LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Bachelor's thesis **Lappeenranta, April 7th 2008**

Equity index funds and ETFs: Performance comparison between passive investing alternatives

Joonas Turunen

Table of contents

1 Introduction 2
2 Previous Research 3
2.1 Tracking error costs 3
2.1.1 Portfolio adjustments 4
2.1.2 Cash drag 6
2.1.3 Dividend policy 6
2.2 Non-tracking error costs 7
2.2.1 Management fees 7
2.2.2 Transaction costs for investors 7
2.2.3 Taxation 8
3 Study outline and Hypothesis 9
3.1 Tracking error measurements 9
3.2 Sharpe ratio 12
3.3 Jobson & Korkie test statistic 12
3.4 Jensen's alpha 14
4 Empirical Data 15
4.1 Fund samples and descriptive statistics 15
4.2 Monthly total returns 17
4.3 Risk-free rates 19
4.4 Tracking errors 20
4.5 Risk adjusted returns 23
5 Results 28
6 References 30

1 INTRODUCTION

It has been well established, that active mutual fund management is rarely able to outperform or even match the performance of the underlying benchmark. This has been credited to the higher expense ratios of actively managed funds. Stock picking through data analysis seldom is enough to create sufficient alpha i.e. to outperform the benchmark index to offset the higher management and trading costs. Harper et al. (2003) investigated the differences in returns between passively managed exchange-traded funds (ETFs) and actively managed closed-end funds (CEFs). They found that due to the different structures of the ETFs and CEFs, the ETFs exhibited returns that, in vast majority of examined countries, outperformed those of the CEFs. Statistically significant higher risk-adjusted returns of the ETFs were found to be the result of lower expense ratios. Against this backround, it is not surprising that during the last decade the number of passive investing instruments such as ETFs and open-end index funds has skyrocketed. According to a recent report by ICI Statistics & Research,[1] the amount of capital invested in ETFs around the world was nearly $560 billions in January 2008, which is over a 25% increase compared to last year. The supply of ETFs and index funds has risen with demand and expanded outside the traditional benchmarks such as equity indices. There is not, however, a consensus on which of these two fund classes ultimately outperforms the other or if, in fact, they are equally good at matching the performance of the underlying benchmark. Studies conducted on this topic have mostly concentrated on comparing a single ETF against a single index fund in one or multiple countries. As other studies have found significant differences in performance between seemingly identical ETFs or index funds, there is an abundance of results that conflict with each other. The most common explanation given when an ETF outperforms an index fund is the tax-advantage an ETF has due to its unique structure.

In this study we will quantify the performance differences between the most popular passive investing alternatives that have been available for investors looking for an equity index-tracking instrument in the United States and in the Euro zone. Our sample time period is just over three years, beginning from October 2004 and ending

[1] Internet: http://www.ici.org/home/etfs_02_08.html March 31st 2008

in January 2008. In total, eleven open-end index funds and six ETFs that either track the S&P 500 index or the Dow Jones EuroStoxx 50 index are examined for past monthly performance. Our findings indicate, that significant differences in performance can be seen both within and between these two asset classes.

2 PREVIOUS RESEARCH

A lot of research has been done in the field of passive versus active investing. Now that the ETFs have existed long enough to provide sufficient statistical data, the focus of many studies has been directed at comparing different passive investing methods. In this chapter we will summarize the key features of ETFs and index funds and explain the reasons why comparisons between these two asset classes are made. Elton et al. (2002) studied Standard & Poor's Depositary Receipts (SPDRs), the most popular and longest running ETFs, from February 1993 to December 1998 and concluded that cash dividend and management expenses are the two primary interference variables of tracking error. When tracking error includes the dividend reinvestment and deduction of management expenses, the NAV return is lower than the S&P index return by 28 basis points. However, Blume and Edelen (2002), by studying S&P500 index funds, argued that the tracking error was minimal and the index had been tracked quite closely. Harper, Madura and Schnusenberg (2003) also found that iShares ETFs had no significant tracking error during their study period.

2.1 Tracking error costs

Tracking error costs are the costs that are directly decreasing a fund's ability to perfectly mirror a benchmark or, in our study, an equity index. In the literature these costs have been divided into three sub-categories. The effect of these separate cost categories on performance as well as of those costs that do not directly influence a fund's tracking ability have been widely studied. There is, however, a lack of consensus on the absolute, quantitative effects of each of these costs.

2.1.1 Portfolio adjustments

Understanding how portfolio adjustments can have an effect on fund performance requires some knowledge on how traditional mutual funds, index funds and ETFs differ from each other in terms of market participation. An exchange-traded fund works in a manner unlike any other fund type. Cash, used as a transaction medium in an ETF is only a fraction of that used in a traditional mutual fund. When acquiring or selling shares, an ETF uses a method called creation in-kind or redemption in-kind. What this means in practice is that a basket of the shares owned by the ETF is traded for another basket of shares of the same value, provided by a certain number of large, institutional investors. These baskets represent the changes in the portfolio that the ETF needs to make in order adjust its ownership to better mirror the underlying index.

Gastineau (2002) points out that a major source of tracking error for index funds are the transaction costs that occur when the composition of the underlying index changes. In theory, replacing a share with another in an index should not have an effect at all on a fund's performance, but in practice, it can significantly affect it. The transaction costs due to index changes stem from the increased volatility of the shares in question after there has been an announcement of an index rebalance. There has been a lot of research to support the argument that index funds prolonging their rebalancing until the scheduled day of index change are worse off compared to those funds that rebalanced immediately after the announcement was made. Blume and Edelen (2002) reached this conclusion in the case of S&P 500 index funds. They claimed that if an index fund or an ETF had adjusted its portfolio immediately after an announcement for an index change was made, it would have always resulted in a better return than postponing the adjustment until the actual day of the change. Gastineau (2004) quantified these costs to be between 200 and 300 basis points annually in the Russell 2000 index. In addition, he concluded that the reason why S&P 500 ETFs underperformed their index fund competitors between the years 1994 and 2002 was partially due to the lack of proactiveness by the ETF managers during index changes. In simple terms, the predictability of the transactions made by ETFs during index changes has caused them to underperform in comparison to index funds with less predictable market participation pattern. One can imagine a situation, where a share that has been relatively thinly traded in the past, becomes very volatile after it is announced to join an index. Part of this volatility is due to the investors, who speculate that if they buy the share now, they will be able to sell it back for a higher price once the index funds are forced to include it in their portfolios. This is called arbitrage trading and there is evidence that this structural weakness of index funds has been taken advantage of. Zitzewitz (2002) estimated that it was possible for these arbitrageurs to earn excess returns between 40% and 70% in international funds at the expense of other shareholders.

Edelen (1999) relates in-and-out trading to liquidity, showing that the indirect costs of providing liquidity to investors in an asymmetrically informed market can have a significantly negative impact on mutual fund returns. Although this problem is not as important for domestic index funds, it can still be a meaningful influence on an index fund's tracking error. Bid-ask spreads and other liquidity costs are the primary source of tracking error for index fund managers. The advantage for the ETFs in these kind of situations is usually that they can acquire their shares without liquidity costs in the form of bid-ask spreads. Furthermore, ETFs restrict the creation and redemption of shares to large in-kind transactions. These in-kind transactions prevent the price of the ETF from deviating from the net asset value through arbitrage and reduce the observed discount and premium found in closed-end funds. If there were to be a substantial premium or discount, arbitrageurs would step in and create or redeem shares, bringing the market price back to equilibrium. Most small investors, however, are unable to meet the size requirements for creations and redemptions in-kind, and must conduct all transactions in the secondary market. The advantages for the large investors are that they can obtain a large number of ETF shares without influencing their price in the secondary market. For example, when there is a large inflow of funds into an index fund, the manager must invest these funds, incurring costs in the form of bid-ask spreads. Likewise, when there are redemptions that cannot be met with the cash available on hand, fund managers have to sell stocks and once again incur costs. Very often, some constituent stocks of an index are illiquid, forcing managers to suffer high costs to trade in them. The movement of cash in and out of index funds is a secondary cause of tracking error.

2.1.2 Cash drag

The term "cash drag" is used to describe the effect of uninvested assets on the performance of a mutual fund. Every traditional mutual fund has to keep a certain amount of cash available to meet daily redemptions. As this cash is not invested in the shares of the underlying index, it can have a significant effect on an index fund's tracking error. For ETFs this is not a major concern, as almost no cash changes hands through daily redemptions. The miniscule cash drag of ETFs has to do with the situations, when the two baskets of shares changing ownership are not exactly equal in value. The difference in values of those baskets has to be compensated with a cash component and for ETFs it can either cause an inflow or an outflow of cash. An inflow of cash is even more troublesome for index funds, as it is impossible to immediately invest all of the incoming funds. The periods between an inflow of cash and rebalancing of the portfolio cause cash drag. There has, however, been some criticism against this concept. First of all, cash drag is said to affect only a fraction compared to the daily price movements of shares. Secondly, most mutual funds use futures to prevent loss of exposure to the market risk. However, the comparison between index investing vehicles is so intense that any handicap, major or miniscule, can have a statistically significant effect on returns.

2.1.3 Dividend policy

One of the most significant feature where index funds have a built-in advantage over the ETFs is the dividend policy. When a dividend is released from a share, the index fund is able to re-invest it within a couple days' time. This is not the case with ETFs, as they are required to accumulate all cash dividends until the end-of-quarter. Only then it can be distributed to the shareholders. Dividend accumulation is another form of cash drag, which in this case works solely against ETFs. Of course, on average, it takes three business days for index funds to receive their dividend payments but that is nothing compared to the dividend cash drag of ETFs. One could also make the argument that this is not as big a problem today as it would have been in the 1960s and '70s, when the dividend yields were significantly higher.

2.2 Non-tracking error costs

2.2.1 Management fees

Management fees are an inescapable cost of indirect investment in the stock market. For active mutual funds, the total expense ratio, which measures management fees as a percentage of total managed assets, can be as high as 2%, but for index funds, expense ratios are usually below 0.5% per year. Exchange-traded funds have been able to offer even lower expense ratios than the cheapest of index funds. For example, the SPDR ETF featured in our study has an expense ratio of 0.12%. The main reason why ETFs are able to offer lower expense ratios is that they are not in charge of shareholder accounting. The task of keeping track of shareholder transactions and other such paperwork is a large percentage of the total expense ratio. For ETFs, these tasks are performed by the brokerage houses of the shareholders. Gastineau (2001): “Shareholder accounting for ETFs takes place at the investor’s brokerage firm, rather than at the fund. This creates no problems for the shareholder, although it does have some significance for the distribution of exchange-traded funds.” According to him, the elimination of shareholder accounting can save ETFs anywhere from 5 to 35 basis points annually in expense costs.

2.2.2 Transaction costs for investors

The very first reason why investors are willing to invest in an index-tracking instrument is the cost efficiency. For example, if an investor was to construct a portfolio to mirror the performance of S&P 500 by buying all of the underlying shares in appropriate weights, the brokerage commissions would be enormous. Buying an ETF reduces the number of these commissions to one, but index funds go even further and usually do not charge a commission on a cash deposit. This is called a no-load fund and the majority of index funds are no-load funds. Unlike mutual funds, ETFs are traded on the secondary markets, which means investors are also exposed to additional transaction costs in the form of bid-ask spreads. For the popular ETFs this is not a problem as they are considered highly liquid.

2.2.3 Taxation

The last factor that distinguishes ETFs and index funds is their tax efficiency. When redemptions exceed additions, the index fund manager is forced to sell stocks and distribute capital gains to shareholders. These capital gains are immediately taxed and create substantial costs for the shareholders. An ETF, on the other hand, rarely if ever distributes capital gains. Because of the creation and redemption in-kind, ETFs always give away the stocks that have appreciated the most and have the most capital gains taxes to be paid, and keep the stocks with the lowest hidden taxable appreciation. When they need to sell stocks in order to rebalance, they can sell those stocks and not incur capital gains because of their high original purchase prices. When the brokerages opt to take ETF shares off the market, the fund hands over stocks, rather than cash. This allows the ETFs to avoid selling their underlying stocks to accommodate investor traffic. By contrast, conventional mutual-fund managers frequently need to raise cash to pay departing investors. Those stock sales can lead to capital-gains distributions and a tax bill for remaining fund investors. In the long run, taxable investors shouldn't ignore this advantage.

3 STUDY OUTLINE AND HYPOTHESIS

3.1 Tracking error measurements

Our measurement for statistically significant differences in performance begins by first calculating the continuously compounded monthly returns for both indices and all of the ETFs and index funds. These returns will then be formed into time series to enable us to make the statistical tests described later in this chapter. From these time series values we will first extract the accumulated total returns over the whole sample period. This will give us a rough estimation of the tracking abilities of the funds. After this we will calculate the tracking errors of these investment vehicles. At this point, however, a caveat is in order. There are multiple ways of expressing tracking error, each with its own supporters. Mutual funds prefer presenting an annualized in-sample value for tracking error, whereas previous academic research in this area has used an accumulated or an out-of-sample value. For display purposes, we first calculate the accumulated i.e. sum of all monthly tracking errors as well as the average monthly tracking error. The first method used to calculate tracking error is subtracting the benchmark's total return from that of the portfolio, in this case the ETF or an index fund.

$$TEa = \sum r_p - \sum r_b$$

where:

TE_a = Accumulated monthly differences in returns over the measured time period
$\sum r_p$ = Sum of monthly portfolio returns
$\sum r_b$ = Sum of monthly benchmark returns

This will give us an estimate of a fund's absolute performance, but not of its risk-adjusted performance. When assessing portfolio risk, the tracking error with risk-adjusted properties is the most often used method. This method uses standard deviations of the returns to measure the dispersion of returns. Tobe (1999) defined this tracking error as the percentage difference in total return between an index fund and the benchmark it is replicating. In other words, tracking error is simply the standard deviation of the portfolio's active return, where active return is defined as the difference between the portfolio return and the benchmark return. According to him, this definition of tracking error is best used for evaluation of a passive investment vehicle such as an index fund. However, it is important to distinguish between the portfolio's average difference in returns and the tracking error. The average difference in returns is the sum of the periodic differences between the returns of the portfolio and those of the benchmark, divided by the number of observations. Average differences can understate volatility of returns where the differences are large and tend to offset each other. For example, if a series of returns contains only large differences that completely offset one another, the average difference will be zero. This will give the wrong estimate for tracking error when a portfolio's returns are clearly more volatile than those of the benchmark, even though they are estimated to sum to zero over time. The reason why tracking error is not to be confused with offsetting differences, is because it is based on the standard deviation of the periodic differences. Offsetting periodic differences will effectively enlarge the tracking error. If this offsetting did not happen, the tracking error could, in theory, be zero, because the value of $(\sum r_p - \sum r_b)^2$ could easily be zero. However, the tracking error used in our study is not distorted by return series where the periodic differences are consistently the same amount with the same sign, either positive or negative (for example: -0.05, -0.05, -0,05). The typical source of consistent negative difference is usually the portfolio's management fee. While the tracking error in our study is the average of the periodic differences, the standard error is the standard deviation of the periodic differences divided by the square root of the number of periods. Standard errors of the funds will then be used to calculate the t-statistics.

In short, we reject the use of standard deviation as the tracking error and define the tracking error used in this study as:

$$TE_{ave} = \frac{TE_a}{N}$$

where:

TE_{ave} = Out-of-sample average tracking error between a portfolio and its benchmark
N = Number of observed periods

Standard error is calculated as mentioned earlier i.e. dividing standard deviation of periodic differences between a portfolio and its benchmark by the square root of the number of periods:

$$SE = \frac{\sqrt{\frac{\sum(r_p - r_b)^2}{N-1}}}{\sqrt{N}}$$

By calculating the t-statistic for the portfolio's tracking error, the notion that the portfolio's returns are similar to those of the benchmark can be rejected or confirmed at a given confidence level using the standard approach for the t-statistic:

$$t = \frac{TE_{ave}}{SE}$$

The larger the portfolio's standard error is, the smaller and thus less significant the portfolio's t-statistic will be. As a result, portfolios with large tracking errors and large standard errors tend to have insignificant t-statistics. Their large standard errors reduce the ability of the investors to estimate their average difference, whether positive or negative. The only remaining explanation for the large tracking errors would, in the investor's eyes, be just random chance.

3.2 Sharpe ratio

Excess returns are defined as returns above the risk-free rate, which in our study are the 1 month Euribor rate for Euro zone funds and 1 month T-bill rate for the U.S. funds. Risk-adjusted returns are then calculated to rank these return series by their performance. The method of choice is the Sharpe ratio, which is a performance index using standard deviation to measure the portfolio risk (Sharpe, 1966). Portfolios with small standard deviations receive higher ratios, which is why this is an excellent tool for measuring and ranking index funds, whose sole purpose is to mirror their benchmark's performance, which in turn means mirroring the benchmark's standard deviation. The Sharpe ratio is calculated as follows:

$$SR_p = (r_p - r_{rf}) / s_p$$

where:

SR_p = Sharpe ratio of the portfolio

$\mathbf{s_p}$ = Portfolio standard deviation

$\mathbf{r_p}$ = Return on portfolio

$\mathbf{r_{rf}}$ = Return on a risk-free asset

3.3 Jobson & Korkie test statistic

To test whether the Sharpe ratios of two different portfolios are statistically distinguishable, we also compute the P-value of the difference, using the approach suggested by Jobson and Korkie (1981) after making the correction pointed out in Memmel (2003). In our study we concentrate on comparing the Sharpe ratio of each fund against the benchmark's Sharpe ratio. The null hypothesis is that there is no difference between the benchmark's and a portfolio's Sharpe ratios:

$$SR_b - SR_p = 0$$

For this to hold, the Jobson & Korkie test-statistic has to follow normal distribution at a given level of confidence. In other words, a portfolio's underperformance compared to the benchmark requires a deviation from normality. The z score is calculated as follows:

$$Z_{JK} = \frac{s_p r_{eb} - s_b r_{ep}}{\sqrt{\theta}}$$

with

$$\theta = \frac{1}{N}\left(2s_b^2 s_p^2 - 2s_b s_p cov_{b,p} + \frac{1}{2} r_{eb}^2 s_p^2 + \frac{1}{2} r_{ep}^2 s_b^2 - \frac{r_{eb} r_{ep}}{s_b\ s_p} cov_{b,p}^2\right)$$

where:

Z_{JK} = Jobson & Korkie test-statistic

s_p = Standard deviation of portfolio returns

s_b = Standard deviation of benchmark returns

r_{ep} = Mean excess return of portfolio i.e. over the risk-free rate

r_{eb} = Mean excess return of benchmark i.e. over the risk-free rate

$cov_{b,p}$ = Covariance between benchmark return and portfolio return

3.4 Jensen's alpha

Jensen (1968) described the method for calculating performance differences by utilizing the Capital asset pricing model (CAPM). This method calculates the alpha i.e. abnormal return in terms of realized rates of return, assuming that CAPM is empirically valid:

$$\alpha_j = r_p - \left(r_{rf} + \beta_p \times (r_b - r_{rf})\right)$$

where:

α_j = Jensen's alpha
r_p = Portfolio return
r_{rf} = Risk-free return
r_b = Benchmark return
β_p = Portfolio beta

This is the difference between a fund's actual return and those that could have been made on a benchmark portfolio with the same risk i.e. beta. It measures the ability of a fund's management to increase returns above those that are purely a reward for bearing market risk. Since our study includes only index funds, it might at first seem strange to measure their performance by Jensen's alpha. However, earlier we established that several index funds have been shown to exhibit active management features such as proactive index adjustments. All of the Jensen's alphas are, like the tracking errors, tested for significance using the t-test.

4 EMPIRICAL DATA

4.1 Fund samples and descriptive statistics

Our sample period includes monthly returns for ETFs and index funds tracking either the S&P 500 composite index or Dow Jones EuroStoxx 50 index from October 2004 to January 2008, resulting in 40 monthly observations. While most of the used instruments were created long before 2004, we use a common starting date for all of them to allow for an easier comparison. All of the data used were collected from Thomson ONE Banker database. Both the index and fund returns are total return series i.e. they are adjusted to dividends and stock splits, making them legitimate for comparison. This resulted in a sample of eleven index funds and six ETFs. Nowadays, more equity index funds and ETFs have been created and are available than are included in the study. However, the history of these funds is much shorter and not enough observations exist at this point to include them in this study. The index funds that could be matched to counterpart ETFs are shown in Tables 1 and 2.

Table 1. List of index funds and ETFs tracking S&P 500

S&P 500	**Mnemonic**	**NAV**	**β**
ETFs:			
iShares S&P 500 Index	(IVV)	$17.83B	1.014
SPDRs	(SPY)	$81.63B	1.009
Index funds:			
Vanguard 500 Index	(VFINX)	$67.95B	1.013
SSgA S&P 500 Index Fund	(SVSPX)	$1.97B	0.994
Fidelity Spartan 500 Index Investor	(FSMKX)	$8.50B	0.998
Fidelity Spartan U.S. Equity Index Inv	(FUSEX)	$23.84B	0.988
Morgan Stanley S&P 500 Index A	(SPIAX)	$542.73M	0.987
Barclays Global Investors S&P 500 St	(WFSPX)	$274.62M	0.999
JPMorgan Equity Index A	(OGEAX)	$653.64M	0.994

Table 2. List of index funds and ETFs tracking DJ EuroStoxx 50

Dow Jones EuroStoxx 50	Mnemonic	NAV	β
ETFs:			
UBS DJ EuroSTOXX 50	(UBS)	€731.45M	0.973
iShares DJ EUROSTOXX 50 (FRA)	(ISD)	€3.84B	1.065
iShares DJ EUROSTOXX 50 (LSE)	(ISE)	€1.96B	1.024
Lyxor DJ EUROSTOXX 50	(LYX)	€4.91B	1.036
Index funds:			
Fidelity Funds EUROSTOXX 50	(FIDEL)	€327.74M	0.992
UBS (Lux) Euro STOXX 50	(UFUND)	€913,74M	1.026
Unieurostoxx 50	(UNICR)	€2.91B	1.007
Credit Suisse Indexmatch Lux Eurostoxx 50	(CRESU)	€156.17M	0.980

It can be clearly seen, that there are extremely large variations in the net asset values both within and between the two fund types. The ratio between the biggest and the smallest ETF is over 76 to 1. The ratio between the biggest and the smallest index fund is even more staggering: approximately 275 to 1. This is most likely due to the relative adolescence of the smaller funds. The mnemonics shown in columns 2 in both tables are not all entirely accurate. Some of them have been created by the author to divide ETFs into 3-letter mnemonics such as (ISD) and index funds into 5-letter ones, for example (UNICR).

4.2 Monthly total returns

The accumulated total returns as well as the standard deviations of those returns for the index funds and for the ETFs during the entire sample period are presented in Tables 3 and 4.

Table 3. Monthly total returns of funds tracking S&P 500

Series	Sum of Total Return	Average of Total Return	Standard Dev. of Total Return
S&P500	27.451%	0.686%	2.458%
IVV	26.840%	0.671%	2.500%
SPY	26.802%	0.670%	2.491%
VFINX	24.463%	0.612%	2.495%
SVSPX	26.497%	0.662%	2.454%
FSMKX	26.208%	0.655%	2.483%
FUSEX	26.465%	0.662%	2.441%
SPIAX	24.284%	0.607%	2.464%
WFSPX	26.838%	0.671%	2.472%
OGEAX	25.651%	0.641%	2.450%

Table 4. Monthly total returns of funds tracking DJ EuroStoxx 50

Series	Sum of Total Return	Average of Total Return	Standard Dev. of Total Return
DJEURO50	40.837%	1.021%	3.576%
UBS	41.152%	1.029%	3.679%
ISD	38.813%	0.970%	3.825%
ISE	40.789%	1.020%	3.672%
LYX	41.184%	1.030%	3.720%
FIDEL	38.909%	0.973%	3.547%
UFUND	39.199%	0.980%	3.693%
UNICR	37.595%	0.940%	3.612%
CRESU	36.293%	0.907%	3.531%

As expected, not a single fund was able to track the S&P 500 index perfectly. In table 3 we can see that both of the ETFs and one index fund (WFSPX) came within a 65 basis point margin, which, from an investor's point of view, seems reasonable. Especially when we consider the 40-month long sample period. Interestingly, all other S&P 500 index funds have lower returns than both of the ETFs. On top of this, the performance differences are much larger within index funds than within the two ETFs. The best index fund (WFSPX) beat the worst index fund (SPIAX) by over 250 basis points, whereas the difference between the two ETFs is only 38 basis points.

Surprisingly, two of the ETFs tracking Dow Jones EuroStoxx 50 in table 4 outperformed their underlying index, albeit by a slim margin. On top of that, the third best ETF underperformed the index by mere four basis points. Index funds did not fare as well, although two of the best performing ones did outperform the worst ETF. The ETFs with the highest average returns are (LYX) and (UBS), with outperformances of 35 and 32 basis points respectively. Once again we can see large differences in performance between the index funds. However, this time one of the ETFs has underperformed by over 200 basis points, setting the gap between the best and the worst at around 235 basis points. For index funds there was even more variation. The performance gap was over 290 basis points.

On average, the results exhibited in Tables 3 and 4, show that the ETFs have higher mean returns than their corresponding index funds. However, they also have higher standard deviations of monthly returns. Nevertheless, there are a few exceptions where the index fund's standard deviation is greater than that of the corresponding ETF. Before we can conclude on the better investment vehicle, the estimates for risk-adjusted performances are needed.

4.3 Risk-free rates

To be able to accurately measure the Sharpe ratios for monthly returns over a multi-year period, we will use the 1-month Euribor rate as a proxy for risk-free rate of return for those funds that are tracking the Dow Jones EuroStoxx 50 index, and the 1-month T-bill rate for the S&P 500 funds. All risk-free returns are converted from their nominal values to a time series consisting of logarithmized monthly return values, as shown in Graph 1 and Table 5. The formula used for converting each annualized nominal rate of return to a logarithmized monthly rate of return was:

$$r_{rf} = LN[(1 + rf)^{1/12}]$$

Graph 1. Illustration of cumulative risk-free rates of return during the sample period

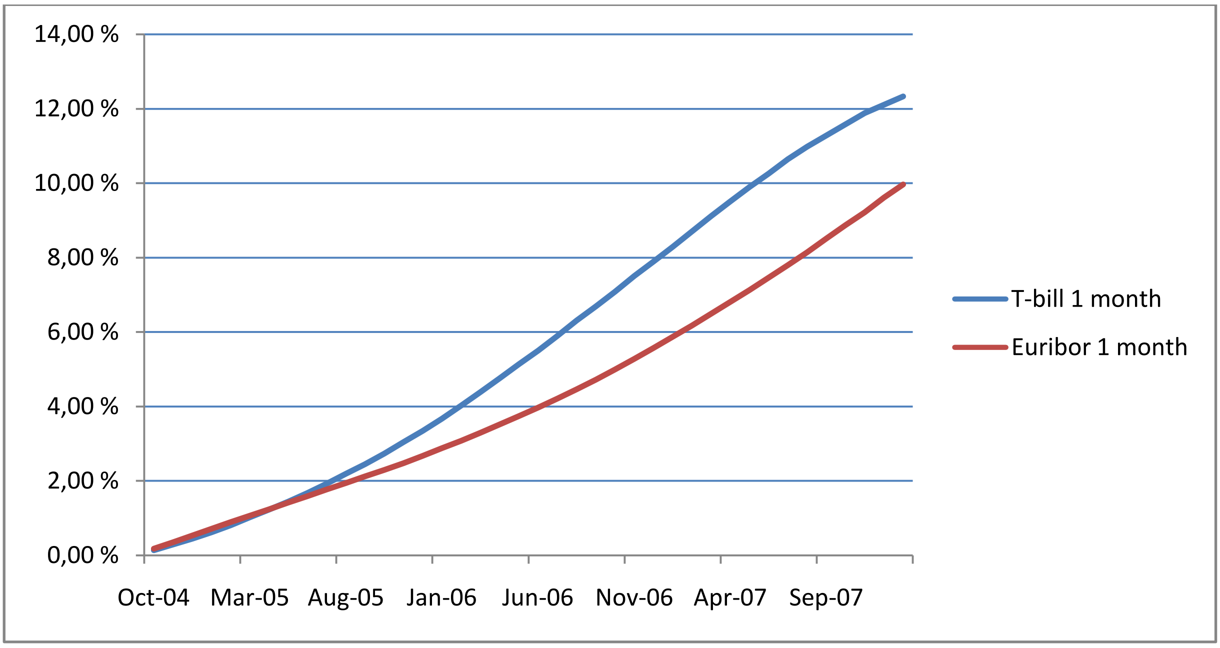

Table 5. Cumulative and monthly total returns of risk-free rates

Series	Sum of Total Return	Average of Total Return
1 Month T-bill	12.331%	0.308%
1 Month Euribor	9.961%	0.249%

Column 3 of Table 5 indicates the average risk-free rates that the funds have had to outperform in order create excess returns. These excess returns will be later used to calculate the ability of the ETFs and index funds to match the market risk premiums set by their benchmarks. But before that, we first quantify the tracking errors using the formulas described earlier in Chapter 3.

4.4 Tracking errors

The arithmetic and average monthly tracking errors for each individual ETF and index fund are reported in Tables 6 and 7. The means of tracking errors are tested for significance i.e. if they differ from zero, in Tables 8 and 9.

Table 6. Monthly tracking errors for funds tracking S&P 500

Series	Sum of monthly TE	Average of monthly TE	Standard Dev. of monthly TE
IVV	-0.6116%	-0.0153%	0.190%
SPY	-0.6491%	-0.0162%	0.235%
VFINX	-2.9880%	-0.0747%	0.155%
SVSPX	-0.9544%	-0.0239%	0.227%
FSMKX	-1.2431%	-0.0311%	0.379%
FUSEX	-0.9865%	-0.0247%	0.233%
SPIAX	-3.1669%	-0.0792%	0.430%
WFSPX	-0.6127%	-0.0153%	0.263%
OGEAX	-1.7998%	-0.0450%	0.195%

Table 7. Monthly tracking errors for funds tracking DJ EuroStoxx 50

Series	Sum of monthly TE	Average of monthly TE	Standard Dev. of monthly TE
UBS	0.3144%	0.0079 %	1.196%
ISD	-2.0246%	-0.0506 %	0.401%
ISE	-0.0482%	-0.0012 %	0.289%
LYX	0.3462%	0.0087 %	0.354%
FIDEL	-1.9281%	-0.0482 %	0.085%
UFUND	-1.6379%	-0.0409 %	0.421%
UNICR	-3.2423%	-0.0811 %	0.281%
CRESU	-4.5446%	-0.1136 %	0.480%

Table 8. Risk-adjusted tracking errors for funds tracking S&P 500

Series	Average of TE	Standard error	T-statistic
IVV	-0.0153 %	0.0308 %	-0.497
SPY	-0.0162 %	0.0381 %	-0.426
VFINX	-0.0747 %	0.0251 %	-2.978***
SVSPX	-0.0239 %	0.0367 %	-0.649
FSMKX	-0.0311 %	0.0614 %	-0.506
FUSEX	-0.0247 %	0.0378 %	-0.652
SPIAX	-0.0792 %	0.0697 %	-1.135
WFSPX	-0.0153 %	0.0426 %	-0.360
OGEAX	-0.0450 %	0.0317 %	-1.420*

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Table 9. Risk-adjusted tracking errors for funds tracking DJ EuroStoxx 50

Series	Average of TE	Standard error	T-statistic
UBS	0.0079 %	0.1940 %	0.041
ISD	-0.0506 %	0.0650 %	-0.779
ISE	-0.0012 %	0.0469 %	-0.026
LYX	0.0087 %	0.0574 %	0.151
FIDEL	-0.0482 %	0.0137 %	-3.515***
UFUND	-0.0409 %	0.0683 %	-0.600
UNICR	-0.0811 %	0.0456 %	-1.776**
CRESU	-0.1136 %	0.0779 %	-1.459*

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

In table 8 we can see that all of the funds tracking S&P 500 were found to have negative tracking errors. After testing these tracking errors for significance, however, we found out that while they are mostly negative, in aggregate, they are not significantly different from zero. Two of the index funds (VFINX) and (OGEAX) were the only ones to have statistically significant negative tracking errors, even though one of those tracking errors is significant only at the 10 % level. For the Euro zone funds shown in table 9 there is more variation between the t-values of tracking errors. The best of the ETFs (UBS) had very insignificant t-value, however, we must remember the methodologies concerning tracking error calculations described in chapter 3. We notice that (UBS) has a much higher standard error than the rest of the funds, which we already described as a cause for an insignificant t-value in Chapter 3. Three out of four index funds are found to have significant t-values for their negative tracking errors, while the ETFs are uniformly not different from zero. The findings in tables 8 and 9 support the claim that the ETFs and index funds closely mirror the underlying index, even though they may not be exact replication of the underlying index. Statistically speaking, though, all of the significant negative tracking errors were those of the index funds and not ETFs.

4.5 Risk adjusted returns

Since both table 3 and 4 indicated that the returns of ETFs are on average, higher than those of the index funds and the standard deviations of index funds are lower than those of the ETFs, a comparison of risk-adjusted returns is appropriate. As described in Chapter 3, we will first calculate the traditional Sharpe ratio to take into the account the different levels of risk i.e. standard deviations of excess returns. One must remember that the excess returns here are the returns over the risk-free rate. The tracking errors exhibited in previous tables were not excess returns. Tables 10 and 11 display the Sharpe ratios for ETFs and index funds. The first column displays the Sharpe ratios for the ETFs and index funds, while the remaining columns display the components of the ratio i.e. means and standard deviations of excess returns.

Table 10. Sharpe ratios for funds tracking S&P 500

Series	Sharpe ratio	Average of excess return	Standard Dev. of excess return
IVV	0.1455	0.3627%	2.4518%
SPY	0.1456	0.3618%	2.4929%
VFINX	0.1219	0.3033%	2.4841%
SVSPX	0.1447	0.3541%	2.4885%
FSMKX	0.1402	0.3469%	2.4478%
FUSEX	0.1452	0.3533%	2.4745%
SPIAX	0.1217	0.2988%	2.4332%
WFSPX	0.1471	0.3627%	2.4654%
OGEAX	0.1363	0.3330%	2.4435%

Table 11. Sharpe ratios for funds tracking DJ EuroStoxx 50

Series	Sharpe ratio	Average of excess return	Standard Dev. of excess return
UBS	0.2109	0.7798%	3.6976%
ISD	0.1876	0.7213%	3.8451%
ISE	0.2087	0.7707%	3.6921%
LYX	0.2087	0.7806%	3.7404%
FIDEL	0.2028	0.7237%	3.5678%
UFUND	0.1969	0.7310%	3.7119%
UNICR	0.1903	0.6909%	3.6313%
CRESU	0.1852	0.6583%	3.5551%

In general, the Sharpe ratios for the ETFs are higher than for the index funds in both markets. One index fund deserves special notice. (FIDEL) was found to have a negative tracking error that was significantly different from zero in Table 9, however, when we look at the excess returns, its very low standard deviation is the key component that raises its Sharpe ratio above other index fund competitors. This causes us to doubt the validity of both tracking error and Sharpe ratio tests conducted in this study so far. Our method of choice for testing the accuracy of the Sharpe ratio for the funds in our sample is the method first introduced by Jobson & Korkie (1981), later corrected by Memmel (2003). Determinations of the Sharpe ratio estimates are exhibited in Tables 12 and 13, where the second column describes the traditional Sharpe ratios, third column lists the Jobson & Korkie test-statistics and fourth column has the p-values of those test-statistics. A p-value less than our level of confidences (α= 0.01, 0.05, 0.10) means that given the risk characteristics i.e. standard deviation, the difference between a fund's and the benchmark's Sharpe ratio is statistically significant i.e. doesn't fulfill the requirement of normality.

Table 12. Jobson & Korkie test statistics for funds tracking S&P 500

Series	SR	Z_{JK}	P-value
IVV	0.1455	-1.4329	0.151
SPY	0.1456	-0.7930	0.428
VFINX	0.1219	-23.7462	<0.001***
SVSPX	0.1447	-0.8980	0.369
FSMKX	0.1402	-0.6494	0.516
FUSEX	0.1452	-0.8534	0.393
SPIAX	0.1217	-1.2812	0.200
WFSPX	0.1471	-0.5136	0.608
OGEAX	0.1363	-2.4644	0.014**

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Table 13. Jobson & Korkie test statistics for funds tracking DJ EuroStoxx 50

Series	SR	Z_{JK}	P-value
UBS	0.2109	-0.0181	0.986
ISD	0.1876	-0.8382	0.402
ISE	0.2087	-0.1130	0.910
LYX	0.2087	-0.1113	0.911
FIDEL	0.2028	-0.3517	0.725
UFUND	0.1969	-0.4814	0.630
UNICR	0.1903	-0.7606	0.447
CRESU	0.1852	-0.7996	0.424

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Interestingly, in table 13 we can see that all of the Euro zone funds have Sharpe ratios that do not differ from zero when compared to that of their benchmark and therefore can't be used for comparison. (FIDEL)'s highly significant negative tracking error with non-excess returns shown in table 9 doesn't correlate here with excess return measurements. However, this is not the case with two of funds tracking the S&P 500. Table 12 shows that two of the index funds (VFINX) and (OGEAX) have risk-return characteristics that produce significantly different Sharpe ratios. From earlier tables we were able to see that both of these funds had low standard deviations and low returns compared to their peers. This time the non-excess return characteristics of the two S&P 500 funds did correlate with excess return characteristics.

Last, but not least, we apply the CAPM and measure the excess return performance by testing the significance of each fund's Jensen's alpha. We divide the monthly average excess returns with the monthly average standard error of those returns to get the t-statistic for each of the fund's Jensen's alpha. This will tell us whether the expected return calculated by the slope coefficient i.e. beta of a fund matches the return observed while adjusting for monthly standard errors of those returns.

Table 14. Jensen's alpha test statistics for funds tracking S&P 500

Series	Jensen's alpha	T-statistic	P-value
IVV	-0.0206 %	-0.6770	0.498
SPY	-0.0196 %	-0.5206	0.603
VFINX	-0.0797 %	-3.2165	0.001***
SVSPX	-0.0217 %	-0.5976	0.550
FSMKX	-0.0304 %	-0.5015	0.616
FUSEX	-0.0203 %	-0.5428	0.587
SPIAX	-0.0744 %	-1.0798	0.280
WFSPX	-0.0153 %	-0.3629	0.988
OGEAX	-0.0426 %	-1.3623	0.173

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

Table 15. Jensen's alpha test statistics for funds tracking DJ EuroStoxx 50

Series	Jensen's alpha	T-statistic	P-value
UBS	0.0285 %	0.1489	0.882
ISD	-0.1011 %	-1.5753	0.115
ISE	-0.0196 %	-0.4221	0.673
LYX	-0.0192 %	-0.3385	0.735
FIDEL	-0.0419 %	-3.0951	0.002***
UFUND	-0.0609 %	-0.9034	0.366
UNICR	-0.0862 %	-1.9136	0.056*
CRESU	-0.0980 %	-1.2740	0.203

*** Significant at the 1% level
** Significant at the 5% level
* Significant at the 10% level

One index fund in both the market areas (VFINX) and (FIDEL) had highly significant negative Jensen's alphas, suggesting their performances during the 40 months observation period have been sub-par. (OGEAX) and (SPIAX) tracking the S&P 500 are also relatively close to the point where their underperformance could be interpreted as statistically significant. Compared to the index funds, the Euro zone ETFs exhibited solid p-values for Jensen's alpha except for the (ISD), which has been underperforming it's peer group throughout the study in every performance measurement. By pure comparison of Jensen's alpha, the highest score among the S&P 500 funds goes to (WFSPX), which has been the only index fund able to compete with and sometimes outperform the two consistently solid ETFs, (IVV) and (SPY).

5 RESULTS

Our objective was to measure and compare the past risk and return performances of several exchange-traded funds and mutual index funds with either S&P 500 or Dow Jones EuroStoxx 50 indices as their benchmarks. We utilized eleven index funds and six ETFs for two equity indices over the sample period between October 2004 and January 2008. The proxies for performance were average total returns and risk-adjusted returns over the whole sample period. Our results show that on average, the ETFs exhibited higher average returns than index funds as well as higher standard deviations of those returns. During the study we verified that both the ETFs and the index funds were on average, competent in mirroring their benchmark's performance. For the majority of the funds, all performance estimates used in this study were found to support the hypothesis that the costs of passive mutual fund investing, being lower than in actively managed funds, are not high enough to cause significant underperformance.

The process of defining statistically significant underperformance required a varied set of tests on the returns and risk characteristics of the funds in question. Throughout the study we noticed that most of the ETFs and index funds were consistently receiving high marks for their ability to track their benchmark's return, while individual funds did the opposite. What was also interesting to see, were the different performance estimates that, depending on the statistical test used, were given to the same fund. This is not surprising for two reasons. First of all, the margins between the returns were very small considering the sample period. Secondly, a consensus on the best fund performance estimation method is yet to be reached, especially in the index fund segment.

In Chapter 2 we reviewed the previous articles written about this topic. More and more often ETFs are criticized for their unwillingness to break the mold and incorporate slightly more active management strategies during times when benchmark compositions are changing or there is higher volatility on the markets. This, according to our results, has not been decreasing their performance. The ETFs examined in this study exhibited much more consistent performance within their group than the index funds did. All of the significantly negative performance estimates were given to certain index funds, that for some reason, could not match the best index funds let alone the ETFs. Based on total returns, tracking errors, Sharpe ratios and Jensen's alpha estimates we conclude that ETFs are, on average, better than their corresponding index funds.

When considering the abilities of investors to diversify among ETFs to reduce risk, the ETFs do serve as a better option compared to index funds when investing in an equity index portfolio. This would indicate that a more passive investment style through ETFs provides better risk-adjusted returns than a slightly more active management of assets in index funds. However, the determinants behind these differences in performance are, and have always been, under heavy debate. Each type of fund has its own advantages if, for example, we look at the expense ratios, dividends and portfolio adjustments. The quantitative effect of these factors could very well be behind the underperformance of certain funds in our sample, but considering the excellent tracking performance of majority of the examined funds, these factors might not play as big of a role as the good old random chance.

6 REFERENCES

Blume, M.E., Edelen, R.M. (2002). *On replicating the S&P 500 index.* Working Paper, The Wharton School University of Pennsylvania.

Edelen, Roger. (1999). *Investor flows and the assessed performance of open end mutual funds*, Journal of Financial Economic, 53, pages 439-466.

Elton, E.J., M.J. Gruber, G. Comer and K. Li (2002). *Spiders: Where are the Bugs*, *Journal of Business*, 75 (3), pages 453-472.

Gastineau, Gary L. (2004). *The benchmark index ETF performance problem*, Journal of Portfolio Management, 30**,** pages 96–103.

Gastineau, Gary L. (2002). *Equity Index Funds Have Lost Their Way*, Journal of Portfolio Management

Gastineau, Gary L. (2001). *Exchange-traded funds: An introduction*, Journal of Portfolio Management, vol. 27(3), pg. 88

Harper, Joel T., Madura, Jeff., Schnusenberg, Oliver. (2003). *Performance comparison between exchange-traded funds and closed-end country funds*, Journal of International Financial Markets, Institutions and Money, vol. 16(2), pages 104-122.

Jobson, J. D., Korkie, R. (1981). *Performance Hypothesis Testing with the Sharpe and Treynor Measures*, Journal of Finance, vol. 36, pages 889–908.

Kostovetsky, Leonard. (2003). *Index mutual funds and exchange-traded funds*, Journal of Portfolio Management. New York: Summer 2003. vol. 29(4), pg. 80

Memmel, C. (2003). *Performance Hypothesis Testing with the Sharpe Ratio*, Finance Letters, (1), pages 21–23.

Sharpe, W. (1966). *Mutual Fund Performance*, Journal of Business, vol. 39, pages 119-138.

Tobe, Chris. (1999). *Tracking Error: An Essential Tool in Evaluating Index Funds*, The Journal of Investment Consulting, pages 8-11.

Zitzewitz, Eric. (2002). *Who Cares About Shareholders? Arbitrage-proofing Mutual Funds*. Available at SSRN: http://ssrn.com/abstract=305410 or DOI: 10.2139/ssrn.305410