2011) studied the performance of ETFs relative to the S&P 500 – a representation of the market. The results indicated that a majority of the ETFs in the study outperformed the market and the returns of the ETFs strongly persist at the short-term level. Three measurements were used – the simply raw return and two types of risk-adjusted return measurements – Sharpe ratio and
Sortino ratio. Regression analysis revealed that the persistence of the tracking error in ETFs is explained by expense, age and risk factors. The tracking error was estimated as the standard deviation in return differences between ETFs and benchmarks. Previous research showed that the variables having a significant impact on the tracking error were cash drag, the spread cost, the dividend payments arising from the stocks of an index as well as the size and timing of the rebalancing process according to Frino & Gallagher (2001). The estimated minimum and maximum tracking errors showed in the results indicates a significant difference in a ETFs ability to track its benchmark. It has been shown that there is predictability in ETFs performance indicated by its raw returns. The top performing ETFs in a year are probably going to perform in a similar way in the subsequent years and confirm that the performance of ETFs is somehow predictable. However Shin and Soydemir (2010) reveals that another study found that 24 out of 26 international ETFs had significantly smaller returns than their benchmarks based on Jensen’s alpha – which also is a risk-adjusted measurement. Further – the tracking error was significantly different from zero and show persistence. A change in the exchange rate is found to be a prominent source of tracking error. It should be noted that these ETFs are benchmarking foreign markets and are thus exposed to exchange rates unlike domestic ETFs. Also in a study published in 2002 Elton et al. managed to identify a 0,28 % underperformance in the SPDRs relative to the S&P 500 index which were contributable to the tracking error.

**Non-Tracking Error Differences**

Dellva (2001) assesses ETFs in the respect of their comparability with mutual funds, suitability for different types of investors, cost and tax effectiveness as well as the reason behind their recent success. Dellva recognizes the traditional viewed qualitative advantages of ETFs to be

- Their ability to be continuously evaluated compared to traditional open end funds which has its NAV estimated at the end of the trading day.
- Direct accessibility to the market via the stock exchange.
- The possibility for symmetric trading by taking short – selling a borrowed security with the expectation that the asset will fall in value – or long positions.
- The capacity to leverage one’s portfolio by margin trades where securities are used as collateral to borrow money for a purchase of securities. Basically a way of magnifying profits (losses) made beyond of one’s own capital.

Dellva also examines the widely praised expense ratios of ETFs compared to traditional mutual funds and how they affect fund performance. The result showed that ETFs tracking broadly diversified indexes had the lowest expense ratios spanning from 0.18 to 0.25%, that there is a clear case of price competition between the two types of investment vehicles and that it doesn’t favour mutual funds.

A cost comparison, that incorporates the commissions paid when buying and selling ETFs, is also conducted. This commission will vary in size depending on which broker is used and the amount and timing of the transaction. The study suggests that the spread between bid and ask orders varies depending on the size of the transaction as well as the market activity / liquidity of the ETF with widely held and actively traded ETFs exhibiting lower spreads and thus lower transaction costs. Dellva merge the trading costs and expense ratios into one single cost comparison between ETFs.
and mutual funds. The comparison takes into account different lengths in holding period and additionally various amounts of total cash investments. The outcome would point out that much of the advantageous expense ratios of ETFs are consumed by their high cost of being traded. With an initial investment sum of $10,000, ETFs would not surpass the returns of mutual funds until the sixth or ninth year depending on which broker are being used. Furthermore, small dollar-cost-averaging investors in ETFs will never be able to overcome the recurrent costs of commissions and bid ask spreads with the lower expense ratio. As previous research suggest, Delva also conclude that ETFs are very tax efficient. This is due to the In-Kind creation and redemption-process not constituting a taxable event while the turnover that occurs in mutual funds trigger capital gain taxes.

The study advocate that the ETF users should be higher net worth investors taking large positions to generate sector or technology exposure, or perhaps day traders that want to free themselves from the restrictions of traditional mutual funds. Tax-deferred, long-term retirement investors will get little to no advantages from ETFs. The brokerage commission costs will make them unappealing to small investors.

Kostovetsky (2003) goes one step further than Delva by quantifying the cost difference in both a single and multiple-period model. Much like Delva, Kostovetsky do not dig too deep into the issue of the relative tracking capabilities of ETFs and mutual funds. Instead, perfect tracking ability is assumed so that the study can hone in on the important choices facing the investor due to the non-tracking error differences between the investment vehicles. By focusing on very liquid ETFs like QQQs and SPDRs; Kostovetsky could omit not only tracking error but also the negligible spreads from his model. The complexity of quantifying the tracking error and including it into the model stand as the reason behind Kostovetsky’s assumption. The non-tracking error costs, or as expressed in the paper explicit transaction costs are:

- **management fees**: Management fees of mutual funds can range from 2 % for actively traded funds to typically management fees in the magnitude of 0.4 % for index funds (www.avanza.se). These costs have on the other hand been even lower for ETFs as some of the most popular ETFs traded on the NYSE has expense ratios as low as 0.09 %.

- **shareholder transaction costs**: Broker’s fee – Many index funds are purchased without a commission. In fact the majority of funds are no-load funds, charging no commission at transacting. In contrast when buying an ETF the investor must make the transaction on the secondary market which involves paying a commission to the brokerage house to compensate for the execution of the contract on behalf of the customer. **Bid-ask spreads** – The second part of shareholder transaction costs is the fee paid to the market makers through the system of the bid ask spread (Kostovetsky 2003). The bid ask spread is the largest of the transaction costs facing the investor and is generally greater the less liquid of a security is being dealt with. Together with the spread, the market-impact cost forms the implicit transaction cost. The market impact cost manifests itself in the form of price concessions for large trades as they move the bid (ask) price downward (upward) (Fleming 1996).
- **taxation costs:** When redemptions surpass the additions to a fund the managers must sell stock and hand out capital gains to shareholders. These capital gains are taxed and create a cost for the shareholder. ETFs differ in this regard as they rarely if ever distribute capital gains due to the ability to circumvent the tax by redeeming stock (which isn’t a taxable event) in the right order so that if the times would call for selling stocks, it can be done without incurring capital gains taxation (Kostovetsky 2003).

The results indicated that ETFs outperformed index funds if sufficient amounts of capital were to be invested and/or the holding period were long enough. By quantifying the costs and running a multiperiod model with continuous reinvestment in the ETF, Kostovetsky illustrate that the investor needs 59,650 USD or in the case of holdings of ten years; 13,019 USD invested in ETFs for them to be preferable to index funds.
**THE BID-ASK SPREAD**

The bid-ask spread is the price impatient traders pay for immediacy. Demsetz (1968) showed that the bid-ask spread can best be understood by considering the ignored this immediacy problem in supply and demand framework. Accurately predicting the needs of immediacy in human action would be an unusual trait, and approximating it incurs costs for the person specializing in standing ready and waiting to trade with the incoming orders of those who demand immediate servicing of their market orders. Demsetz describes the bid ask spread as the mark-up that is paid for predictable immediacy of exchange in organized markets, or inventory mark-up in other markets. Impatient traders buy at the ask price and sell at the bid price. It is the return that dealers/market makers and limit order traders’ facilitating of immediacy yield or the compensation for the cost of waiting. If all traders had the same time preference the equilibrium spread would ensure that traders would be indifferent between using limit or market orders. When the time preferences between traders differ the equilibrium spread sets the supply of immediacy/liquidity equal to the demand for liquidity. Harris (2003) argues that traders who set a high value to their time and trade on material information will tend to use market orders while traders valuing time lower and has a high risk tolerance prefer limit orders. Market order users basically pay the limit order users for the cost of waiting in the form of the bid-ask spread.

**THE SPREAD COMPONENTS**

Stoll (1989) states that the quoted bid-ask spread is made up of three parts:

- order processing costs / transaction cost
- inventory holding costs
- asymmetric information- or adverse selection cost

---

5 An order placed with a broker to buy or sell a specified number of shares at a set price or better. Compared to a market order which aims to sell or buy a specified number of shares at the best available current price.
Order processing costs Due to the fact that buyers and sellers do not necessarily have active orders at the same time the theoretical equilibrium price for a stock will not be realized. An important implication of the introduction of transaction costs or immediacy costs is the notion of two supply and demand curves, reflecting the separated time frames of the trade as seen in Figure 3. One equilibrium point E will lie where it is supposed to in the traditional walrasian framework, constituted by the average demand curve D and the average supply curve S. However the demand and supply lines D and S are those of market participants wanting immediate execution of their orders. Consider the equilibrium point E as the average price for which security X has been trading at. However, a person that wants to sell shares immediately at the equilibrium price cannot rely on the existence of a buyer at that precise moment and vice versa. Therefore the D and S-line do not demonstrate always existing market orders, rather they measure time rates of demand and supply for which no orders need to be present at all times. Imagine a person or other market participant that takes on the task to stand ready to sell or buy at stated prices, instantly upon receiving a matching order. To cover the cost of standing ready this person will be buying (selling) security X at a price below (above) the equilibrium price E, represented by the market makers demand D\text{mm} and supply S\text{mm} curve. The total price difference constitute the bid-ask spread. Market makers are individuals or institutions which specialize in providing the service of buying and selling share at certain stated price levels immediately. This difference is visually presented by the two new intersections in the figure giving birth to the two new equilibrium prices; A for immediate sales and B for immediate purchases. This “mark-up for immediacy”-component constitutes what other authors would call the order processing cost or transaction cost portion of the spread.
**Inventory holding costs** According to Madhavan and Smidt (1993) by providing the service of liquidity, by standing ready for immediate exchange, to the market – dealers bear unwanted inventories and the risk that follows since the value of the stock might deteriorate. Ho and Stoll (1983) showed that in the face of inventory carrying costs; risk adverse market makers will adjust the size of their inventory towards the desired levels by the process of regulating the price. This part of the spread is the inventory holding costs stated Stoll in a later paper (1989). The inventory holding component of the spread is narrowest when the market maker does not bear unwanted inventory and increases as the inventory deviates from the optimum level. Disequilibrium of market makers inventories manifests itself not only in the price spread but also in the rise in market-impact costs (i.e. large movements in the price from trading) at the end of the trading day as the market makers need reimbursement for carrying overnight risk. Furthermore Madhavan (2002) shows that the amount of diversion from the optimal inventory level is correlated to capital constraints which might lead to greater inventory effects for smaller dealers therefore possibly resulting in a larger bid-ask spread given low enough competition.

**Asymmetric information cost** Later views of the spread relied more on informational arguments such as Glosten & Milgrom’s study from 1985. The asymmetric information component of the bid ask spread covers the recovery from the losses incurred by trading with informed traders. These informed traders make investment decisions based on private information only available to them or publicly available information. They compare their value estimates with the market prices and buys if they find undervalued securities and sell overvalued securities. Harris (2003) emphasises that informed traders are the only traders who make prices move towards their fundamental value, all other traders add noise to the prices and are therefore called noise traders. Glosten & Milgrom showed that the presence of traders with superior or private knowledge in a market leads to a positive spread since dealers need to compensate for transacting with better informed traders. This is mainly an asymmetric information problem where badly informed investors lose money to better informed ones. Increases in trading volumes will decrease the spread while an increase in the number of traders with superior knowledge will increase it. The market maker opts to set the bid ask levels to maximize the difference between the losses sustained by dealing with informed traders and the gains attained by dealing with liquidity-motivated / noise-traders according to Copeland and Galai (1983).

It has been argued by authors such as George et al. (1991) that the order processing costs component represents approximately 90% of the total spread. The same survey also showed that the remaining part consists of the adverse selection component which means that these findings contest previous literature concerning the dominance of the inventory cost component. George et al. showed that previous estimates of the spread were biased due to the autocorrelation in expected returns and that the adverse selection component of the spread was smaller than previously thought. As an example of the contrary, Madhavan showed in 1993 that fluctuations in inventory make market makers adjust bid and ask order levels to induce mean reversion towards a target inventory level. Based on the works of Demsetz, Stoll (1989) would create a regression model for the spread where the cross sectional relation of the spreads to firms’ trading characteristics remained true to Demsetz’s order processing and inventory cost focused view of the spread. Variables such as volume, number of trades and firm size increase the chance of finding a counterparty to trade with and would thusly reduce inventory holding risk. The variance in the stocks’ return measures the risk of an adverse price change of a security in one’s inventory. The result indicated that the quoted proportional spreads are negatively related to measures of trading.
activity such as volume and also negatively correlated with the stock price. And as the volatility of stock-returns increased so did the bid-ask spread. Stoll noted that the empirical relation is particularly strong between the spread and the trading activity variable, volume. The author continues to stress the importance of understanding the determinants of the spread and considers improvements in financial disclosure, improvements in trading systems and uniformity between traders’ access to markets as means to reduce the overall market spread.

**THE SPREAD DETERMINANTS**

Demsetz (1968) meant that for any degree of competition, differences in spread will indicate differences in the cost of quick exchange. In this view, the fundamental force affecting the size of the spread is the time rate of transactions – as the frequency of transacting goes up the lower the cost for waiting and accordingly the lower the spread will be. In many markets, trading activity or liquidity is an excellent determinant for the spread and thusly many authors would have it act as the primary driver behind the spread. However in his book Trading and Exchanges – Market Microstructure for Practitioners, Professor of finance and business economics Larry Harris (2003) divvies and categorized the determinants of the bid ask spread to better preserve the distinction between theoretical and observable driving factors. To be able to wrap my mind around the vast ocean of academic works done on this subject and to present it to the viewer in a digestible manner, the determinants of the spread will follow Harris’s structure. The spread will be divided into primary determinants, each with several secondary determinant-subgroups, each and every one should be rationally traceable back to the constituent parts of the bid-ask spread.

**The Primary Spread Determinants**

There are three primary drivers affecting the quoted spread on any given market;

- Asymmetric information
- Volatility
- Utilitarian Trading Interest

While they each work on their own to impact the spread they do not so independently as the impact on the spread can send ripple effects back to the driving factor. One example of this is when the information asymmetry increases on the market, so will the spread. However, the increase in spread will discourage uninformed traders, reducing the traded volume as a result. Since traded volume is a secondary determinant and negatively correlated with the spread, the spread will widen even more.

- **Asymmetric information** isn’t just one of the constituting components of the bid ask spread but is described as a driver in itself. Just as earlier literature suggests, markets with asymmetrically available information for its participants will lead to greater spreads. This effect is pronounced when informed traders possess material information about the value of securities that would greatly affect the market price if it were freely available information. As explained above, market makers will adjust their pricing to cover for the cost incurred from dealing with these “in the know” counterparts.

- **Volatility** in the price of an asset will increase the spread. Traders widen their spreads when trading volatile securities to minimize the value of the timing option, since volatility increases the limit order option value. Secondly, holding inventory of volatile instruments is
not preferred to holding instruments with stable prices for the risk adverse dealer leading to an increase in the inventory holding cost component of the spread. It is also probable that volatility increases the asymmetric information component of the spread. This is because volatility in prices and uncertainty is closely related as traders must be certain they have all available information when they make value estimates. It is harder for traders to be fully informed about volatile instruments. Volatility would therefore also be a good proxy for the measurement of the asymmetric information component.

- **Utilitarian trading interest** will influence the spread in two ways. Utilitarian trading is defined as all trading that is not strictly profit seeking meaning traders seeking to hedge risks, buying and selling for educational purposes, moving money forwards or backwards through time or simply for the thrills. The comparison to noise traders could be made without too much of a stretch. Markets would not exist without utilitarian traders since purely profit motivated traders cannot all profit when trading only amongst themselves. Utilitarian traders bring activity to the marketplace. Activity or high frequency trading reduces the width of the spread as explained below but also proclaimed by Demsetz in 68. Furthermore, since the utilitarian traders are much like the noise traders in the fashion of possessing and utilizing fundamental information for trades, they dilute information in the order flow when they trade. Remember that it is the informed traders that base their investment decisions on relevant data on firms’ true value, driving the market price closer towards that “true” value. The dilution of information of the order flow will reduce the adverse selection cost of the spread. When utilitarian trading interest is strong, market activity increases enabling dealers to spread their fixed costs of doing business over more volume by high frequency trading compared to if they traded infrequently hence a reduction of the inventory cost component of the spread should follow from higher market activity. Further reduction in the inventory holding risk component is to be expected from higher liquidity but for a different reason – as it becomes easier and faster to trade securities similarly it will become easier to quickly readjust inventory holdings resulting in a reduction of the inventory holding risk. Dealers can therefore quote narrower spreads for more liquid instruments than they can for illiquid ones. Liquidity also breeds further liquidity in the consolidation of traders who perceive their limit orders more likely to be filled in a quick manner. The timing option that limit order give up are lessened in active markets making limit orders even more attractive. An increase in the popularity of limit orders will decrease the size of the spread. Conversely inactive markets will not be able to support as many dealers giving the existing ones some form of market power discouraging new entries and giving an increase in the spread of quoted bid and ask prices.

**Secondary Determinants**

There are a number of associated secondary factors that are less complicated to gauge than the primary determinants. More precisely, asymmetric information and the presence of utilitarian traders would both be outlandishly hard to measure unless divided into more tangible sub elements. These sub elements are often sufficiently correlated to the spread itself that they would, according to Harris, be classified as primary drivers by several authors. Yet, Harris’ distinction between secondary and primary determinants is not made based on their relevancy to the spread but by the degree of them being observable. The secondary determinants below are all arranged by
the primary spread determinant that they most closely represent. Note though, that many of the secondary factors do correlate with more than one primary determinant, rendering the classification somewhat arbitrary.

**Proxies for asymmetric information**

**Information Disclosure** Rules The requirement of greater information disclosure acts to reduce information asymmetry on the marketplace. Exchanges that demand listed corporations to release consistent, trustworthy and comprehensive financial reports will exhibit lower spreads than those which don’t. Stock markets do on several occasions go out on a limb to go up and above the local legal requirements for financial reporting, audits and accounting standards as an effort to reduce information asymmetry.

**Market Condition Reports** Markets that collect and distribute information of market supply and demand conditions will have a narrower spread than they otherwise would have.

**Analysts** Securities that attracts many analysts will have smaller spreads than otherwise. This is because analysts act as agents for levelling the information gap between market participants by evaluating the financial conditions and economic prospects of financial instruments. Since this information often is readily available for a relatively small fee, the playing field of gathering and appraising information becomes fairer.

**Information Vendors** Securities and commodities that the financial media covers closely will have reduced spreads compared to less noticed ones. The financial press acts much like the analysts in the way they spread information throughout the marketplace, working as an agent against information asymmetry.

**Major Commodity Contracts** Since the information asymmetry source of the spread relies on the fear from trading with informed traders, traders with private or superior information than the general market, spreads in contracts concerning economy wide risks are quite small. The reason for this is because it is very hard for traders to possess private information that would significantly affect the price. While it is conceivable that a market participant possess private information about the local supply and demand conditions of commodities such as oil, gold, wheat or copper it will generally not be enough to move the global price level much.

**Diversified Portfolios** Stock index futures generally have smaller spreads than individual stocks. Private information about a particular stock would be significant in the scope of buying and selling that stock. However in the scope of a well-diversified portfolio, the private information about a particular stock will be significantly less impactful on the overall return of the portfolio. Also, evaluating portfolios are generally simpler in the aspect of mistakes of evaluating individual stock being averaged out and offset by mistakenly evaluating another stock to the opposite.

**Diversified Stock** Corporations with a well-diversified business profile will experience smaller spreads since private knowledge about one aspect of the firm's business do not impact its net worth in the same way private information about a highly homogenous firm would.

**Established versus Emerging Industries** Firms that operate in established industries and markets are easier to evaluate than ones operating in emerging industries and markets since established markets have more certain growth prospects and established industries generate more easily predictable cash flows. Since the difficulty increases with how established the industry the
information gap between traders will also increase as they have different aptitudes for obtaining and evaluating this information. Firms in established industries are accordingly more likely to have smaller spreads than firms in emerging industries.

**Age of the Firm** Younger firms are harder to evaluate than older firms due to reasons such as new firms utilizing newer technology. New tech might be promising but is also less certain in how valuable it will be in the end. Even if the technology aspect would be the same between two firms, the new firm will have a less mature management system which brings uncertainty to the valuation. The inference from the above reasoning is that the stock of new firms will experience wider spreads than old firms.

**Insider-trading Rules** Insider trading rules are designed to hinder traders to utilize certain types of private information when exchanging on the marketplace. Markets can and often do enforce certain rules as counter measurements to stop insider trading. This will ensure to minimize the adverse selection problem for the market makers and dealers that provide liquidity. Such rules reduce the size of the spread.

**Expectations of Material Information** The bid ask orders tend to widen as the market is about to get hit by relevant financial information. This is because dealers expect that some market actors do possess private information and trade upon this information right before it becomes public information. The spread stays widened a while after the information is made public as some expect certain dealers to be able to better evaluate the information.

**Proxies for volatility**

Compared to the other two complementing determinants encompassing the total spread, volatility is quite straightforward to measure and does not elude us with proxy factors. There is however several different methods to predict future volatility partly depending on the type of instrument concerned. For corporate stock the most important factors for volatility is the extent of financial leverage within the firm, its operating leverage and uncertain growth opportunities. Commodities depend on aspects such as changing weather impacting the supply or low inventories spiking the price.

**Proxies for Utilitarian Trading Interest**

**Trading Activity** Without traders who are willing to trade without there being an obvious profit emanating from it, markets would have a hard time sustaining high volumes and frequency in trades. Markets that face a lot of trades throughout the day with high volume will have a high presence of utilitarian traders and therefore lower spreads.

**Firm Size** Large firms tend to have narrower spreads than small firms. This is basically a case of high liquidity attracting even more investors since they know their stock is going to be actively traded. The large size also attracts a lot of attention from the financial media and exhibits less volatile cashflow. These all contribute to reduce the spread for larger firms.

**Debt Size Issue** Large debt issues attract the attention of many investors and tend to be traded on active exchanges. Since many investors buy and hold until the debt matures there will be an initial spike in interest that tapers off as times go by. This is mirrored in the spread which initially is small and widens as demand for the debt falls.
**Risk Replication** Certain commodity contracts that replicate the risk profile that bothers potential traders attracts hedging interest and will therefore have small spreads. Such contracts are the most popular when the natural hedging interests on the long and short side are approximately equal.

**Volatility** Aside from the direct effect on the spread volatility also acts upon it indirectly as it catch the attention of gamblers who trade out of mere entertainment and thrill. Volatile instruments might also attract traders interested in hedging as these traders often has to adjust their positions in more volatile markets. While the direct effect of volatility acts as a steroid for the spread these two indirect effects works in the opposite way, to lessen it.
MEASURING THE SPREAD

Since Demsetz paramount work on the bid ask spread in the late 60s many has tried to quantify the effects of the bid ask spread and its constituent components into a comprehensive model. For example, Tinic (1972) tried to create a better understanding for the order-processing costs specifically. Others are Ho and Stolls paper from 1981 and Glosten and Milgroms work from 1985 trying to model inventory holding costs and adverse information costs respectively. However later studies have continuously improved on early measures. An example of this is George et al. (1991) study showing that early estimates showed downward bias of the spread but also the order processing cost component.

IMPLICIT TRANSACTION COST

Before the different approaches to quantify the spread are touched upon it is paramount to first establish what is meant by the term spread and what is meant to be measured. As explained above the spread and the liquidity term is very closely related so consequentially many of the endeavors trying to model the spread come from liquidity researchers trying to quantify transaction costs on the market. As such models for the bid ask spread will in many cases also incorporate the price impact cost of transacting. The price impact is the movement caused in the price of an asset from large buy or sell orders ploughing their way through an order book with finite order volume. Together with the bid-ask spread the price impact make up the implicit transaction costs that arise from traders affecting the price through exchange.

Compared to the explicit trading costs (commissions, fees, and taxes) the implicit ones are hard to measure because to understand the magnitude of the effect that results from trading one must have a benchmark that displays what would happen without the trade taking place. The use of the specified price benchmark method is not the only one however. More complicated econometric methods can be deployed but are more useful to estimate average transaction costs on a whole market as to evaluate the costs from an individual trade.

BENCHMARK METHODS

When using the price benchmark method the per unit transaction cost for the trade has to be calculated first. This trade price will thereafter be subtracted with the price benchmark and multiplied with the trade volume to get an estimation of the transaction cost incurred from the purchase. In the case of a sale it is the benchmark price that is subtracted by the trade price instead as shown below

\[
\text{Estimated cost} = \text{Trade size} \left\{ \frac{\text{Trade price} - \text{Benchmark price} }{\text{Benchmark price} - \text{Trade price}} \right\} (\text{purchase}) \]

A price benchmark that exhibit the price level as it would be were it not for the trade would be the ideal since there would be no doubt that the difference between it and the actual trade price would reveal the price impact of the trade. Given that the price cannot be observed in some alternative dimension where the trade never tookplace, such a benchmark is unfortunately merely hypothetical and real life studies have to settle with alternative price benchmarks that try to imitate the ideal one.

The quotation midpoint is often used in the comparison with the trade price to get the estimated transaction cost. This midpoint price is the average of the quoted bid and ask prices at a certain
time and depending on when the quotation midpoint is calculated, different kinds of cost estimates will be the result.

**Effective Spread**

If the midpoint price is calculated at the time of the trade the effective spread or liquidity premium estimate follows. The effective spread measure would show the cost of buying and then immediately selling a security on a market with sufficient depth that the quotation midpoint wouldn't change. If a trader buys a security at the ask price and immediately sells it to a bid order his loss per share would equal the bid-ask spread omitting for explicit costs. Since two transactions were executed the cost per trade or liquidity premium would be half of the effective spread. The effective spread is equal to the quoted spread unless trades occur within that range through price improvement acts from brokers. This estimate is the most commonly used by small investors by simply eyeing financial quotations.

**Realized Spread**

The realized spread is twice the difference between the trade price and the quotation midpoint price at some specified time after the trade has taken place. Traders often raise prices in as a response to aggressive buyers (buying ask orders) and lower prices in response to aggressive sellers. Because of this, realized spreads are oftentimes smaller than effective spreads. The realize spread would interest bigger dealers and liquidity providers whose profits depend on the prices they establish and liquidate positions at. Since dealers sometimes has to deal with informed traders the realized spreads will be less than the quoted spreads and the difference between the effective spreads and the realized spreads measures dealers' losses to well informed traders.

**Daily price data**

Another manner to create a benchmark to which the trade price is compared with is by using daily prices. The most common benchmark based on daily prices is the volume weighted average price (VWAP) and it is the average trade price of the day with each observation weighted by its trade volume. By comparing with the VWAP one can get an idea of how costly ones transaction was compared to the average trader on the market. The VWAP is computed by dividing the total dollar value of all trades by the total trading volume, like:

$$VWAP = \frac{\text{Dollar Volume}}{\text{Trade Volume}} = \frac{\sum_t \text{Trade Size}_t \times \text{Price}_t}{\sum_t \text{Trade Size}_t} = \sum_t \omega_t \text{Price}_t$$

Where

$$\omega_t = \frac{\text{Trade Size}_t}{\text{Trade Volume}}$$

Another way to do it is to measure transaction costs relative to opening and closing prices which would allow traders to compare their trade prices to prevailing prices before and after their trades. The daily data required for these benchmarks are relatively easy to acquire making these methods preferable compared to the quotation midpoint methods which require information about intraday trades and quotations. Also, since financial information presented during the trading day can severely alter the closing price of the asset but have nothing to do with the actual transaction costs of a trade executed earlier during the day, the closing day price benchmark would not be very accurate. The greater the time between the trade and the determination of the benchmark price the
greater the risk of inaccuracy becomes. An average price estimator exhibit less risk for inaccuracy but all daily data benchmarks have to deal with this kind of problem to some extent.

ECONOMETRIC METHODS

It is also possible to utilize statistical methods to measure the impacts traders have on prices with econometric transaction cost measurement models. Econometric methods work in the absence of bid ask order data and usually either examine price reversals or the link between order flow and price changes. This kind of approach is generally used when the transaction cost of a whole market is of interest whereas the benchmark method is often used for individual trades.

When traders buy and sell they can tend to cause price reversals. These price reversals occur due to an imbalance of uninformed traders causing the price to either rise or fall. This movement in the price is corrected as the informed traders realize the opportunity of making profit from it. This will display as a zig-zag movement or small recoils in the overall trend. This phenomenon can be over within a day or span several months. Another form of price reversal is the bid-ask bounce. This is the price oscillation that occurs in the bid-ask interval as impatient buyers and sellers demand liquidity.

Average Absolute Transaction Price Change

A very simple estimator for transaction costs is the average absolute transaction price change. This estimate is perfect when bid and ask prices are constant and when all trades take place at the bid and ask orders exclusively. But since bid and ask prices vary in reality and since not all trades occur at the bid and ask prices this estimate tend to either underestimate or overestimate the spread. This estimate would basically be the ultimate estimator for measuring the bid-ask bounce.

Roll's Serial Covariance Spread Estimator

More advances estimators of transaction cost are based on the serial covariance in the price change of assets. The efficient market hypothesis assumes that there is no such thing as covariance in prices over time, unless there are market frictions or transaction costs making prices correlate in time. The serial covariance of a sequence of price changes is the average of the product of adjacent price changes. When prices reverse such as it would in a price reversal, the serial covariance will be negative. Negative covariance would according to the efficient market hypothesis indicate that there are transaction costs on the market. Roll’s Serial Covariance Spread Estimator tries to use this insight. This spread proxy is unbiased when the sample size is large, when aggressive buyers and sellers arrive at random with equal probability and when the sequence of arriving traders is uncorrelated with changes in the effective bid and ask prices. Otherwise this estimator tends to underestimate the spread and is therefore only used when quotation data isn’t available.

The Roll estimator is as follows:

\[
\text{Effective Spread} = 2\sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})}
\]

THE IMPACT ON TRACKING ERROR

While there has been a great many attempts to quantify the tracking abilities of financial vehicles, one of the more recent ones is Frino and Gallagher’s model. It tests the significance of a number of variables that represents the tracking error in its whole:

\[
|e_{pt}| = \alpha + \beta_1 CF_{pt} + \beta_2 SPR_{t} + \beta_3 VOL_t + \beta_4 DIV_{pt} + \beta_5 INOUT_{t} + D6FR_{pt} + \varepsilon_{it}
\]
The tracking error or difference in performance can be expressed as \( R_{pt} - R_{bt} \), where \( R_{pt} \) is the return of the index portfolio \( p \) in period \( t \) and \( R_{bt} \) is the return of the benchmark index \( b \) in period \( t \).

CF represents the absolute value of the funds net monthly cash flow scaled by the index funds size. It is the cash flow of the funds that is not contributable to capital gains or dividends. In other words, the working assumption is that capital gains and dividends are kept in the fund leaving the cash flow above and beyond that amount as the estimated cash flow-variable.

New cash flow to a fund can be calculated in a few ways, two are described below. The common denominator for both methods is that new cash flow is defined as the change in total net asset value minus the appreciation of the fund's assets. The appreciation is expressed as the total NAV at the beginning of the period multiplied with the rate of return the fund earned during this period. The result of this exercise will be denominated in absolute terms which can be problematic given that large funds tend to have large cash flow regardless of their performance. To alleviate this bias an alternate method, outlined by Gruber (1996) can be deployed – the normalized cash flow – a growth rate which weights each cash flow by the NAV of the fund at the beginning of the period. However the second measure tends to magnify the reported cash flow for funds with undersized NAVs.

The DIV variable is the dividend yield of the securities that comprise the index. INOUT represents the percentage market capitalisation of the equities excluded and included from the index/indices. FR is a dummy variable with either the value 1 or 0 depending on if the observation is from a fund with a full replication strategy or not.

VOL is a measure of the volatility of the benchmark index and is simply the standard deviation of daily returns of the fund each month in accordance with Frino & Gallagher's (2002) work. Alternative measures of volatility can also be employed – for example Parkinson's high-low estimator as Wiggins (1991) paper explains. This technique was developed as a way to better tackle the problem of forecasting random walk series where let's say a price of an asset in period \( t \) will essentially equal the price observation on \( t-1 \) plus a white noise term, \( \varepsilon_i \). This is true under the assumption of a fully efficient market as explained by Fama in 1969. Parkinson (1980) showed that a plain standard deviation analysis of the close-close price data will miss the information provided by incorporating the lowest and highest price observation during the period.

Frino & Gallagher (2002) explains that when dealing with daily observations the SPR would be defined as the market capitalization weighted mean bid-ask spread for all securities in the index. For each stock in the index the daily percentage bid-ask spread can simply be calculated by dividing the spread by the stocks theoretical price demonstrated by McInish & Wood (1992). Aber et al. (2009) also conducted a research where the volatility and tracking ability of ETF was compared to index funds. Since it was concluded that the closing price should not have as much influence as the price at which the largest volumes are traded Aber et al. used a theoretical midpoint price intended to alleviate this problem – to extract an intra-day price at which trading occurs in the highest volume. This midpoint price was not weighted by intraday volume but simply by adding the closing price \( P_C \) to the opening price \( P_O \) over 2.

\[
\text{Midpoint Price} = \frac{P_C + P_O}{2}
\]
Frino & Gallagher (2002) then weighted the variables by each stocks market capitalization share of the index. An issue with such a midpoint price could be that it relies on the assumption of non-zero skewness in the distribution of the volume over the daily price range. However this is not necessarily evident as Hong and Stein (2003) showed that high trading volumes lead to greater negatively skewed returns since some traders are short sales-constrained which leads to the unfolding of previously hidden information during market declines.

**IMPORTANT MARKET CHARACTERISTICS: LIQUIDITY**

In my attempt to discern the effects liquidity have on the relevant transaction costs for my study I encounter some of the inherent difficulties when dealing with the liquidity term. The definition of liquidity, how it should be measured and how the symptoms of it is exhibited are all notions that are far from set in stone. This becomes painfully evident as the very thing I aim to explain by the employment of the liquidity term is often used as a measurement of the liquidity concept.

As the inner workings of the bid-ask spread and its influencing factors are laid bare, Harris would have them separated into three major groups as displayed above. However, keep in mind that this distinction is the effort of one particular author to accumulate and present academic literature spanning several different fields in a sensible way. While this setup is very useful to help the reader grasp the concept of the bid-ask spread, it does nonetheless hide some of the intricacies of the spread that’s revealed by studying each field by its own.

Two field of inquiry is closely related to the bid-ask spread and its origins. None of them should come as big surprises considering the earlier exposition on what the spread is and what it stems from but instead seem like obvious extensions of the earlier explanation. First I touch on the subject of liquidity which roughly translates into the utilitarian trading interest or trading activity as Demsetz would call it. Liquidity research basically studies the ease of transacting on markets and the costs arise from it. The liquidity term is also strongly correlated with volatility since volatility tend to increase as liquidity drains from the marketplace. The second one, market microstructure theory, would encompass the asymmetric information factor but is also described as the origin of liquidity.

The spread is a part of larger market friction concept called liquidity. Many drivers of liquidity affect the magnitude of the spread and accordingly the factors that affect the spread also affect liquidity. The literature on liquidity is vast and established. As we are trying to formulate the mechanics of the spread and how to amount for it, it behooves us to lend an ear to a very comparable field of inquiry and maybe take some of its solutions to heart.

The liquidity term has many dimensions and means different things to different people. Regulators measure liquidity to determine which market structures are the best since high liquidity suppresses high volatility. Traders measure liquidity to determine if their trading strategies can be implemented without taking too heavy of a hit from transaction costs. Everyone in the market deal with liquidity in some way; impatient traders or market order users take liquidity. Dealers or limit order users provide liquidity to the market and the brokers and exchanges organize liquidity. Even though they all try to measure the same concept, a single all exclusive measurement for capturing liquidity has not yet been developed according to Amihud et al (2005).
The dimensions of liquidity all contribute to the confusion about it and can be summarized by:

- **Immediacy**, immediacy refers to how quickly trades of a given size can be executed at a given cost.
- **Width** refers to the cost of doing a trade of a given size. For small traders, the bid-ask spreads constitute the major part of this cost even though it also includes brokerage fees.
- **Depth** refers to the size of a trade at a given cost and is measured in units available at a given price of liquidity (Harris 2003 pp. 394-409). Market depth is very connected to the quoted bid-ask spread. Higher depth implies better liquidity, i.e., greater ability to execute trades without large fluctuations in the price (Madhavan 1992).

Amihud et al. (2006) define market liquidity as “the ability to trade large quantities of an asset quickly, at low cost, and without moving the price”. During illiquid market circumstances investors may be unable to construct optimal portfolios and the potential benefits of diversification could be lost (Fleming et al 1996).

In defining the “liquid market” of finance theory, which is a premise for the option equation to hold, Black (1971) noted that this is a market in which a “bid-ask price is always quoted, its spread is small enough, and small trades can be immediately executed with minimal effect on price.” Grossman and Miller (1988) pointed out that we can measure the liquidity of a market by looking “the ability of executing trades under the current price quotes price – and time-wise.”

**Past and Present Measures**

Traditional measures of market liquidity include trade volume or the number of trades during a specified period of time. Past studies of market liquidity have mainly focused on what is called static indicators such as such as turnover, bid-ask spread, as measures of market liquidity.

However, Muranaga and Shimizu (1997) argues that in order to examine how market liquidity affects the price discovery process in an actual market, not only should the static aspects of market liquidity be examined, but also dynamic ones; probability of quote existence, price volatility, gross order book volume (buying order volume plus selling order volume), and net order book volume (buying order volume minus selling order volume). They also calculate the volatility of gross order book volume, and this indicator can be regarded as representing the ease with which the order book is restored to its original state after certain decrease in orders, it’s the resiliency of the market.

However Goyenko et al. (2009) states that another way of measuring the liquidity is the Amihud (2005) measure. Using cross-section and time-series on the “daily price response associated with one dollar of trading volume” the price impact can be calculated quite accurately.

Keim and Madhavan (1996) asserts that when dealing with mid-size transactions, the bid-ask spread is an agreeable measurement for estimating the illiquidity cost proxy, while Kraus and Stoll

---

6 The process by which demand and supply forces equate and clear.
(1972) show that the price impact is a more suitable cost proxy when dealing with upstairs trading.

Korajczyk and Sadka (2008) highlight the problem of many liquidity-measures requires quote data and continues that while market volume might be a good proxy for liquidity additional proxies for liquidity can be mixed together to complete the picture. The turnover measure is a common way to calculate liquidity and is basically a measure of the ratio of the volume during time period $t$ to issued shares at the end of the period. In the expression below $\text{vol}_{i,j}$ stand for the volume in security $l$ and $\text{SO}_{i,t}$ is the number of outstanding shares at the end of the period. A high turnover rating indicates high liquidity.

$$\text{Turnover}_{i,j} = \frac{\sum_j^t \text{vol}_{i,j}}{\text{SO}_{i,t}}$$

Another measure for liquidity is the $Q\text{spread}$ which is basically a measure of the bid-ask spread expressed in percentage and is negatively related with liquidity. First off the bid ask spread range for each observation is divided by the bid-ask midpoint which is an average value of the ask and sell price. All these are then summed up and multiplied by the ratio of 1 to the number of observations $n$.

$$Q\text{spread}_{i,t} = \frac{1}{n_{i,t}} \sum_{j=1}^{n_{t}} \frac{\text{Ask}_{i,j} - \text{Bid}_{i,j}}{m_{i,j}}$$

Where $\text{Ask}_{i,j}$ and $\text{Bid}_{i,j}$ are the ask and bid price for security $i$ for observation $j$.

And where $m_{i,j}$ is equal to

$$\frac{\text{Ask}_{i,j} + \text{Bid}_{i,j}}{2}$$

**FOR PRICE DISCOVERY**

When market efficiency is considered in the context of a market’s price discovery function and the information content of price, market liquidity can be regarded as a factor which effects market price uncertainties—uncertainties in the sense that market prices do not reveal all available information or in the sense that market price will temporarily diverge from the market-clearing equilibrium price—or price discovery function, and will, as a result, effect market efficiency. Studies which review market efficiency and market liquidity from the viewpoint of the uncertainties of transaction execution price and information reflecting process on price such as Brown & Zhang (1997) and Easley & O’Hara (1992) underscores how market efficiency can be improved by increasing market liquidity. Specifically, a decline of market price uncertainties accompanied by an increase in market liquidity will, through the decline of liquidity premiums such as the bid-ask spread and market impact, improve market efficiency, resulting in efficient fund and risk allocation. Muranaga and Shimizu (1997) asserts that if the mechanism by which market liquidity affects the

---

7 Contracts for an exchange-listed stock that is not executed through the exchange. The transaction is performed with different terms and pricing than would be available on the regular market.
price discovery process can be clarified, it is believed that it will provide a good reference in considering measures designed to improve market efficiency.

Madhavan (1992) demonstrate that the process of finding market clearing prices, differ in types of orders, the times at which the trades take place, the degree of information on the market at the time the orders are submitted and the dependence on market makers to provide liquidity.

**FOR TRANSACTION COSTS**

The measurements used by researcher to capture the transaction costs of market participants are numerous and a consensus on which measure is the best does not exist. In their paper from 2009 Goyenko et al. (2009) tested a set of widely used proxies for liquidity to see if they actually measured transaction costs. The results confirmed that previous research had developed good measurements that actually captured the transaction costs on the market. The first test compared the monthly and annual low frequency data with high-frequency data in capturing liquidity measurements. In most of the cases the correlations were high enough and the mean squared error low enough that the investment in using high-frequency data was not worth the cost.

Three measurements dominate, with their high correlation and low mean squared error predictions, in estimating the monthly and annual effective and realized spreads. The simplest of these is the analytic Effective Tick, which is a proxy of the effective spreads based on observable price clusters. The measure that delivers the best performance is Holden, which combines serial correlation with the Effective Tick. The third measure is LOTY-split, both performance and computational requirements lies between the previous two measurements. Considering the ease of computing the Effective Tick and its performance, Goyenko et al. considers this measure to be the best to use for estimating the monthly and annual effective and realized spreads.

Amihud and Mendelson (1986) implemented the quoted bid-ask spread as a measure of illiquidity. They found evidence that asset returns include a significant premium for the quoted spread and were the first connection made between market microstructure and long term asset pricing. This evidence was later questioned by Eleswarapu and Reinganum(1993) by showing that the return premium was mainly a seasonal effect, now known as the January effect.

Madhavan demonstrated in 1992 that high volume trading shortens the holding period for market makers and consequently lowers the inventory control cost fragment of the bid-ask spread. Markets with high volume will ceteris paribus have lower costs than other markets – thus attract even more capital. This consolidation-process will result in a decrease of informed based trader presence and a further decrease in spreads.

In 2005 Amihud, Mendelson and Pedersen measured illiquidity as the cost of immediate execution. The spread contains a premium for immediate buying and selling. The relative spread on stocks has been found to be negatively correlated with liquidity characteristics such as the depth, the number of shareholders, the number of market makers trading the stock and the stock price continuity. Their paper presents evidence of average portfolio risk-adjusted returns increase with their bid-ask spread. Further, the study highlights the important link between stock market microstructure and determining asset returns and how structural policies leading to liquidity increases can reduce the firm's cost of capital and ease the trading and exchange process.
As explained above, liquidity doesn’t only give a good estimation of the presence of the utilitarian trading interest portion of the spread but also gives an indicator how the likelihood of high volatility. Remember, liquidity is negatively correlated with volatility under normal circumstances.

**IN MARKET MICROSTRUCTURE THEORY**

The importance of market microstructure as a main determent for liquidity has become more evident in recent time by works from academics such as Amihud, Mendelson & Pedersen (2005) and Parlour & Seppi (2003). The market microstructure effects the bid ask spread in many ways. For example the spread will depend on the size of the trade. Huang and Stoll (1997) suggested that large trades tend to have larger spreads because information leakage from upstairs trading.

When market liquidity is discussed in market microstructure theory, it is often the case that more practical concepts are introduced, such as the “cost of changing positions (tightness),” the “trade size or thickness of the order book-profile 8 required for changing prices (market depth),” and the “required period of time to recover from price fluctuation caused by a sudden shock or to reach a new equilibrium (market resiliency)” according to Kyle (1985).

---

8 Order book refers to a panel which provides traders with bid-ask prices and volume offered per price.
Market microstructure theory deals with the how market architecture affect exchange. It examines how operating procedures such as regulation of management and dissemination of financial information influence transaction costs, price discovery, liquidity etc. Transparency and disclosure of financial data such as prices, quotes, volumes, sources of order flow and market participants lay as the foundation for a great deal of theoretical and empirical research. Madhavan (2002) work shines light on the intricacies of this field and will borrow its setup to the following section of this paper. Market microstructure theory is useful in building an understanding of the level of asymmetric information and utilitarian trading interest on markets. How information is handled and the rules surrounding trading will be pivotal for attracting or discouraging noise traders and the economic calculation that market makers do when setting the spread levels.

However, determining the effects of market architecture on transaction costs is complicated since high quality data of sufficient amount is lacking. Adding to this problem is the fact that traders adjust their strategies in response to certain market structure elements.

Market architecture includes a set of characteristics which includes but are not limited to:

- **The degree of continuity** Periodic systems limit trading to specific points in time while no such limitations are present in a continuous system.
- **Dealer presence** In order-driven markets no party function as a market maker, individual investors create and display buy and sell orders openly while in a dealer market there is dealer intermediation and only the bid and ask prices of market makers and specialized dealers are displayed.
- **Price discovery** Some exchanges contains assets that follow the process of price discovery that happens on other markets. The second kind of market innovate prices by having superior information share.
- **Automation** Markets are more or less automated in the sense that in one end there are floor trading exchanges where traders complete the transactions on the floor and on the other end there are screen based electronic systems.
- **Protocols** Program trading, minimum tick, when to halt trading, circuit breakers and rules adopted for opens reopens and closes; these rules or protocols that can vary between exchanges.
- **Pre- and post-trade transparency** Transparent systems provide pre trade information about quotes and depths and post trade information about actual prices, volumes etc. Non transparent systems fail in this regard.
- **Information dissemination** The extent and speed of information distribution separate markets.
- **Anonymity** Some markets offer a great deal of anonymity for example regarding identities and hidden orders while other markets do not.
- **Off-market trading** Does the market permit off market trading, after hours trading or not?
- **Fragmentation** Fragmented markets in which asset prices can deviate from the general or global price and centralized markets in which all the orders are addressed to the same place (Blais 1993).

The information characteristics listed above can be grouped into pre-trade and post-trade dimensions where pre-trade transparency implicate information such as large trade order imbalances as well as disclosure of current bid and ask quotations, depths and limit orders away from the best price. Post trade transparency on the other hand refers to the timely and widely available declaration of past trades, execution timings, volumes, prices and buyer / seller identifications.

The impact of the differences between order-driven and quote-driven markets has been explored by Ahn et al (2002) who find that both adverse selection and order handling costs on Tokyo Stock Exchange (TSE), which is order driven, differs from the quote-driven New York Stock Exchange (NYSE). First, the adverse selection costs arise at different times during the trading day. Second, the study on the TSE finds that adverse selection costs increase with trade size while order-processing costs decrease with it. Previous research shows evidence that medium trades on the NYSE contain more information about the fundamental price than large trades. This difference can be traced back to how the two exchanges deals with large order flows. On the NYSE, large order flows are sent upstairs and made through a search-brokerage mechanism. On the TSE, all transactions are made electronically on the same setting with the investor identity unrevealed.

**INFORMATION ASYMMETRY**

It has been argued that delayed disclosure of large trades on London stock exchange has played a role in inter-market competition and order flow migration. Floor systems in contrast to electronic systems are not very transparent as they tend to hide fringe customer limit orders, or the ones away from the best quote. Oppositely, electronic limit order book systems tend to reveal limit orders away from the best ones which tend to decrease information asymmetry. Disclosure of identities of traders and their motivation, revelation of order imbalances at the open or during trading halts, front running when brokers trade ahead of customer orders, upstairs and off-exchange trading has been discussed as a force that affects a range of aspects such as post-trade transparency. Generally, the more post-trade information is withheld the less transparent the market will be and the higher the information asymmetry will be.

Transparency is less of an issue in automated trading systems where the constituent parts of order flow cannot be distinguished. Presentation of some degree of order flow composition transparency in floor based systems is however conventional practice.

It has been shown that transparency will allow dealers to identify traders that possess private information thus reducing the adverse selection component of the bid-ask spread. Madhavan contest this view by arguing for the case that disclosure of financial information help investors estimate the extent of liquidity traders thus making the market more prone to be affected by the presence of asymmetric information.

Post trade transparency in the form of volume and price information has been demonstrated to lower the information asymmetry. Traders also believe that changes in asset prices are more fundamentally sound if they are supported by big volumes. Higher post trade transparency thus leads to a reduction of the spreads and an increase in volume traded. On the other hand, too much
Post trade transparency can push traders to off-market venues and invoke fragmentation to the marketplace. These negative effects of transparency are however likely to be larger in markets that lack depth according to Madhavan (1996).

Concerning pre-trade transparency however there have been studies done indicating a detrimental effect of high transparency. The display of the limit order book on the FTSE 100 Index (FTSE) leads to a reduction of liquidity due to an unwillingness to submit orders in a highly transparent system by limit order traders. Informed traders will favour anonymous trading systems whereas noise traders or liquidity traders will prefer greater disclosure.

**Quote vs Order Driven, Periodic vs Continuous Markets**

Madhavan studies the process of price discovery under two alternative forms of market microstructure: an order-driven system and a quote-driven system. The quote driven system is a continuous system since order will be executed as they arrive. The order-driven system can function either under a continuous exchange, where order will be executed as they arrive, or a periodic system, where order are executed at certain times during the trading day. The results showed that a quote-driven system provides a more efficient price discovery process than a continuous market. Further, during certain circumstances periodic trading system will be able to function where a continuous market fails. However, a periodic trading system inflicts higher information costs for traders who must gather their own information as they can't observe price quotations. Also of note is that a high degree of information asymmetry can cause a continuous market to fail, but closing the market in an attempt to solve it can actually exacerbate the problem.

**Decimalization**

With a reduction of the minimum increment in which stocks prices are quoted – decimalization – the profitability of providing liquidity goes down. Empirical studies of market decimalization have shown that the overall market liquidity will fall along with the quoted spread. Comparisons of the dealer heavy National Association of Securities Dealers Automated Quotations (NASDAQ)-market with NYSEs' specialist oriented system provide insights of why the expectations of the bid-ask spread is not reflecting the competitive system of the NASDAQ exchange.

One of these explanations is that dealers on the NASDAQ exchange by e.g. order flow preferring collude to set wider spreads than what would be in a competitive environment according to Biais (1993). A change in the tick-size has been shown to effect the bid-ask spread by Harris (2003) when he published the study on how the aftermath of a change in the tick-size from $1/8 to $1/16 resulted in tighter spreads, decreased quotation sizes and higher trading volumes for stocks on the NYSE and American Stock Exchange (AMEX) with a price below $10. However, Ahn et al. (1996) failed to find evidence of significant higher trading volumes after the change in the tick-size and the degree of the narrowing in spread was much smaller than previously predicted by Harris (1994). Moreover, Ahn et al. could not find the effect of an increased market depth previously predicted by Harris.

**Fragmentation**

When a marketplace is fragmented one or several stocks will be listed and traded on other exchanges. This raises the question of how the price formation processes of the same asset on two market places interact with each other. One of the markets will act as a guide to the other by incorporating new information faster and showing the corresponding price correction which the
other(s) follow. Because of the arbitrage opportunities which can surface between the markets the price difference are contained within certain parameters that market friction allows for. The market which innovations drive the reaction of other markets has greater information share. A market could possibly have a larger information share than another and still display wider spread between bid and ask orders as long as the innovations on that market drives the reactions on the one with a narrower spread. The spreads are assumed to be positively correlated with the information component of a trade, remember informed and noise traders, and therefore the market which has the greater price discovery, being most informative, can exhibit a wider spread.

Hasbrouck (1995) made a study comparing the price discovery of commonly traded stocks on NYSE and smaller regional markets and found that the information share on NYSE was larger but still displayed the lowest spread on average. Additionally, Biais (1993) study also discredited this idea when the expected bid ask spreads in fragmented markets were the same as in centralized markets although the variance of the spread is larger in the latter one.

AUTOMATION

A study focusing on order flows as a proxy for the spread as the QQQ ETF proliferated from AMEX to other markets. It found that the bid-ask spreads decreased when ETFs was made available on NYSE. The NYSE trading has also reduced the variance of the pricing error which is evidence of higher market quality. Electronic communication networks (ECN) made a large contribution to the price discovery process as the increase in competition for order flow improved market quality and reducing the spread according to Tse and Erenburgs (2003) paper.
THE HYPOTHESES

Hypothesis 1: If the market microstructure and or the liquidity of the OMXS and the NYSE differ then the size of the transaction costs will also differ.

Hypothesis 2: If the size of the transaction costs differs on the OMXS then the cost relationship between ETFs and index mutual funds will be different from past research.
METHOD

CHECKING FOR SYSTEMATIC DISPARITY IN LIQUIDITY

Among all the difference methods of measuring transaction costs the Roll’s covariance measure are deemed best suited for this paper and will be employed to determine if there is a systematical difference in liquidity between OMXS and NYSE. This is mainly due to the ability of econometric bid ask spread estimates to handle large sample size, as in the example of a market wide investigation. The Average Absolute Transaction Price Change measure, while being of the econometric nature, does rely on the assumption of constant bid ask prices and does not seem fit for the task. The average absolute transaction price change measure is best suited for measuring the bid ask bounce and not transaction data over longer a period of time. Furthermore, the benchmark methods require data of specific trades to which a constructed benchmark is compared.

Roll’s covariance measure will serve the purpose of revealing whether there is a systematic liquidity disparity between the two markets. But if the markets are to be compared to each other the estimates have to be normalized as Roll’s measure will take the absolute value of the indices into account. To do this the covariance of the percentage change in price will be used rather than the absolute price change.

QUANTIFYING THE SPREAD’S IMPACT ON TRACKING ERROR

Now, while the Rolls estimate will give me an insight on whether or not there is a systematic difference in liquidity between the two markets it will not reveal to what extent this is plays a role in the performance of index funds. Remember that the transaction cost component of ETFs and index funds fall on different sides of the implicit / explicit transaction cost divide.

The transaction costs of ETFs constitutes an added cost on the shareholder level and can be treated in an arithmetic fashion as its effect is very direct. This is not the case for index funds. In this case it is the fund manager who pays the transaction costs, not the shareholder, and this manifests itself in a much less straightforward way of accounting for differences. Keep in mind that the owner of index fund shares do pay the price of transaction costs in the end. This is ultimately experienced in the form of the reduced return the fund produces as the increased cost of managing said fund trickles down to the end consumer.

But to what extent an increase in the cost of transacting on a market affects index funds ability to perform / track an index, is not as apparent. Frino and Gallaghers regression model will be employed to gauge whether the spread component is a significant contributor to performance deviations in the two index funds, Vanguard 500 and SPP Aktieindexfond Sverige. The impact of the spread differs between the two instruments. For ETFs it is the investor who bears the cost directly at purchase of shares on the secondary market. In the case of index funds, the implicit trading cost will express itself as an inability to track the underlying benchmark. A few clarifications of how it will be deployed follows.

To start the daily return R_RAW of the ETFs, Index funds and corresponding index are calculated by subtracting the price P of the funds or index on day t – 1 from the price on day t. The difference is divided by the price on day t – 1, generating the daily raw return percentage.
To get the dividend variable, \( DI\), the index-dividends expressed as a percentage will be multiplied with the market capitalization of the index. That is, the dividends will be weighted with each stock's share of the total index value so that it is proportional. In the case of S&P 500 the daily percentage change were multiplied with the daily market capitalization values of the index. This procedure was not done with OMXS30 which instead was subtracted from the SIX30RX, which is the same as the OMXS30 index but contains dividend-yields, to get the daily percentage dividends yield.

Two variables are going to be omitted from the Frino and Gallagher model. The first variable is the percentage market capitalisation of the equities excluded and included from the indices, \( \text{INOUT} \). The first problem encountered with this variable was the extent of the information needed for S&P 500 where a considerable amount of equities had entered and exited – whereas OMXS30 had stayed unchanged during the same period. The sheer workload of minutely sift through a 500 stock sized index for current and past listed companies and their weighing is not only time consuming but the information required proved to be hard to come by.

The second variable that was omitted is the dummy variable indicating full replication or not (https://secure.msse.se, https://personal.vanguard.com). While the Vanguard 500 fund does aim to fully replicate the S&P500 index by holding roughly the same composition and weight of stocks SPP Aktieindexfond Sverige does not. While existing on two different sides of the dummy variable it will still not be included for the simple reason that the sample size of funds is merely two. If the study were larger and included several funds the dummy variable might be useful in picking up explanatory information.

The regression variables will be tested for heteroscedasticity with White’s test. White (1980) tests the null hypothesis: \( H_0: \sigma_i^2 = \sigma^2 \). That variance is consistent throughout the sample and doesn’t show signs of patterns across the data.

**ACCOUNTING FOR THE SPREAD’S IMPACT ON THE COST RELATIONSHIP**

Having answered the question whether there are systematic differences in the transaction cost due to market structural differences, and whether it differs to a degree at which it should be taken into account for estimating index fund performance, the crux of the problem is ready to be handled. It is the competitiveness of the two instruments for the end consumer that are in the epicenter of this study and equipped with the newly acquired knowledge of the importance of the implicit transaction cost, we can now try to give a more complete answer to this question.

Earlier studies comparing the cost relationship between ETFs and index funds have been conducted by both Dellva (2001) and Kostovetsky (2003). Dellva compares the qualitative characteristics of ETFs to mutual fund, elaborates on their aptness to fill different types of investors needs but also their tax and cost effectiveness. Dellva takes into account commissions and expense ratios, different lengths in holding period and various amounts of total cash investments.

Advancing the work of Dellva, Kostovetsky (2003) undertook the task of creating a quantifiable single and multi-period model for explaining the explicit cost relationship between the instruments. Much like Dellva, Kostovetsky assumed perfect tracking capabilities for both of the instruments as a
way of both simplifying the exercise but also an attempt to focus the study in the dichotomy of explicit and implicit costs.

**KOSTOVETSKY’S ONE-PERIOD MODEL**

Even though the cost differences between the two types of securities are small they are still essential to analyze. Kostovetsky (2003) argues that one can always validate paying steeper costs associated with actively managed funds with better management. For the passively managed funds, which with I am dealing with, costs are the only factor in deciding which instrument to pick.

This econometric set-up that Kostovetsky used to compare the performance of ETFs and index funds on the NYSE relies on the assumption of perfect tracking ability of ETFs and index funds so that the non-tracking error costs will be the only deciding aspect between the instruments. Explicitly, if you are certain that two assets will exhibit identical performance in the future but differ in the transaction costs then these costs will be the only factors that separate them. Kostovetsky firstly weighed the costs of index funds against the costs of ETFs in a one period model.

Imagine an investor who wants to invest an amount $I$ in an index tracking instrument for a time period $t$. The whole amount of the investment, $I$, is thusly not invested at once but is spread out over smaller investments through time. This could simulate either a dollar cost averaging strategy or simply a smaller investor, investing part of his or her monthly salary as it is received. The brokerage commission $C$, is paid at each given purchase of $N$ numbers of assets at prices $P_0, P_1, \ldots, P_{n-1}$ ($P_i$ is the price of the fund at each transaction). $C$ is assumed to be constant, independent from the initial investment but dependent on the number of purchases. This is untrue for ETFs on the secondary market which have bid-ask spreads. Kostovetsky study only deals with highly liquid ETFs like QQQs or SPDRs “which trade at negligible bid-ask spreads compared to commissions charged by brokers”.

A share $\alpha$ of the capital gains $k$ is dealt to the shareholders, at the end of the period. A fraction $ak$ of the capital gains earned by the investor will be reinvested into the fund, but not before a percentage $(\alpha k)\tau$ in taxes are paid to the state. At the end of the period a part of the initial investment is also distributed as dividends $d$. Before reinvesting the distributed dividends the investor must pay taxes. The percentage of the dividends paid in taxes is denoted by $d\tau$. $\alpha$ and $d$ are not dollar values but rather percentages of the total post-commission investment ($I-CN$).
Lastly the remaining value is charged an expense ratio $e$. The expense ratio includes management fees and can include tracking error factor if it was to be taken into account for. Capital gains $k$ is measured as the percentage increase in price during period $t$.

With the above described variables Kostovetsky now shoulders the task of developing a formula to get the final value of the investment, be it in ETFs or in mutual funds.

The price distribution is directly related to the capital gains earned by (Expression 1):

$$k = \left(\frac{1}{P_0} + \frac{1}{P_1} + \ldots + \frac{1}{P_{n-1}}\right) P_n - 1$$

The value of the investment at time $t$ before dividend and capital gains taxes are paid is (Expression 2):

$$(I - CN)(1 + k + d)$$

This basically is the value of the investment minus the initial costs (flat rate of commission at each purchase) – plus the untaxed return of the investment. The untaxed returns are constituted by the capital gains which are the increase in the security’s value or price during the period and the dividend yield that are paid from corporations to shareholders. But the investor will not get the full worth of the capital gains- and dividend distributions seeing as they are subjects to taxation. The value of the capital gains tax that has to be paid is (Expression 3):

$$\{ (I - CN)k \} \tau_k$$

...where the tax-rate $\tau_k$ is multiplied by the share $\alpha$ of capital gains $k$ which the worth of one’s initial investment after commission fees $I-CN$ is based on – by multiplying the investment with the capital gains ratio $\{(I-CN)k\}$ I get the amount susceptible for distribution to the shareholder. And the value of the dividend taxes that have to be paid is (Expression 4):

$$\{(I - CN)d\} \tau_d$$

-$d$ denotes the yearly percentage rate at which I multiply the initial investment after commission fees $I-CN$ to get the yearly dividend yield. The dividend tax-rate $\tau_d$ times the nominal dividend distribution $\{(I-CN)d\}$ signifies the value of the dividend taxes.

The value before expenses is (Expression 5):

$$(I - CN)(1 + k + d) - \alpha \{ (I - CN)k \} \tau_k - \{ (I - CN)d \} \tau_d$$

Expression (5) is expression (2), the value at time $t$ before taxes minus expressions (3) and (4) – the capital gains- and dividend distribution tax equations.

The $I - CN$ term will be taken out from expression (5) and multiplied by $(1 - e)$ to get the final value of the investment – a mere arithmetic exercise (Expression 6):

$$\text{Final value} = (1 - e)(I - CN) \times \{ 1 + k(1 - \alpha \tau_k) + d(1 - \tau_d) \}$$
This formula, used to calculate the final value, can be used on both ETFs and index funds. It is basically the initial invested sum of money less the commission fees paid at each instance of investing and subsequently adjusted to account for the funds expense ratio. This is then multiplied with the capital and dividend returns ratio, adjusted for tax. If equation 6 were to be used to calculate the final value of a mutual fund investment, \( C \) would be set to 0 since Kostovetsky assumes that no-load funds are readily accessible to the investor. And when used in the case of an ETF investment the capital gains ratio would be equal to 0 since ETFs don't distribute capital gains.

Expression (7) subtracts the Final value – expression (6) for index mutual funds with the same expression for ETFs. Essentially Expression (6) for Mutual funds subtracted with Expression (6) for ETFs is what Kostovetsky call The investor choice equation. The result is the value difference at time \( t \) after investing a sum \( I \), \( N \) times during one period after taxes and expenses. Since \( k \) and \( d \) is assumed to be the same – expression (7) encompass the one period cost comparison model by Kostovetsky. More on this equation below. (Expression 7)

\[
FV_i - FV_e = [(1 - e_i)(I - C_iN)(1 + k(1 - \alpha_i \tau_k) + d(1 - \tau_d))] \\
- [(1 - e_e)(I - C_eN)(1 + k(1 - \alpha_e \tau_k) + d(1 - \tau_d))]
\]

Kostovetsky continues to divide the variables into three big categories depending on which investors and funds they affect. Since tracking error is hard to model Kostovetsky assumes perfect tracking which makes the variables \( k \) and \( d \), capital returns and the dividend yield ratio, global in the sense that they are the same for all funds tracking the same index. Even if perfect tracking were not assumed, \( k \) and \( d \) could still be considered global if the tracking error difference is included in the expense ratio \( e \). This is easier to grasp if one were to consider the underperformance of a fund due to tracking error as a cost borne by the investor, at which point it would be as if the fund had a greater expense ratio.

The semiglobal variables are the ones that are the same for all the investors but differ amongst assets. Such are the brokerage fee \( C \), the expense ratio \( e \), the capital gains distribution ratio \( \alpha \). Lastly the variables that vary between investors, the local variables. These are the initial investment \( I \), the number of purchases \( N \) and the tax rates \( \tau_d \) and \( \tau_k \). Kostovetsky notes that it is the semiglobal variables that tend to constitute the only differences between index funds and ETFs. The investor will choose an index fund over an ETF if the Final Value for an index fund – Final value for an ETF > 0. Since only semiglobal variables differentiate the two expression (7) encompasses all the semiglobal variables in a cost comparison model. The semiglobal variables for index funds is signified with \( i \) and an \( e \) for the ETF variables.

**MODIFYING THE ONE-PERIOD MODEL**

To accurately incorporate the spread components exclusive contribution to the profitability - relationship between ETFs and index mutual funds the cost comparison model of Kostovetsky needs to be reworked.

The manner in which the spread functions on index mutual funds profitability compared to how it effect the investor dealing with ETFs differ in the respect that the costs of paying the bid ask spread will fall on the fund managers in the case of index mutual funds – resulting in an inaccuracy of its index tracking ability consequently transferring the cost on the investor. ETFs unlike mutual funds are traded on the exchange – much like the name “exchange traded” suggest – with a market maker setting bid and ask prices to cover the costs of immediacy. An investor of ETFs thus pays the
margin between the theoretical price of the security and the actual price that is paid as a spread cost – coming into effect the moment the purchase takes place.

The cost of the bid ask spread in percentage of its theoretical price $S$ will occur after the cost of commission have been subtracted from the sum invested. This percentage component has to be multiplied with the investment amount $I$ to capture the nominal spread cost. Observe that the spread expenditure cut into the amount of purchased shares – given a fixed $I$ – requiring that $S$ is deducted ahead of the remaining costs.

$$FV = (1 - e)((I - CN) - IS) * (1 + k(1 - \alpha \tau_k) + d(1 + \tau_d))$$

In the case of index mutual funds the bid ask spread is not paid on a shareholder level but on the fund level thus requiring a different way of modeling it. In this case – the amount of securities received given a certain $I$ will not depend on $S$ since the cost of market friction in this form are borne by the fund resulting in a seemingly continuous and cumulative cost which materializes as a underperforming fund – ceteris paribus.

$$FV = (1 - e - S) * ((I - CN) * (1 + k(1 - \alpha \tau_k) + d(1 + \tau_d))$$

An investor will choose to invest in an index mutual fund if the Final value of the index fund – Final value of the ETF $> 0$.

**Kostovetsky’s Multi-Period Model**

While the single period model can establish which instrument the value-maximizing investor should prefer for an investment period of a single year – does this relationship hold true if one hold these instruments for a longer period of time, spanning several years? To measure how well these two instruments’ relationship stands the test of time – whether there is more to their workings than a holding period of one year discloses – I need a model that incorporates multiple time periods. To do this the theoretical framework of Kostovetsky will once more be used as his multiperiod model is deployed with a few alterations to incorporate the bid ask spread.

By using this model I can determine whether different time horizons have positive or adverse effects on the Final value assessment of the two instruments. The model describes the investor initially making the same kind of investment as was made in the one period model. That is; periodic investment in the security on a monthly basis totaling $N$ times – each at which a broker’s commission $C$ is paid – during the investment period of one year and the final cumulated investment sum $I$. After the first year the investor stops adding new capital to the venture although all the distributed capital gains and dividends that the security generates during the entire period $t$ are reinvested. After $t$ amounts of years the investor sells the shares. At realization – capital gains taxes $\tau_k$ are paid on the difference between the final value and the cost basis, not to forget the brokerages commission. All the variables are assumed to stay the same as in the one period model.

Equation (8) resembles the expression of the Final value of the one period model but is situated in the space time continuum indicated by the yearly cumulative management fees-term for instance which is multiplied with the value of the initial investment less the broker’s fees. The $\prod_{t=1}^{t}[1 + k_t(1 - \alpha \tau_k) + d_t(1 - \tau_d)]$ represent the compounded interest effect that continuous reinvestment of capital gain- and dividend-distributions leads to. Expression (8) is in a nutshell the value of the investment after $t$ amounts of years including taxable proceeds from year $t$. (Expression 8)
To evaluate how much of a tax burden the investor will have to cope with at time period \( t \) I need first to calculate how much of value \( t \) – expression (8) – that is exempt from capital gains tax. This is because some of the money has already been taxed and some of it belongs to the initial invested sum. Note that constant values have been chosen for all \( k_t \) and \( d_t \) – changes are only due to regional factors – not \( t \). The undistributed capital gains are demonstrated by expression (9). The cost basis is the part of the final value that is not taxable since it was either already taxed or part of the initial investment. (Expression 9)

\[
\text{Cost basis}_t = \{(1 - e)\text{Value}_0\} + \sum_{t=1}^{t} \{\text{Value}_{t-1}\{\alpha k_t (1 - \tau_k) + d_t (1 - \tau_d)\}\}\]

The final value of the investment at redemption, after the final capital gains taxes at liquidation and the last brokerage commission are paid is expressed by equation (10). Here capital gains taxes are calculated on the latest untaxed capital gains-distributions.

\[
\text{Final value}_n = \text{Value}_n - \tau_k (\text{Value}_n - \text{Cost Basis}_n) - C
\]

**Kostovetsky’s Sample Selection and Non-Tracking Error Assumption**

If the research concerning the complications of indexed investing is limited as Frino and Gallagher (2001) states: “Despite the significant attention to active funds in the performance evaluation literature, empirical research evaluating index funds is surprisingly scarce” then academic studies evaluating ETFs is even more so. Agapova (2011) states that “few studies are available on ETFs due to limited data given their short period of existence” in her paper and while some study has been conducted of the ability of ETFs to track indices, no other study that I am aware of is building off of Dellva’s and Kostovetsky’s cost-centric approach to comparing the instruments. The surrounding costs that are affiliated with transacting these instruments didn’t really come into light until Kostovetsky presented his cost comparison model.

Kostovetsky did his research based on the largest exchange traded funds and index mutual funds trading on the largest financial market in the world. While doing so might increase the relevancy and reproducibility of the study it also raises the question of whether the same results would be attained on a “lesser” market. The large capitalization and highly liquid SPDR- and QQQ-ETFs would trade at what Kostovetsky considered “negligible bid-ask spreads compared to commissions charged by brokers”, or roughly 2%. The spread is however recognized as a component contributing to the overall transaction cost for ETFs and the author also acknowledge that smaller ETFs are much less liquid and would therefore trade with a greater bid-ask spread.

Below I explain what the bid-ask spread is, what causes it and why there is reason to believe that had Kostovetsky’s research been conducted on a different market or under different circumstances, the impact of the bid-ask spread could possibly be more than negligible. The concept of liquidity and how it relates to the spread is also touched upon.

**MODIFYING THE MULTI-PERIOD MODEL**
Just as in the one period model – the spread will be situated in different places of the value-expression depending on the type of security. In the case of ETFs the spread will be subtracted before management fees are paid, since the spread occurs instantaneously at the time of acquisition and because management fees are an annual percentage.

$$Value^{ETF}_{t} = (1-e)^{(1-CN)} \times \prod_{t=1}^{T} \left[ 1 + k_{t}(1 - \alpha \tau_{k}) + d_{t}(1 - \tau_{d}) \right]$$

The bid-ask spread effects index funds differently as has been have shown above. Just as the management fees the bid ask spread can be regarded as an annual fee that effects the performance of the fund over time – irrespective of N.

$$Value^{Indexfund}_{t} = (1-e)^{(1-CN)} \times \prod_{t=1}^{T} \left[ 1 + k_{t}(1 - \alpha \tau_{k}) + d_{t}(1 - \tau_{d}) \right]$$

AQUIRING THE DATA

The Sample

Due to the intention of researching the liquidity on the U.S. And Swedish markets, the largest funds were selected, which subsequently ought to be the most liquid ones. In Sweden the largest ETF can be found in XACT OMXS30 which tracks the 30 largest companies on the Stockholm stock exchange – in terms of trading volume. There are plenty of index funds that track OMXS30 but the one with the lowest management fee and one of the largest NAV were chosen – namely SPP Aktieindexfond Sverige.

The chosen American equity indices is Standard and Poor’s 500 tracking the 500 largest securities on the US markets, largely from the NYSE, with SPRD S&P 500 being the largest ETF tracking it and Vanguard 500 being the largest index fund with the same underlying benchmark index.

To get a sample of daily data points that is large enough to produce a fairly sized number of monthly observations when aggregated but still being manageable a time period of 5 years has been chosen, ranging from March 16, 2006 to March 15, 2011.

Data on equity indices

The dataset consists of monthly bid-ask spreads, market values and historical prices for all equities on the two sample indices, OMXS30 and S&P500. DataStream also provided data of return index for both OMXS30 and S&P500; the return index assumes all dividends and distributions are reinvested in the respective index. The data has been downloaded from Thomson Reuters DataStream and contains all equities registered on OMXS30 and S&P500 during the sample period.

Data on equity index funds

The monthly NAV was retrieved, historical prices and shares issued of Swedish equity index funds from the Swedish Financial Supervisory Authority (http://www.fi.se). The funds report their data to the Swedish Financial Supervisory Authority quarterly and the average issues was calculated by
dividing total fund value by the price and then calculate the monthly NAV by multiplying the number of issues with that month's average price in order to create monthly time-series.

**Data on ETFs**

Thomson Reuters DataStream provided all monthly historical prices and NAV for the ETFs in the sample. Their management fee and other information were gathered from the Swedish and American pages of Morningstar (morningstar.com, morningstar.se).

**Information about Market Architecture**

Material concerning the structural characteristics of the Swedish and American trading systems will be gathered from their official websites which provides ample information about procedures and protocols.
RESULTS AND ANALYSIS

DIFFERENCES IN THE NYSE AND OMXS TRADING SYSTEMS

DEALER PRESENCE

NYSE

On the NYSE specialists or market makers stabilize stock prices by buying when the public is selling and selling when the public is buying. As a broker, the specialist executes limit orders left on the book when they represent the best price. The dual role of the specialists as brokers for the orders left with them to be executed and as dealers trading for their own account is subject to frequent criticism. Their insight and access to the book and the trading give them an advantage that might be used to their own benefit at the expense of the orders they execute as brokers for the less informed traders. However, the new electronic systems are giving the investors the ability to customize their limit and market orders to be executed at a pre-determined price or cancelling the order if it is outstanding after a certain time. As execution is customized and automated, Stoll (2005) argues that customers benefit since the specialist’s option to delay execution of orders and use the information from the order book to their own benefit is eliminated. One has to purchase a license to trade directly on the NYSE. However NYSE does not only rely on the market makers but has supplemental liquidity providers that compete with other quote providers to ensure that the market stays liquid.

OMXS

Scandinavian stock markets are order-driven, i.e. limit-order markets, in contrast to U.S. stock markets. An order-driven market relies only on public limit-orders to bring liquidity and is a significant market structure difference from U.S. stock markets which have designated market makers. Due to the lack of market makers, the liquidity on order-driven markets is relatively volatile and may even be scarce or even non-existent for short periods according to Söderberg (2009). However Anand et al (2009) tells us that since 2003 firms on the OMXS are able to contract liquidity providers to continually quote prices within a contractually determined spread.

All through the continuous trading session trading are made On Exchange which basically mean that trades are matched in the order book in accordance with the NASDAQ OMX NORDIC Member Rules or alternatively executed outside the order book and reported to the exchange as a manual trade. Orders are suitable for either full or partial execution in one or several steps. Exchange traded funds are just like regular shares traded throughout the whole session.

AUTOMATION

NYSE

On April 20, 2005, NYSE announced it was going to become a public owned company and to merge with Archipelago, one of the most successful ECNs. The merger was a sign of the general increase of electronic trading in stock markets. NYSE is considered a hybrid market, integrating the automated ECNs and the NYSE specialists being a part of the traditional floor trading. The specialists are part of the order-driven auction market and facilitate the order-processing states Stoll (2005).

OMXS
Equities and equivalent instrument that are traded on NASDAQ OMX Stockholm takes place on the electronic trading system INET Nordic - a relative new system which apart from normal trading also enables algorithm trading\textsuperscript{9} (www.nasdaqomx.com(A)). The INET Nordic trading system provides continuous and simultaneous transmission of data to all authorized customers in real time. This includes data displayed in the form of order books, market summaries, finished trades and index information. Buy and sell orders issued by members of the exchange or by their clients go through the Direct Market Access\textsuperscript{10} (DMA), are cataloged in the order book and then matched with a mirrored order. The automated matchmaking process prioritizes orders according to a set list in which price has outmost precedence.

**Continuous Trading**

**NYSE**

The auction format of the NYSE is a continuous one throughout the day from 09:30 to 16:00

**OMXS**

Continuous trading starts at 09:00 along with the automated matchmaking of the INET system and lasts to 17:25.

**Protocols**

**NYSE**

*Off market trading:* NYSE opens with an auction at the beginning of each trading day, prior to the continuous trading which starts 9:30 a.m. and ends 4:00 p.m. Monday through Friday. NYSE does allow aftermarket trading and it starts at 16:00 and ends at 20:00. The orders that arrive during the out-of-hourstrading have to be executed within a 1% price range around the last traded price (www.euronext.com). All orders entered for the day trading session will be canceled at 16:00 and at 20:00 all limit orders entered after 16:00 are canceled (http://www.nyx.com). While being restricted to large and institutional investors for a long time it is now also open to the public

*Decimalization:* The tick size on the NYSE Euronext\textsuperscript{11} exchanges depends on the value of the stocks. For stocks priced below €10 the tick size is €0.001. For stocks between €10 and €49.99 the tick size is €0.005. Prices from €50 to €99.99 have a tick size of €0.01 and stocks priced of €100 and above have a tick size of €0.05.

*Circuit Breakers:* In response to the market turmoil in the late 80's, the New York Stock Exchange instituted circuit breakers, which is basically a halt in trading activity, to reduce volatility and promote investor confidence. Depending on how big of a market decline that occurs during the opening hours of the exchange the NYSE management can implement up to three stops in trading with the last stop being in effect for the remainder of the day (https://usequities.nyx.com).

**OMXS**

\textsuperscript{9} Algorithm trading refers to placements, changes or cancellations of orders in the order book by automatically generated orders in response to specific pre-programmed factors in software.

\textsuperscript{10} Direct Market Access enables a member of the NASDAQ OMX Group to electronically and automatically forward clients' orders to the trading system, INET.

\textsuperscript{11} Multinational financial corporation that operates multiple security exchanges, NYSE being one. NYSE Euronext is now a part of IntercontinentalExchange.
Off market trading: Trading on the NASDAQ OMX Stockholm exchange are divided into three sessions, each with distinctive specifics of how trading is conducted. During the pre-trade session which last from 08:00 to 09:00 CET a call auction takes place to facilitate price formation by managing and matching orders with crossing prices in a procedure called the uncross – the highest bid price must be equal to or higher than the lowest ask price for there to be an uncross. DAY, GTT, GTC, IOC, On-open and On-close orders are eligible for entry (but are not matched until 09:00 CET) during the pre-trade but the focus of order management is to reduce volume. Any manual trades conducted during this session must be reported before the execution of the uncross.

At the end of the continuous trading session – 17:25 CET - the prep for the post-trade session kicks in during which the closing call uncross takes place. On-close orders 12: Market-On-Close orders, Limit-On-Close orders and Imbalance-On-Close order (IOOC) 13 orders can be put in until the closing call uncross. The post-trading session begins at 17:30 and ends at 18:00 CET. During this session no automatic matchmaking takes place and allowed actions are limited to – Order cancellation, off hour’s transactions and limited order updates. Manual trades are entered outside of the order book between members of the NASDAQ OMX and clients. The post-trading session makes public the same financial data as during the continuous trading session.

Decimalization: The tick sizes for instruments listed on the Stockholm market within NASDAQ OMX Nordic depends on the price of the underlying asset. For equity ranging at 0.00 – 4.99 SEK in price the tick size is 0.01 SEK. It is 0.50 SEK for the 150.00 – 499.50 SEK-range and 5.00 SEK for 5,000.00 SEK assets and above.

Circuit Breakers: Trade on the Stockholm bourse can be halted, limited, delayed or postponed in case of “circumstances or events which, in the opinion of the NASDAQ OMX Nordic, have a material influence on the NASDAQ OMX Nordic’s ability to maintain well-functioning trading operations”. Other measures NASDAQ has in their toolbox in the face of such an event are – the limitation of members’ electronic connections to the exchange and the termination of suchlike connections - restraints on the volume of trades per electronic connection or specific ones. Irregular price movements or suspicion of unequal information in the market can give cause of a trading halt (www.nasdaqomx.com(A)).

**POST AND PRE-TRADE INFORMATION**

NYSE

NYSE is a fairly transparent market. Pre-trade information consists of all market orders and the bid-ask spread, number of orders and total order quantity of limit orders. During the auction trade hours the exchange continuously publishes the theoretical opening price and the potentially executable volume at that price. The post-trade information is published immediately and consists of quantity, price and time of execution of each transaction (www.euronext.com).

However in 2007, a new electronic trading system called MatchPoint came into effect. MatchPoint is a market where traders can make transactions anonymously. The system is primarily set up for institutional investors and other large block traders who want to trade anonymously to prevent

---

12 Orders specifying that the trade is to be executed at the close of the market or as close to it as possible.
13 Imbalance-On-Close order – Imbalance orders are used in the auction and accepts the equilibrium price reached by the price determination criteria specified in NASDAQ OMX Nordic Market model.
price impact and information leakage. These trades are referred to as "dark" pools or "dark" liquidity, because the electronic trading information is not publicly displayed (www.nyse.com).

OMXS

The information flow from the exchange varies throughout the trading day. Net order imbalance information is not publicly available during the pre-trade session forth to 15 minutes before opening call while trade information containing volume, price, member ID and trade type is public (www.nasdaqomx.com (B)). During the opening call order dissemination of the order book becomes available (www.nasdaqomx.com (C)). This information consists of market-by-level, equilibrium price and the five best levels on each side of the equilibrium price. During the opening call no information on past trades is given but individual trades executed will be published right after a cross (www.nasdaqomx.com (B)).

The continuous trading session allows for reporting in real-time of trades carried out. Counterparty information of trades is also reported. The composition of the trade information is much like the one presented during the pre-trade stage. Order dissemination continues to being offered but with a slight alteration to composition. Instead of market-by-level as in the previous session, market-by-order is now publicized. When the closing call takes place the same type of data which is publically available during the opening call is provided (www.nasdaqomx.com (B)).

Block trades are large orders relative to the average daily turnover and are allowed to be reported outside of the weighted average spread in the market which is the range between the reference prices which would have been paid if the trade had been executed in the order book (www.nasdaqomx.com (B)).

**Expected Effects from Market Microstructure Differences**

In this comparison of the market architectures of the two stock exchanges the similarities are obvious. Three very important attributes that prevail on both markets are the small tick-sizes, continuous trading and the fully automated efficient electronic systems these attributes have been shown to increase the market liquidity. These three factors ensure that trades are executed without any delays and at small transaction costs. As a consequence it could be expected that both markets will exhibit relatively high liquidity. NYSE is however considered to be a hybrid market committed to preserving the relic of the floor trading system. This could have an adverse effect on market transparency, increase information asymmetry and transaction costs.

A major difference between these two markets is that the NYSE is a dealer-driven market while OMXS is an order-driven market. OMXScould therefore be constrained by periods of illiquidity due to lack of satisfactory amount of market orders. This problem is however met with a solution – the largest enlisted OMXS companies’ hires liquidity providers that guarantee a fixed negotiated amount of available orders of the stock with a certain tick-size. As previously described, market makers will set place bid and ask orders at a narrower interval than would prevail in a market with orders solely from traders, effectively narrowing the bid ask spread. Since OMXS only house liquidity providing market makers for a few stocks it stands to reason that the spread on OMX is wider in general.

Both OMXS and NYSE allow for afterhours trading which drives the markets towards less transparency, but equally so. While it was restricted to a select few for a long time it is now open for
anyone using a broker that provides that service. This may perhaps dampen the effect that the off market trading activity has on the prevalence of asymmetric information.

Both OMXS and NYSE provide fairly equal amounts of both post and pre trade information about quantities, quote data and execution times and prices. However both exchanges have implemented systems for large trades or traders to execute orders anonymously or outside the book. OMXS does this with its special procedures for block trades and NYSE with its MatchPoint system. This increases the information asymmetry and increases transaction costs.

After comparing the trading architectures on the Stockholm and New York exchanges it is apparent that they are similar in many ways. However, two important differences exist. The absence of dedicated market makers on the Stockholm stock exchange leads us to the reasonable expectation that the OMXS have a less favorable structure for liquidity. On the other hand, if we take into account the old-fashioned floor trading arrangement that exists alongside the more efficient electronic system on the NYSE it is no longer obvious that the microstructure on the OMXS is less favorable. How the net change in transaction costs looks like is not clear since the two differences work against opposite directions. Even though the NYSE is the largest exchange market and should attract more liquidity by this fact alone and subsequently has smaller spread, this aspect is not to be considered when analyzing the market microstructure. The conclusion is that the market microstructure differences in the two systems leaves open the possibility of systematic difference in the general spread.

**Disparity in Liquidity: The Roll’s Estimate**

After calculating the percentage price change during the time period of 2006-04-03 to 2011-04-01 of the two indices the covariance could be extracted. In both cases there were negative covariances (\(-1,20113E-05\) for OMXS30 and \(-3,05981E-05\) for S&P500) indicating systematic price reversal behavior in the price series and by proxy the presence of transaction costs. The Roll’s covariance measure for OMXS30 came out to be 0,0069 and 0,0110 for S&P500. Since both estimates are based on normalized numbers they can be compared with one another. S&P500 displays a close to 60% greater indication of transaction cost presence suggesting that market friction prevail despite it being based on the larger market.

While it is true that S&P does not solely include stocks from NYSE but also include stock from the NASDAQ, NYSE is still the largest stock market in the world with a good margin and dominates the S&P500 index. While there was room for difference in the spread due to market structural factors between the markets this result still comes as a surprise considering the sheer size of the NYSE. One might have had assumed that the trading activity on such a market would put enough downward pressure on the spread to nullify any market structure differences.

**Quantifying the Spread**

To determine whether the spread is a significant variable for the tracking error observed in the selected funds, all the different costs- factors are fitted to an OLS regression on the dependent variable, tracking error estimated as the RAW return difference between the fund and the index. The regression result yielded coefficients for each variable, which enables quantification of the
factor’s impact on tracking error but it also yields test statistics so that the individual variables can be checked for significance. The regressors made up of cash flow CF, the bid-ask spread SPR, the index-volatility VOL1 calculated as the monthly standard deviation and an alternative volatility measure VOL2 delineated by Parkinson and the monthly dividend yield DIV measured as the average monthly difference in returns between the going price index and the corresponding return index.

This is a four factor model with one alternate measure for volatility for two sets of market data. These sets of variables should have the explanatory power to correctly account for the variance and if not the Parkinsons refined volatility measure, used in a parallel pair of regressions with the VOL2 measure, might. Table 2 contains the result from the OLS regression of the monthly average return disparity between SPP Aktiefond Sverige and the benchmark index OMX 30 on the constituent sources of tracking error. Table 3 displays the result based on the US market where the Vanguard 500 fund is regressed on the S&P500-index. Values for the alternate model with Parkinson’s volatility measure are presented in parentheses.
### TABLE 2

**Index Fund: SPP Aktieindexfond Sverige, Benchmark: OMXS30**

<table>
<thead>
<tr>
<th></th>
<th>CF</th>
<th>SPR</th>
<th>VOL1</th>
<th>VOL2</th>
<th>DIV</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coefficient</strong></td>
<td>.5631685</td>
<td>1.398708</td>
<td>-</td>
<td>-</td>
<td>.0616641</td>
<td>.0086673 (-)</td>
</tr>
<tr>
<td></td>
<td>(.5464432)</td>
<td>(1.263871)</td>
<td>.0769074</td>
<td>.0000462</td>
<td>(.0873608)</td>
<td>(.0013997)</td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>7.38 (7.15)</td>
<td>0.42 (0.38)</td>
<td>-1.56</td>
<td>-0.95</td>
<td>0.34 (0.46)</td>
<td>0.66 (-0.13)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>0.000 (0.000)</td>
<td>0.674 (0.707)</td>
<td>0.124</td>
<td>0.347</td>
<td>0.735 (0.646)</td>
<td>0.511 (0.893)</td>
</tr>
<tr>
<td><strong>R² adjusted</strong></td>
<td>0.4629 (0.4480)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F-statistics</strong></td>
<td>13.71 (12.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>p-value (F-stat)</strong></td>
<td>0.0000 (0.0000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The p-value for the f statistic (0.0000) indicates that, at a 99% level of confidence one or several of the explanatory variables are worth keeping in the model. One or more of the coefficients are significantly different at this confidence level – confirming that the overall model is significant. The Swedish adjusted R² – R² decreased from 0.4629 to 0.4480 along with the critical F statistic –
13.71 to 12.97, although the F-test still held at the 0.005 confidence level. The model using the Parkinson measure for volatility raised the adjusted goodness of fit R² measure slightly in the US model – 0.8211 with the standard deviation measure and 0.8236 with the Parkinson estimator – and also the critical F value – from 68.68 to 69.89.

The result does not support the hypothesis that the spread is a significant source of tracking error. The p-values for the SPR variable is 0.674 – with a 1.398708 coefficient – on the Swedish market and 0.31 on the U.S. – with a 4.651393 coefficient – which leaves SPR insignificant at a 90% confidence level. These results are contradictory to the findings of Frino & Gallagher (2002) who found that the spread was a significance part of tracking error for index mutual funds. To examine whether this was due to the volatility measure that was being used I employed the alternative variable (VOL2) to see if it made a difference. The alternate Parkinson volatility estimator VOL2 did not improve the significance of the SPR variable either – the p-value for SPR in the Swedish model went up to 0.707 – with a 1.263871 coefficient – while it increased to 0.519 in the US – with a 3.114118 coefficient. The generally higher spread coefficients that are observed on the Swedish market might be due to the market microstructure characteristics of the Swedish stock exchange that might increase the average transaction costs.

As seen from Table 2 the cash flow CF is the only consistently significant variable which can be seen as the driving component of tracking error for index funds. However – the cash-flow component is not relevant for the thesis of tracking error due to the bid-ask spread. DIV is significant at a 99% confidence level in the US model with the alternate volatility measure.

Whites test on the Swedish market showed a total p-value of 0.38. A small p-value indicates that the null hypothesis can be rejected. But in this case the null hypothesis cannot be rejected.

This would indicate a presence of heteroscedasticity so to confirm the suspicions the residuals versus the fitted values were plotted against each other. The graph contradicted the white’s statistic and as a result the tests for heteroscedasticity were inconclusive but by inference from two contradictory tests suggesting that the residuals should not have severe variance, if any. The p-value of White’s test on the U.S. market was 0.0011 and the null hypothesis could not be rejected at a 99% confidence level. To correct for the probable heteroscedasticity a robust regression was employed which makes the new results aligning with the assumptions of an ordinary least squares regression.

### TABLE 4

<table>
<thead>
<tr>
<th>Robust regression: SPP Aktieindexfond Sverige</th>
<th>CF</th>
<th>SPR</th>
<th>VOL</th>
<th>DIV</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>.7405177</td>
<td>4.651393</td>
<td>.7934104</td>
<td>-3.835369</td>
<td>.0684132</td>
</tr>
<tr>
<td>t-stat</td>
<td>15.38</td>
<td>1.02</td>
<td>1.27</td>
<td>-2.50</td>
<td>2.67</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.311</td>
<td>0.211</td>
<td>0.015</td>
<td>0.010</td>
</tr>
<tr>
<td>R² adjusted</td>
<td>0.8211</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistics</td>
<td>68.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value (F-stat)</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 5

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>CF</th>
<th>SPR</th>
<th>VOL</th>
<th>DIV</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-stat</td>
<td>7.11</td>
<td>0.66</td>
<td>-1.61</td>
<td>0.91</td>
<td>0.72</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000</td>
<td>0.513</td>
<td>0.114</td>
<td>0.369</td>
<td>0.476</td>
</tr>
<tr>
<td>R² adjusted</td>
<td>0.4697</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistics</td>
<td>14.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value (F-stat)</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results conclude that the bid-ask spread can be disregarded in the cost-analysis model because it is not a significant component of tracking error. The spread component is statistically not differentiable from a 0 value at a 95% confidence level. A significant spread variable would have been included in the cost analysis model and could have made an impact on the index fund costs. Omitting the spread cost is only relevant for index funds since ETFs have the spread cost on the investor level.

THE COST COMPARISON

SINGLE PERIOD COMPARISON

In this section the results from the cost-analysis models will be presented. In the one year model a sum has been invested in the chosen index fund (FVi) and ETF (FVe) on the Swedish market. The corresponding costs of each fund have been extracted from the initial investment during one year and have then appreciated with the same growth rate. Throughout this section both the original cost comparison and the altered one which factors in the spread component will be graphed. FVi-FVe* denotes the modified cost comparison.

In Figure 4 the value of Equation (7) for the Swedish market with its variables held constant is graphed as the initial investment I changes. Positive values of FVi-FVe indicate that the index mutual fund has a higher Final Value than its ETF counterpart on an annual basis. It is evident that regardless the size of the initial investment Swedish ETFs will never be more cost beneficial than Swedish index mutual funds as Figure 4 clearly shows. In fact as the initial investment increases the excess value of index funds does too. An apparent shift in the relative profitability of the two securities takes place as I exclude the bid-ask spread cost. Although ETFs gains a boost to its relative profitability the chart never crosses into negative territory – indicating the same basic
relationship as the previous result.

Figure 5 graphs the nominal difference of the two measurements of relative profitability. It is obvious to the reader that the nominal spread cost $S_I$ is positively related to the size of the investment $I$. As more is invested the greater is the impact of market friction. For example – a person investing $70,000 commits to a spread cost of $288.
To better appreciate the nature of the value disparity that prevails these instruments the value difference are plotted as a percentage of the initial investment in Figure 6. Note that the discrepancy is nonlinear and negatively related to the invested sum, $I$. The greatest impact of the difference strikes at the lower sums which decrease as $I$ increases and stagnates just above 0.5%. The same pattern can be seen in the curve graphing the value dissimilarity excluding the spread cost – although at a lower percentage level of the investment.

The difference between these two curves are plotted by the \((FV_{i}-FV_{e})-(FV_{i}^{*}-FV_{e}^{*})\)-line. This line indicates that the spread components impact on the relative value difference is constant with the invested sum $I$. 

**FIGURE 5**
The variable table number 6 presents the effect on the final value as the variables are being tampered with.

Swedish Index funds do not distribute capital gains because of the tax-relieve effect – thus the capital gains ratio – $\alpha$ is adjusted on the index fund side only while holding it at zero for ETFs. As a result the $\alpha$-row is not as interesting for this intent. A higher management fee for index fund could have a noteworthy effect on the final value – as seen in the variable table. The results from variable C and N in the table are expected and show that increasing the commission fee and numbers of transactions have an adverse effect on the relative additional value of ETFs. Since capital gains distributions are unbecomingly in Sweden the tax rate on capital gains will as expected not have an effect on my results. The overall results from the changing variables indicate an insensitive Swedish fund market. This can be due to the fact that the tax-structure and costs behind index funds and ETFs are very similar.
Cost-analysis between XACT OMXS30 and SPP Aktieindexfond Sverige

**TABLE 6**

**Variable table:** $FV_i - FV_e$ = results presented in parentesis ()

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$\alpha_i = 0$</th>
<th>$\alpha_i = 0.1$</th>
<th>$\alpha_i = 0.2$</th>
<th>$\alpha_i = 0.3$</th>
<th>$\alpha_i = 0.4$</th>
<th>$\alpha_i = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I =$ 70.000</td>
<td>$(386)$</td>
<td>$(361)$</td>
<td>$(335)$</td>
<td>$(309)$</td>
<td>$(283)$</td>
<td>$(257)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$C_e =$</th>
<th>$C_e =$</th>
<th>$C_e =$</th>
<th>$C_e =$</th>
<th>$C_e =$</th>
<th>$C_e =$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I =$ 70.000</td>
<td>$(368)$</td>
<td>$(386)$</td>
<td>$(494)$</td>
<td>$(620)$</td>
<td>$(810)$</td>
<td>$(1630)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$e_i =$</th>
<th>$e_i =$</th>
<th>$e_i =$</th>
<th>$e_i =$</th>
<th>$e_i =$</th>
<th>$e_i =$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I =$ 70.000</td>
<td>$(497)$</td>
<td>$(386)$</td>
<td>$(349)$</td>
<td>$(239)$</td>
<td>$(91)$</td>
<td>$(204)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$N =$</th>
<th>$N =$</th>
<th>$N =$</th>
<th>$N =$</th>
<th>$N =$</th>
<th>$N =$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I =$ 70.000</td>
<td>$(370)$</td>
<td>$(371)$</td>
<td>$(374)$</td>
<td>$(377)$</td>
<td>$(386)$</td>
<td>$(405)$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>$\tau_k =$</th>
<th>$\tau_k =$</th>
<th>$\tau_k =$</th>
<th>$\tau_k =$</th>
<th>$\tau_k =$</th>
<th>$\tau_k =$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I =$ 70.000</td>
<td>$(386)$</td>
<td>$(386)$</td>
<td>$(386)$</td>
<td>$(386)$</td>
<td>$(386)$</td>
<td>$(386)$</td>
<td>$(386)$</td>
</tr>
</tbody>
</table>
When I graph the value gap of US index fund Vanguard 500 and the SPDR 500 ETF an entirely new relationship is exposed – where the index fund does not reign supreme but lose its competitiveness as the investment amount I increase. Contrary to the Swedish case – the line does cross into negative territory giving rise to a comparative quantitative advantage of ETF investments as more money is devoted to the venture. Explicitly – if one where to invest $33.337 or more in an index tracking vehicle then ETFs are preferable over index mutual funds. If the spread is excluded ETFs becomes a viable alternative at $23.667 – a whole $9.670 decrease from when the spread is included. At $70.000 investment the spread component reaches $27 (61-34).

However – the intercepts on the $FVi-FVe*$-axis are highly dependable on the constant which are chosen for the model – making these figures susceptible to major fluctuations. Note this does not imply that the cost-impact of the spread is in the thousands – only that the slopes of the lines are so moderate that covering a spread cost of mere $10 have a tremendous leverage on the Investment-axis.
The percentage cost difference figure in Figure 8 has the same general characteristics as the Swedish conditions. The curves are downward sloping and decaying at an exponential rate, settling at investment sums beyond $65,000. Although a few minor distinctions can be observed – namely that the curves enters negative space whereas Swedish ETFs never outperforms index mutual funds. Albeit \((FV_i-FV_e) - (FV_i-FV_e^*)\) is constant it resides at a much lower level – 0.039 % – than the Swedish counterpart leaving us with the conclusion that the spread component is a much lesser of a factor in the U.S. Cost-analysis between SPDR and Vanguard 500 with the spread excluded.
The variable table shows the results of holding a certain level of invested sum while the other variables are adjusted one at a time. Due to the existing capital gains distribution taxation praxis in the US a change in the capital gains distribution-ratio $\alpha$ has a slightly smaller effect than an equal sized increase in Sweden. Furthermore – this praxis has a more evident manifestation in a change of the tax rate of capital gains $\tau_k$ where it does not have an effect at all in the Swedish model. At 0,5% expense ratio for the index mutual fund it is increasingly profitable to invest in the classic index tracking vehicle but if the spread cost is excluded ETFs yield a higher end value as $l$ increases.
MULTI-PERIOD INVESTMENT

The multiperiod model for Sweden shows a scenario where index funds consistently outperform ETFs. Figure 9 graphs the relationship of different initial investment over a 20 year time period – with all the other variables having the same values as the one period model. This holds true throughout the full range of investment sums and does not change over time. The higher the initial investment and the longer the investor's time horizon is – the more cost efficient are Swedish index fund compared with ETFs – money makes money if you will. This relationship cannot be explained by the existing tax laws due to their similar tax structures. Taxes on capital gains and dividends are equal and have the same effect on these investments – in contrast to the U.S. where these taxes depends on income and the investments time horizon. The explanation can instead be found in higher management fee and commission costs for ETFs where the annual compounded effect carries an exponential relationship. Eventually, an investment of $20,000 in an index fund and an investment period of 20 years will result in a $1,000 difference in final value compared with an ETF. The results for the Swedish models indicate that Swedish ETFs are traded with a premium that has to be weighed against any qualitative features that the investor might seek.
The multiperiod model for U.S. funds is shown in Figure 10. As seen in the graph above American ETFs and index funds turns the Swedish relationship – where index funds outperformed ETFs – on its head. Instead American ETFs will outperform their counterparts if the investor’s initial investment is large enough and their holding period is long enough. Similar to the one period model – investors on the US market should prefer ETFs over index funds when considering larger investments.

Similar to Kostovetky’s results – I did also find a relationship between index mutual funds and ETFs that favors the former as the initial investment sum increases. This relationship was however only found in the US market while it was absent for the Swedish instruments. The reason for this diversion is due to the fact that ETFs have a larger management fees than index funds in Sweden as well as the tax benefits seen on the US market is non-existing. Because of these facts investors seeking Swedish ETFs will have to value the qualitative benefits of ETFs more than their US counterparts. The qualitative value of being able to buy or sell ETF shares during the trading hours – short-selling and using ETFs to hedge a portfolio.
CONCLUSIONS

In this paper I’ve set out to answer the question of how market dimensional differences between the OMXS and the NYSE affect implicit transaction costs and how these in turn alter the competitive relationship between the two popular and in many ways similar, financial instruments – Exchange Traded Funds and conventional Index funds.

By examining the rules and protocols that prevails on each market it became apparent that the market microstructure designs on OMXS and the NYSE are very similar in many ways but do differ in respect to the lack of dedicated market makers on the Stockholm exchange and the hybrid system of floor trading arrangement and electronic system on the NYSE. These differences work against each other in influencing the size of the transaction costs and since their effects have as of yet only been tested to be qualitative in nature, this market structural study would not let me prognosticate any quantifiable differences. What such an exercise does reveal though, is that the market microstructural differences between the OMXS and the NYSE does leave the door open of market architecture explaining systematic differences in transaction cost. It’s not out of the picture.

The other two market characteristics that affect the spread component are the closely related liquidity and volatility. Both of which I cover with the Roll’s estimate of market friction. The Roll’s estimate revealed 60% stronger prevalence of price reversals in the S&P500 indicating that market friction is greater on the US market.

The market structural study indicates that significant disparity in transaction cost might be actualized from the information asymmetry component in the spread and the Roll’s estimate shows that the utilitarian trading and volatility component of the spread are greater on the NYSE. What we now know is that a crucial component in comparing the profitability of the two financial instruments should be expected to differ on several levels of its constituting parts.

The nominal spread difference between the where in the order of a spread cost of $288 for $70,000 investment on the Swedish market and a spread cost at around $30 on the US market. The cost of the bid ask spread when purchasing the Swedish ETF is notably higher than when the US counterpart is bought. However, from the regression that was run, I can found that the bid-ask spread turned out to be an insignificant part of the tracking error of the chosen index funds – both on the Swedish and US markets. This came as both a surprise and not, at the same time. Big international financial centers ought to have very low cost of transaction by the merit of their sheer size and activity but at the same time my study gave other expectations. The Roll’s estimate indicated that the spread ought to be greater on the US market and market microstructure theory indicated that the antiquated floor trading system on the NYSE might increase the asymmetric information component of the spread.

Furthermore I found that the Swedish ETF do not get more profitable as the investment sum increases. No matter the initial investment, Swedish ETFs will never be more cost beneficial than Swedish index mutual funds. This result contradicts the findings on the US market where ETF’s becomes more and more preferable as the invested amount increases and is likely a consequence of the Swedish tax-laws for capital gains as well as the higher levels of management fees for ETFs. Because of the tax-relieve effect Swedish Index funds do not distribute capital gains. This causes the tax rate on capital gains in Sweden having minimal effects on the results and leaves the Swedish
fund market insensitive to such alterations. However changes in the capital gains distribution-ratio on the US market have somewhat smaller effects than in Sweden.

I also found out that Swedish index funds convincingly outperformed Swedish ETFs over a 20 year time period, unlike the opposite condition on the US market, and does not change with the invested amount of capital. They have similar tax structures, as in taxes on capital gains and dividends are equal, so this relationship cannot be explained by the existing tax laws. This is not, however, true in the US where these taxes depend on income and the time horizon of the investment. This is instead supported by the annually compounding effect of higher management fee and commission costs for ETFs. Investments in Swedish ETFs should only be preferred over a comparable index fund if the benefits of its qualitative benefits exceed the lower profitability. Whereas an ETF investment on the US market will include the qualitative advantages and indeed outperform a comparable index fund over time.

It’d be a novel idea that ETF investors on the world’s many exchanges broaden their understanding of the competitive nature of this relatively new financial vehicle in the light of this paper. I’ve demonstrated how different local market dimensional factors and fixed costs associated with these instruments create a different cost relationship than previously detected and that it is not always prudent to invest in ETFs when dealing with large amounts of cash.

Further inquiries into this subject should recognize the importance of the data collecting process since it may have played a role in the bid ask spread being insignificant. I needed monthly data of the NAV for each fund but this was not available because I could only get hold on the price of a share of the fund but the total amount of issued shares for that month were lacking. This meant that I had to collect the data for the amount of shares from the fund’s annual and quarterly reports. The amount of shares issued can vary considerably from month to month – however this procedure I had to use led to an annual or quarterly average of the total amount of shares issued. The lack of this particular data could have led to a result where the cash flow variable correlates all too well with the tracking error variable.

My study did not support the notion of the spread playing a significant part in the tracking disparity of Swedish ETF’s. This was the result in spite of the examination of the market dimensional factors demonstrating that there are reasons to believe that it would. I’d like to think that there still is merit to this line of inquiry and I’m certain that where this study to be done with other means or maybe in some other manner, there are discoveries to be made.
REFERENCES


Durbin, J. and Watson, G. S. (1951) Testing for Serial Correlation in Least Squares Regression. II. Biometrika, 38, 159-177


Markowitz H. (1952) Portfolio Selection, The journal of Finance, 1, 77-91


Seha M. Tinic (1972) The economics of liquidity services. The quarterly journal of economics, 86, 79-93


http://www.ici.org/etf_resources/research/etfs_02_14 (2014-03-20) (A)

http://www.ici.org/research/stats/trends/trends_02_14 (2014-03-20) (B)


http://www.investopedia.com/terms/i/index-investing.asp (2014-05-01) (C)


