A study of the performance of exchange traded funds

Auteur : Mignolet, Arthur
Promoteur(s) : Hubner, Georges
Faculté : HEC-Ecole de gestion de l'ULg
Diplôme : Master en ingénieur de gestion, à finalité spécialisée en Financial Engineering
Année académique : 2015-2016
URI/URL : http://hdl.handle.net/2268.2/1416

A STUDY OF THE PERFORMANCE OF EXCHANGE TRADED FUNDS

Jury:
Promoter:
Georges HÜBNER
Readers:
Marie LAMBERT
Nicolas DUPONT

Dissertation by
Arthur MIGNOLET
For a degree of Master in Business Engineering specializing in Financial Engineering
Academic year 2015/2016

Acknowledgements

I would like to express my deepest gratitude to my supervisor Professor Georges Hübner for his remarks, comments, and knowledge sharing. He encouraged me to develop my own ideas and oriented me in the right direction when I needed it. His support and expertise helped me go through this master thesis. I would also like to thank Professor Marie Lambert and Nicolas Dupont, as readers of this master thesis, for their very valuable comments. I thank my friends from HEC-ULG and Kansai Gaidai University who made my years at University unforgettable. Finally, I would like to express my gratitude to my parents and family for their continuous support and encouragement through my years of study.

Thank you very much.

Arthur Mignolet

Table of contents

1. Introduction 1
2. Literature review 5
   2.1. Active and passive portfolio management 5
   2.2. Exchange traded funds 7
   2.3. Performance measures 9
   2.4. Value-at-risk, conditional value-at-risk and modified value-at-risk 17
   2.5. Performance measure test 20
3. Data and methodology 21
   3.1. The Sample 21
   3.2. Motivation for a more elaborated tracking error: The non-normality of excess returns 22
   3.3. The new performance measure 25
   3.4. Testing the measure 29
      3.4.1. Parametric tests 29
      3.4.2. Non-parametric tests 30
4. Results 33
   4.1. An example: Application on ETFs tracking the S&P 500 in 2015 33
   4.2. Results for the sample 36
      4.2.1. Results 36
      4.2.2. Rolling window analysis 37
5. Results - The efficiency of the measure 41
   5.1. One-year periods 42
      5.1.1. Parametric test 42
      5.1.2. Non-parametric tests 45
   5.2. Two-year periods 51
      5.2.1. Parametric test 51
      5.2.2. Non-parametric tests 52
   5.3. Interpretations 54
6. Possible extensions 61
7. Conclusion 63
8. References 65

1. Introduction

Exchange traded funds (henceforth ETFs) have known a growing interest over the past years. They are very similar to mutual funds; they both are collective investment vehicles. The main difference between them is that an ETF is traded on an intraday basis like a stock. Therefore, the price of an ETF will change throughout the day. Mutual funds, however, can be traded only once a day at net asset value calculated after the close (Deville, 2008).

An ETF is a collective investment vehicle, which means that the ownership for the underlying assets is divided among the shareholders. As a result, they indirectly own these assets. This investment structure, similar to mutual funds, represents an alternative for investors. Moreover, ETFs are usually more liquid and cost less than mutual funds in terms of management fees. Shareholders hold a part of the underlying assets, hence they have a share on profits that are distributed through dividends or capital gains when investments are sold.

Since the popularity of ETFs has been increasing recently, the study of the performance of this type of funds deserves attention. While much research has been done in order to measure active performance, fewer papers are dedicated to the measurement of passive performance. Moreover, the performance measures that have been proposed in the literature present some limitations. Firstly, most of these measures are not well fitted for passive management. Indeed, it would not be relevant to use a performance measure based on absolute returns to measure the performance of ETFs. That is because the goal of an ETF is to track the performance of a benchmark and not to produce high absolute returns. Secondly, the performance measures that have been proposed for ETFs do have their drawbacks.

The information ratio, as it is considered by Hassine and Roncalli (2013), does not work well for ETFs that have negative excess returns and it ignores the magnitude of the tracking error (Roncalli, 2014). The ETF efficiency indicator first introduced by Hassine and Roncalli (2013), as well as the information ratio, assumes that the excess returns of ETFs over their benchmarks are normally distributed. This is not surprising according to Fabozzi, Neave, and Zhou (2012) who state that the normal distribution is usually assumed in finance theory but that it does not correspond closely with the distributions that can be observed in real-world financial markets.

Through this thesis, I will develop an approach to improve the above measures and to propose a performance measure that can be used to efficiently assess the skills of an ETF manager. I will show that, for the sample that is used, the excess returns of ETFs are not normally distributed. Therefore, the information ratio and the ETF efficiency indicator both make unrealistic assumptions on the probability distributions.

Moreover, when considering the performance measurement of ETFs, the relative performance of the benchmark with respect to the risk-free rate is often forgotten. Hübner (2012) tackles this issue and proposes a performance measure that takes into account the excess performance of the benchmark. Since the objective of ETFs is to replicate the performance of a benchmark index, the latter should be taken into account when measuring the performance of an ETF.

The lack of consideration for the topics mentioned above motivates the need for a new performance measure for ETFs that will take into account the excess performance of the benchmark and the non-normality of excess returns. Through this thesis, I will try to answer the following question: Is it possible to propose a new efficient performance measure suited to ETFs that will take into account the relative performance of the benchmark and the non-normality of excess returns? In principle, such a performance measure should give better results than the information ratio or the ETF efficiency indicator since it takes into account more variables and make fewer assumptions. The efficiency of this new performance measure can be evaluated by persistence tests and compared with a widely accepted measure for passive management, the information ratio.

In order to answer the question introduced in the previous paragraph, I will organize the thesis as follows. The second section will be dedicated to what has already been done in the literature in terms of performance measurement and I will analyze and criticize those results. In the third section, I propose a new performance measure that aims to answer the concerns raised in the literature review and in this introduction. The new performance measure that I propose in this thesis will therefore take into account the skewness and kurtosis of the distribution of excess returns. This will be done through the use of the Cornish-Fisher expansion (Cornish & Fisher, 1937). In order to create the new performance measure, I will first build a modified tracking error that takes into account the skewness and the kurtosis of the distribution of excess returns. The elaboration of a modified tracing error will be justified by the non-normality of excess returns of the ETFs from the sample. Then, I will consider the research done by Hübner (2012) to take into account the relative performance of the benchmark with respect to the risk-free rate of interest.

The fourth and fifth sections of this thesis will be dedicated to the tests of the new performance measure. I will apply it on a sample of 30 ETFs and I will use different rolling windows in order to graphically analyze the results. Finally, I will test the robustness of the new performance measure in its measurement of the performance persistence. In order to achieve that, I will run a range of statistical tests and try to find evidences of dependency between the performances in the different periods to conclude that there is persistence in the rankings. I expect the measure to show evidences of performance persistence through the period of study and to be able to highlight the skills of the superior managers. Since the new performance measure will take into account more parameters than the information ratio, it should have a better power in selecting funds managed by superior managers with persistent performance. Therefore, the tests of persistence should give better results.

This new measure of performance will obviously present some limitations. Indeed, I expect some drawbacks with respect to the modified tracking error since the Cornish-Fisher expansion requires some conditions to be fulfilled in order to be applied (Cavenaile & Lejeune, 2012). Moreover, I expect the new performance measure not to be easy to apply since it involves many variables and calculations. This is why I will conclude this thesis with a section that will be dedicated to hypothetical extensions to improve the performance measure.

To sum up, this thesis will focus on the elaboration of a new performance measure for ETFs. The motivations for a new performance measure are driven by the drawbacks from the performance measures that have been proposed in the literature so far. They are the normality assumption regarding the excess returns of an ETF over its benchmark and the lack of consideration for the relative performance of the benchmark with respect to the risk-free rate of interest. Finally, I will test the new performance measure and assess its efficiency. This thesis aims to explore a topic that deserves more attention in the literature. This topic is the study of the performance of ETFs.

2. Literature review

2.1. Active and passive portfolio management

This thesis emphasizes on ETFs, which are mostly passively managed funds. The different portfolio management strategies fall into three categories, which are active portfolio management, passive portfolio management and a mixed strategy of the two (Jeurissen & van den Berg, 2005). Active portfolio managers believe that the markets are not efficient and that it is possible to consistently beat the market. Therefore, they will try to produce an added value, the alpha, which is the difference between the returns of the portfolio and the returns of the benchmark. Passive portfolio managers, however, believe that the markets are efficient and that it is not possible to consistently produce a positive alpha. They will try to replicate the performance of a benchmark index, an approach that is referred to as indexing or tracking. Reilly and Brown (2012) propose the following breakdown of the total actual return that a portfolio manager tries to produce in order to highlight the difference between active and passive portfolio management.

$$Total\ Actual\ Return = [Expected\ Return] + [\text{"Alpha"}]$$

$$= \underbrace{\underbrace{[Risk - Free\ Rate + Risk\ Premium]}_{\text{Passive}} + [\text{"Alpha"}]}_{\text{Active}}$$

According to Sharpe (1991) the performance of active portfolios can be assessed by comparing them with their comparable passive alternatives. The breakdown from Reilly and Brown (2012) illustrates it well since the alpha is defined as the difference between the returns from the active and the passive portfolios. Therefore, a good active portfolio manager will be someone able to generate a positive alpha and to consistently outperform the market.

Active and passive management styles both have their strengths and weaknesses. Active strategies imply higher fixed costs and high transaction fees. If the portfolio managers are competent, these fees will be offset by the returns. Passive strategies however benefit from lower fixed costs and transaction fees but do not provide excess returns over the benchmark. The returns are therefore disappointing when the index that is being replicated performs poorly.

Recently, there has been a growing interest for passively driven portfolios in the US and in Europe. This trend is due, among others, to empirical analyses showing that most of the actively managed funds have failed to outperform their benchmark in the long term (Beasley, Meade, & Chang, 2003). Currently, there is an interesting debate in the literature over active and passive management. Some authors highlight the benefits and the efficiency of the passive management style (Malkiel, 1995, 2003; Rudd, 1986). Others highlight the benefits of active portfolio management. For example, Shankar (2007) shows that actively managed indexes perform better than passively managed indexes. The very existence of passive management is even put to question by Fuller, Han and Tung (2010). This debate has brought an interest on passive management. This is the reason why index and tracking funds are becoming more and more important. Hence, there is a growing interest for investment vehicles, such as ETFs, which are mostly passively managed.

There are two main strategies to build an index or a tracking portfolio. The first one consists in a full replication of the benchmark by buying exactly the same stocks in the same proportions as in the index. Such a strategy does not seem very achievable given the high costs that it would generate. The portfolio would need to be rebalanced each time there is a change in the composition of the index. Moreover, if stocks that account for a very small percentage of the index were to be included in the portfolio, it would be administratively heavy and not really necessary (Beasley et al., 2003).

This is why, most of the time, a partial replication strategy is preferred in order to reduce the transaction costs. However, as the portfolio is composed of fewer stocks than the benchmark, there will be a tracking difference. Beasley et al. (2003) call the index tracking problem the minimization of the tracking difference and of the transaction costs. One of the goals of an ETF will then be to have a tracking error (the volatility of the tracking difference) as low as possible for a given amount of management fees.

Although most ETFs are passively managed, some of them can be actively managed. Schizas (2014) studies the performance of the first active ETF and compare it to the performance of passive ETFs, mutual funds and hedge funds. The results show that there is a strong link between active and passive ETFs but that passive ones perform better. However, he states that there is room for improvement because active ETFs are a rather new concept.

2.2. Exchange traded funds

ETFs have grown in popularity over the last few years. Abner (2013) show that they experienced a significant rise in terms of assets under management and in terms of number of products. He states that one important factor responsible for the growth in the ETF market is that investors have learned their mechanisms and are now able to use more ETF products.

ETFs present the advantage of offering arbitrage opportunities. This is due to the fact that the ETF and its underlying asset are traded during the day. For instance, if an investor can buy shares of an ETF for less than its underlying assets, he will buy the shares of the ETF and sell the underlying. Ben-David, Franzoni, and Moussawi (2012) have shown evidences of arbitrage practices between ETFs and their underlying securities.

The main goal of most ETFs is, as index funds, to replicate the performance of a benchmark. Therefore, they are mostly passively managed and benefit from lower management fees. However, as mentioned before, actively managed ETFs have emerged recently. There also exist leveraged ETFs whose goal is not to exactly replicate the performance of a particular benchmark but to replicate 2, 3 or -2 times its performance. Examples from Avellaneda and Zhang (2009) involve ETFs that offers a daily exposure to 2 or -2 times the Dow Jones Financial index.

Benchmark tracking can be achieved through full replication or partial replication. The partial replication of an index will generate a tracking difference because the returns of an ETF will never be exactly the same as the returns of its benchmark unless a full replication strategy is used. Hence, one of the goals of the ETF managers will be to minimize the difference between the performance of the ETF and the performance of its benchmark. The minimization of this difference is referred to as the tracking error problem.

ETFs represent an interesting alternative to index mutual funds and present some advantages:

- They are more liquid than mutual funds since they can be traded on an intraday basis.
- The composition of an ETF is known and there is transparency.
- They open the door for arbitrage opportunities (Avellaneda & Zhang, 2009).
- ETFs are by far more tax-efficient than mutual funds (Russel, 2013).

In the literature about ETFs, empirical studies have been performed in order to assess the efficiency of the market. The United States represents almost 70% of the ETF market in terms of assets under management, although the number of exchange traded products in the U.S. accounts for only 30% of the world total (Abner, 2013). The size of the market has been increasing tremendously since 2000. According to Roncalli (2014), the European ETF market is less efficient than the U.S. ETF market. Moreover, ETFs have a bigger importance in the United States. Indeed, the market share of ETF products in the U.S. compared to other investment product is also bigger than their market share in Europe, respectively 9.3% against 4% in Europe (Roncalli, 2014).

He shows that the ETF market is much more concentrated in equities and commodities than are mutual funds. He also states that the U.S. market is more concentrated than the European market since the top 5 largest ETFs represent 22.7% of the market in the United States. Leaders on the ETF market include SPDR, iShares, db X-trackers, PowerShares and Vanguard.

2.3. Performance measures

The performance of ETFs will be assessed by their ability to meet the goals mentioned above. There are a lot of ways to measure the performance of a portfolio. A lot of performance measures have been proposed in the literature so far. According to Jensen (1968), a measure of portfolio performance should be able to assess a manager's ability to increase the return of the portfolio through successful prediction and his ability to minimize the unsystematic risk of the portfolio. Reilly and Brown (2012) presented some of the most widely used measures to assess the performance of a portfolio, which are the Sharpe ratio (Sharpe, 1994), the information ratio (Grinold, 1989), the Jensen's alpha (Jensen, 1968) and the Treynor ratio (Treynor, 1965).

The Sharpe Ratio: $S = \frac{\bar{d}}{\sigma_d}$ where $\bar{d}$ is the expected value of $\tilde{d} = \widetilde{R_p} - \widetilde{R_b}$

The Sharpe ratio is expressed as a measure of the expected excess return, between a portfolio and the benchmark, per unit of risk associated with this return (Sharpe, 1994). However, Sharpe (1994) states that this ratio should only be used to compare portfolio between them, not to decide if one particular portfolio is worth investing or not. He explicitly says that the Sharpe ratio should be used to measure the expected return per unit of risk only for zero investment strategies.

The Information Ratio: $IR[r(A)] = \frac{E[r(A)]}{Std[r(A)]}$

Grinold (1989) proposes the information ratio, which is calculated by dividing the expected active returns of the portfolio, the alpha, by the standard deviation of the portfolio's active return. "The denominator, also called "tracking error", reflects the cost of an active management" (Cogneau & Hübner, 2009). The expected active return is the difference between the expected return of the portfolio and the expected return of the benchmark. One can notice that this ratio is very similar to the Sharpe Ratio. Indeed, they are both indicators of performance per unit of risk. Moreover, the information ratio is equal to the Sharpe ratio when the benchmark is the risk-free rate of interest.

Jensen's Alpha : $\alpha_i = R_i - R_f - \beta_i(R_M - R_f)$

Jensen (1968) proposes a measure of portfolio performance that is derived from the CAPM. The alpha in this formula can be interpreted as the part in the returns of the portfolio which is attributable to the ability of the manager to generate above average returns adjusted for risk

(Reilly & Brown, 2012). However, unlike the Information ratio or the Sharpe ratio, the Jensen's Alpha does not assess the manager's ability to reduce the unsystematic risk since it takes the beta as risk parameter.

<u>The Treynor Ratio</u> $TR_P = \frac{E(R_p)-E(R_b)}{\beta_P}$

This ratio was first presented by Treynor (1965). It is similar to the Sharpe Ratio, except that it only takes into account the systematic risk. Therefore, it is obvious that it will equal the Sharpe ratio for well diversified portfolios. The Treynor ratio measures the expected differential return, between a portfolio and the benchmark, per unit of systematic risk.

Although these measures are among the most famous in portfolio management, they obviously do not represent the only ways to assess the performance of a portfolio. Cogneau and Hübner (2009) identified more than 100 ways to measure portfolio performance. However, a deep analysis of those measures is beyond the scope of this thesis. Moreover, there is an important problematic that I have not mentioned yet, that is the relevance of these ratios for passive management.

The measures presented above do have their limits and should only be used when appropriate. It would be irrelevant to use a measure that focuses on absolute returns to assess the performance of a passively managed ETF. Indeed, in passive management, the main goal is to replicate the performance of a benchmark. Hassine and Roncalli (2013) state that, for passive investors, absolute performance is meaningless. In this context, performance measures such as the Jensen's alpha and the Sharpe ratio seem irrelevant. From there, it can be easily seen that active and passive portfolio management, because they have different goals, should use different performance measures or should give them different interpretations. The measures presented above mainly work well for the case of actively managed portfolios. A performance measure for ETFs will have to be able to assess their efficiency and to distinguish good ETFs from poorly managed ETFs.

Roncalli (2014) defines a good ETF as "a fund that presents the lowest risk in relation to the index that it replicates". According to him, the efficiency of an ETF can be assessed by looking at three factors, the tracking difference, the liquidity spread and the tracking error. However, managers should also focus on management fees as they represent an advantage that ETFs have over actively managed funds.

While the tracking difference and the tracking error are topics that are well discussed in the literature in the context of the index tracking problem (see Roll, 1992; Jeurissen & van den Berg, 2005; Beasley et al., 2003; Barro & Canestrelli, 2013)[1], the liquidity issue is more often forgotten. However, as liquidity is supposed to be one of the main advantages that ETFs have over mutual funds, the topic requires no less attention. The tracking difference is the difference, positive of negative, between the returns of an ETF and the returns of a benchmark over a given period of time. The tracking error is the volatility of the tracking difference and is usually measured by the standard deviation of the tracking difference as in Roncalli (2014).

When considering ETFs, the performance of the benchmark should be taken into account. Hübner (2012) expresses the performance of a portfolio as a difference of a fraction of the alpha and a fraction of the excess return over the risk-free rate. The performance is expressed as "excess return of the leverage portfolio with an equivalent risk to the benchmark" (Hübner, 2012). The rational investor wants to maximize this equation:

$$\pi_P = \frac{\sigma_B}{\sigma_P}\left(R_P - R_f\right) - (R_B - R_f) \tag{1}$$

The return of a portfolio can be expressed as the sum of 3 components, as in Equation (2). The first one is the alpha, which represents the average excess performance over the return of the benchmark. The second one is the return of the benchmark and the third one is the volatility of excess return, which has already been introduced in this thesis as the tracking error.

$$R_P = \alpha_P + R_B + \varepsilon_P \tag{2}$$

Merging Equations (1) and (2) gives:

$$\pi_P = \frac{\sigma_B}{\sigma_P}\left(\alpha_P + R_B + \varepsilon_P - R_f\right) - (R_B - R_f)$$

[1] Roll (1992) states that, in the case of active portfolio management when there is benchmark, minimizing the tracking error is not compatible with making the portfolio dominant in the mean/variance sense. He focused on the minimization of the tracking error for a given level of excess return.
The index tracking problem is a problem of portfolio optimization. In order to deal with it, several solutions and models have been proposed in the literature. Jeurissen and van den Berg (2005) propose to use a hybrid genetic algorithm to select the best tracking portfolio. In 2003, Beasley, Meade and Chang present an evolutionary heuristic approach to solve the index tracking problem. Their model can include a constraint on transaction costs as well as a constraint on the maximum number of stocks that can be used to replicate the index. Barro and Canestrelli (2013) propose a multi-period model to minimize the tracking error. They also take into account the downside risk. The goal in passive portfolio management is to replicate an index. However, even if the tracking difference and the tracking error are really small, if the benchmark drops, the fund will replicate this poor performance. The double tracking error model that they present aims to protect the investor against a significant drop in the benchmark.

$$= \frac{\sigma_B}{\sigma_P} \alpha_P - (1 - \frac{\sigma_B}{\sigma_P})(R_B - R_f)$$

$$=> \pi_P = \frac{\sigma_B}{\sigma_P} \alpha_P - (\frac{\sigma_P - \sigma_B}{\sigma_P})(R_B - R_f) \quad (3)$$

Which is the one factor performance measure introduced by Hübner (2012). The quantity $\frac{\sigma_P - \sigma_B}{\sigma_P}$ is defined as normalized idiosyncratic risk and is linked to the tracking error by the following relationship:

$$\iota_P = 1 - \frac{\sqrt{\sigma_P^2 - \sigma_\varepsilon^2}}{\sigma_P} \quad (4)$$

This performance measure for active portfolio management, characterized by Equation (3), presents the advantage of taking into account the excess performance of the benchmark over the risk-free rate of interest. For example, if two funds have the same alpha and the same tracking error, they will have the same information ratio. However, depending on the excess returns of their benchmarks over the risk-free rate, one manager will bring a more important return per unit of risk to the passive portfolio. Therefore, their performances should be different. According to Hübner (2012), these differences in benchmark returns exist in practice, and should not be denied. This framework has important implications for performance measurement, especially for passive management since it proposes a performance measure that takes into account the relative performance of the benchmark with respect to the risk free rate.

Hassine and Roncalli (2013) consider using the information ratio to compare the performance of ETFs. In their paper, they highlight what drives investors' decision when choosing an ETF, which is to choose the one that most accurately follows the benchmark. Therefore, in that context, performance measures focusing on absolute return are not relevant for passive investors (Hassine & Roncalli, 2013). They propose a way of measuring the performance that is an adaptation of the Markowitz approach for selecting portfolios. Indeed, they think that this approach is less relevant when there is a benchmark.

Back in 1952, Markowitz proposed a theory to choose efficient portfolios. This theory states that there is not only one, but a set of efficient portfolios. Markowitz uses the return-variance of return rule, making the assumption that investors want to maximize E, the expected return and V, the variance of expected return, which is used as measure of volatility. This theory however is not really suited for benchmark replication strategies. That is why Hassine and Roncalli (2013) tried to adapt it. They replaced the volatility of the expected return of the portfolio by the tracking error, which is the volatility of the tracking difference between the returns of the fund and the returns of the benchmark. They also chose to consider the expected excess return, which is computed as the difference between the expected return of the fund and the expected return of the benchmark.

$$Tracking\ difference = \sum_{i=1}^{n} x_i R_i - \sum_{i=1}^{n} b_i R_i = (x-b)^T R$$

$$Expected\ tracking\ difference = \mu(x|b) = (x-b)^T \mu \quad (5)$$

The investor will obviously try to maximize this expected tracking difference while trying to minimize its volatility. As I already mentioned, the volatility of this tracking difference is referred to as the tracking error.

Hassine and Roncalli (2013) compute the tracking error using the matrix notation:

$$\sigma^2(x|b) = (x-b)^T \textstyle\sum (x-b) \rightarrow \sigma(x|b) = \sqrt{(x-b)^T \textstyle\sum (x-b)} \quad (6)$$

Finally, they consider the information ratio as defined in Grinold and Kahn (2000), which is:

$$IR(x|b) = \frac{\mu(x|b)}{\sigma(x|b)} = \frac{(x-b)^T \mu}{\sqrt{(x-b)^T \sum (x-b)}} \quad (7)$$

They show that, when considering two portfolios that track the same benchmark, the one that should be preferred is the one with the greatest information ratio. However, this way of comparing the performance of two passively managed portfolios presents two drawbacks (Roncalli, 2014). Firstly, if a portfolio provides a slightly lower return than the benchmark with a very low tracking error, it will have a lower information ratio than a portfolio that slightly outperforms the benchmark with a significant tracking error. However, it does not mean that the first portfolio provides the investor with a less good tracking. The first problem with this measure is that it does not always work when comparing portfolios with positive and negative excess returns.

Secondly, this measure ignores the magnitude of the tracking error (Hassine & Roncalli, 2013). For example, a portfolio with high excess returns and a high tracking error could have a better information ratio than a portfolio with low excess returns and a low tracking error. However, the second portfolio would do a better job at replicating the index.

These problems are illustrated in Table 1. Here, four ETFs tracking the same benchmark are considered (the amounts are expressed in bps).

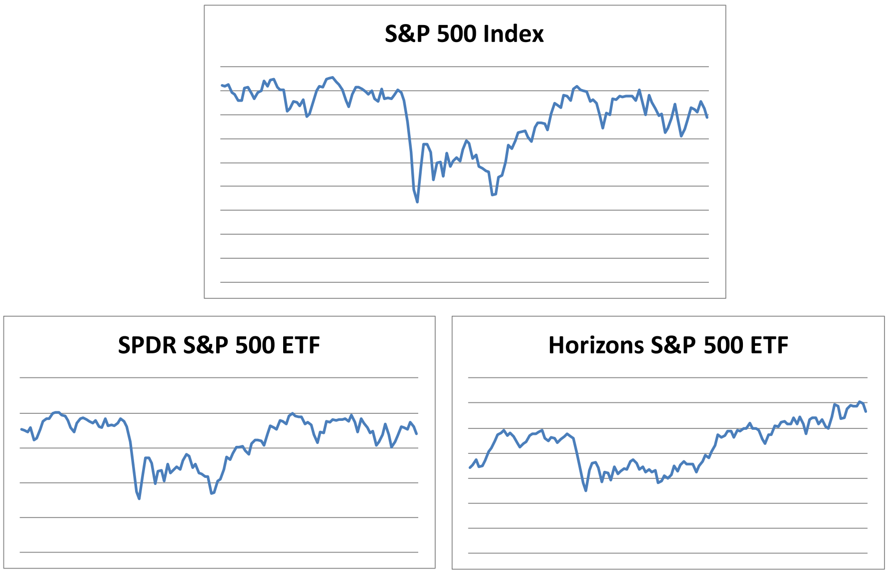
Table 1: Weaknesses of the information ratio

ETF	Excess Return	Tracking Error	Information Ratio
#1	-1	1	-1
#2	3	5	0.6
#3	2	15	0.13
#4	70	80	0.88

According to the information ratios, the ETF #4 seems to be the best one and the #1 appears to be the less efficient. However, it is clear that the ETF #4 does a very poor job at replicating the benchmark. This fund provides indeed very good relative returns but this is achieved at the expense of a very high volatility of the tracking difference. Therefore, for a passive investor who mainly wants to track an index, the ETF #1 would be much better.

To give a more realistic example of those weaknesses, I take 2 ETFs tracking the S&P 500, namely the SPDR S&P 500 ETF and the Horizons S&P 500 from July 2015 to December 2015. Their respective information ratios are 15.7296% for the SPDR ETF and 16.6422% for the Horizon ETF. Clearly, it appears that the Horizons S&P 500 ETF is a better ETF than the SPDR one. However, from the graphs in Figure 1, it can be easily inferred that the SPDR ETF does a better job at replicating the index. The graphs represent the daily adjusted close prices from the 1st of July 2015 to the 31st of December 2015.

Figure 1: Information ratio and tracking efficiency



In order to tackle these issues, Hassines and Roncalli (2013) propose a new performance measure for trackers and, more particularly, for ETFs. They compute the relative PnL of an investor with respect to a benchmark:

$$\Pi(x|b) = Tracking\ difference - bid - ask\ spread$$

The bid-ask spread is the difference between the highest price that a buyer is willing to pay for an asset and the lowest price for which a seller is willing to sell it. The tracker efficiency measure that they propose is a risk measure applied to the loss function of an investor $(-\Pi(x|b))$. They use the value-at-risk as risk measure and define their new efficiency measure as follows:

$$Efficiency\ measure = -\{inf\ \zeta : \Pr(\{L(x|b) \leq \zeta\} \geq \alpha\} \qquad (8)$$

In Equation (8), $L(x|b)$ is the loss function of an investor and is equal to $-\Pi(x|b)$. If F is the probability distribution function of the loss function, the efficiency measure equals $-F^{-1}(\alpha)$ and can be interpreted as the maximum relative expected loss that an investor can make. Therefore, the final performance measure than Hassine and Roncalli present is based on this value-at-risk framework and is computed as follows:

$$\zeta_\alpha(x|b) = \mu(x|b) - s(x|b) - \Phi^{-1}(\alpha)\sigma(x|b) \qquad (9)$$

Roncalli (2014) refers to this new efficiency measure as the ETF efficiency indicator.

In order to be able to compute the ETF efficiency indicator, one important assumption must be made. Indeed, for this measure to be relevant, the excess returns of the ETF over its benchmark must be normally distributed. Therefore, this performance measure becomes irrelevant when the excess returns are not normally distributed and when the skewness and excess kurtosis parameters of the distribution are significantly different from 0. I will demonstrate that, for the sample I use, the excess returns are not normally distributed and that we cannot use Hassine and Roncalli's performance measure. This normality assumption is a big drawback of this ETF efficiency indicator. That is because, in reality, returns are usually not distributed according to a normal distribution (Fabozzi et al., 2012).

This thesis will focus on improving this measure in order to eliminate this drawback and to propose a performance measure suited to ETFs. Therefore, it will be needed to first propose a more elaborated tracking error which will take into account the skewness and the excess kurtosis of the distribution of the excess returns.

I will therefore go one step further by considering the Modified VaR framework in order to take into account the 3rd and 4th moments of the distribution of excess returns. The proposition of a modified tracking error will be the first step toward the elaboration of a new performance measure.

As a result, the value-at-risk, the conditional value-at-risk and the modified value-at-risk will all be considered to build a modified tracking error that will take into account the skewness and the kurtosis of the distribution of excess returns. That is why the literature about the value-at-risk also needs to be considered and studied.

2.4. Value-at-risk, conditional value-at-risk and modified value-at-risk

The value-at-risk (VaR)

The VaR can be interpreted as the maximum expected loss that an investor will make at a given level of confidence. For example, a VaR at the 99% level of confidence means that an investor would have a probability of 1% to lose more than the VaR. The VaR for a level of confidence of 1-α can be rigorously described as:

$$P(r_t \leq VaR_{1-\alpha}) = \alpha \ where \ r_t \ is \ the \ return \ at \ time \ t$$

There are usually three ways of calculating the VaR, namely the historical method, the delta normal method and the Monte Carlo simulation[2]. These methods are presented and applied in Bohdalová (2007). The historical method, unlike the others, is a non-parametric method as it does not rely on any assumption regarding the distribution of the returns. This technique assumes that what happened in the past will happen again in the future. For example, the VaR at the 99% level of confidence is simply the 99^{th} percentile return. It means that there is a 1% probability to have lower returns than the 1% worst returns from the past.

The delta normal method is a parametric one since it assumes that the returns are distributed according to a normal distribution. In this case the VaR mainly depends on the volatility since it is computed by multiplying the volatility, measured by the standard deviation, by the value of the quantile for the appropriate level of confidence. For example, assuming that the returns are normally distributed and that the volatility is 2%, the VaR at the 99% level of confidence equals $0.02 * (-2.3263) = -4.65\%$. This means that an investor would have a 1% probability to make a loss of more than 4.65%. This method is the one used by Hassine and Roncalli (2013) to compute the VaR in their ETF efficiency indicator.

The VaR is useful because it represents a consistent measure of risk across all kind of markets and can take into account interrelationships between different risk factors (Bohdalová, 2007). However, it lacks subadditivity and is thus difficult to optimize (Uryasev, 2000). Therefore, the conditional VaR or expected shortfall, an alternative to the VaR, is often used in practice.

The conditional value-at-risk (CVaR) or expected shortfall (ES)

If the VaR represents the maximum expected loss at a given confidence level, the CVaR represents the expected loss when the VaR border is violated. Therefore, it is straightforward to deduce that the CVaR is always superior to the VaR. Given the lack of subadditivity and

[2] As the Monte Carlo simulation will not be used in this thesis, it will not be presented in details.

convexity of the VaR, the CVaR, which is a more coherent measure of risk, is much easier to optimize (Rockafellar and Uryasev, 2000). Maillard (2012) defines the CVaR for a level of confidence of 1-α as follows:

$$CVaR_{1-\alpha} = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_\alpha^2}{2}}$$

<u>The modified value-at-risk</u>

The modified VaR or Cornish-Fisher VaR uses the Cornish-Fisher expansion and, therefore, takes into account the 3^{rd} and 4^{th} moments of the distribution. In 1937, Cornish and Fisher proposed a way to transform a standard normal random variable into a non-normal random variable taking into account the skewness and the kurtosis of the distribution. This is very useful as returns are not usually normally distributed. The Cornish-Fisher expansion can be written as such:

$$z_{CF} = z + \frac{1}{6}(z^2 - 1)S + \frac{1}{24}(z^3 - 3z)K - \frac{1}{36}(2z^3 - 5z)S^2$$

Where z is a standard Gaussian random variable and z_{CF} a non-Gaussian random variable. S and K are the skewness and the excess kurtosis parameters of the initial distribution respectively. The modified VaR at the 1-α level of confidence is presented in Cavenaile and Lejeune (2012) as follows:

$$MVaR_{1-\alpha} = \mu + z_{CF,\alpha}\sigma$$

Where $z_{CF,\alpha}$ represents the α-quantile of the transformed distribution. The strength of this risk measure is that it takes into account the moment of order 3 and 4 of the distribution.

Maillard (2012) defines the conditional VaR of the transformed distribution as follows:

$$MCVaR_{1-\alpha} = CVaR\left[1 + z_\alpha\frac{S}{6} + (1 - 2z_\alpha^2)\frac{S^2}{36} + (-1 + z_\alpha^2)\frac{K}{24}\right]$$

It can be inferred that the Cornish-Fisher expansion provides an easy way to express the modified VaR and the modified CVaR as a function of the skewness and excess kurtosis parameters (Maillard, 2012).

In 2012, Cavenaile and Lejeune studied the consistency of the modified VaR for various confidence levels and found out that it was not consistent with the widely used 95% level of confidence. Because the VaR represents the maximum expected loss, its value will be negative. Therefore, rational investors want to have the highest VaR as possible. The quantile of the transformed distribution should then be an increasing function of the skewness and a decreasing function of the kurtosis. By taking the first derivative of the quantile of the transformed distribution with respect to the kurtosis, they show that the modified VaR is not consistent for levels of confidence below 95.84%. Moreover, when they compute the first derivative with respect to the skewness, they find minimum thresholds for the skewness at each level of confidence that is above 95.84%.

2.5. Performance measure test

Performance measurement is well discussed in the literature. However, tests of the performance measures are not less important. Indeed, once a performance measure is created, it needs to be assessed. Hübner (2005) assesses the quality of a performance measure on two dimensions. The first one is its precision in reproducing the true ranking and the second one is the stability of the ranking it produces under alternative asset pricing models. He shows that the generalized Treynor ratio is better at ranking funds than the information ratio or the alpha.

In order to assess the quality of a performance measure, one needs to study its robustness in the measure of the persistence of the performance as in Hübner (2005). Hence, a performance measure will be efficient if it can well identify ETF managers with superior skills. In order to do that, one can test the persistence of a sample of funds over a given time horizon and compare it with other performance measures.

Several studies have been done to assess the persistence of funds over time. Kuo and Mateus (2008) study the performance and the persistence of country-specific ETFs. They considered two 30 month periods. Their results show that there is performance persistence and that past results can predict future results. That is because winners tend to stay winners and losers tend to stay losers.

Brown and Goetzmann (1995) study the performance of mutual funds and show that there exists persistence in risk-adjusted performance of mutual funds. In order to show the persistence of the funds, they create the odds ratio, which is the ratio of the funds that repeat performance to the funds that do not repeat. Malkiel (1995) showed that mutual funds have underperformed the market from 1971 to 1991. He also studied the persistence of the mutual funds and found that there was strong persistence during the 1970s but not in the 1980s. He also introduced his own test statistic to test the persistence of the performance.

Other performance persistence studies have been performed in particular markets. Babalos, Caporale, Kostakis and Philippas (2008) study the persistence in mutual fund performance on the Greek market. Their results show persistence from 1998 to 2001 but they find no evidence of persistence after 2001. Ferruz, Sarto & Vargas (2004) study the persistence of Spanish investment funds and find evidences of persistence in short-term fixed-income mutual funds.

3. Data and methodology

In this section, I will first present the sample that I will use in order to test the new performance measure that I propose. I will also run a few normality tests on this sample to prove that the excess returns are not normally distributed for a given period and to justify the need for a new performance measure that takes into account the skewness and the kurtosis of the distribution. Then, I will describe how I will construct this new measure. Finally, I will present the various statistical tests that I will use in order to test the efficiency of this new performance measure.

3.1. The Sample

In order to apply this new measure of performance and to assess its efficiency, I selected 30 ETFs[3] tracking various benchmarks. The period of study goes from 2012 to 2015. The data was retrieved from Yahoo Finance. I selected ETFs that were tracking different benchmarks in different countries in order for the sample to be as representative as possible. However, these 30 ETFs do not represent all the ETFs that are traded in the world and, therefore, the conclusions that will be drawn from this sample might not be valid for other ETFs. The sample that I will use is presented in Table 2.

Table 2: The sample

ETF	Benchmark
SPDR S&P 500 ETF	S&P 500
iShares Core S&P 500	S&P 500
Vanguard 500 ETF	S&P 500
Horizons S&P 500 ETF	S&P 500
iShares Russell 3000	Russell 3000
SPDR Russell 3000 ETF	Russell 3000
iShares Russell 2000	Russell 2000
iShares Nasdaq Biotechnology	NASDAQ Biotechnology
PowerShares QQQ ETF	NASDAQ-100
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX
iShares S&P 100	S&P 100 INDEX
iShares Russell 1000	RUSSELL 1000 INDEX
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX
iShares PHLX Semiconductor	PHLX Semiconductor
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH
db x-trackers - DAX UCITS ETF (DR)	DAX
Lyxor DAX (DR) UCITS ETF	DAX
ComStage - DAX TR UCITS ETF	DAX
iShares US Financial Services	Dow Jones U.S. Financials Index
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite
Vanguard REIT ETF	MSCI US REIT INDEX
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50
SSGA SPDR AEX EUR	AEX
ISHARES AEX EUR	AEX

[3] At least 30 ETFs were needed to meet the required conditions of the approximation of the binomial law by the normal law. This will be discussed later when I will consider several testing procedures to assess the quality of the measure.

3.2. Motivation for a more elaborated tracking error: The non-normality of excess returns

The first thing that I will test for these ETFs is the normality of their excess returns with respect to their benchmarks. Indeed, the non-normality of excess return, for at least some funds, would prove the relevance of a modified tracking error in the new performance measure. However, if all the above ETFs happen to have normally distributed excess returns, the modified tracking error would lose its credibility since the standard deviation alone will be more than sufficient to measure the volatility of excess returns. My goal will be to show that the ETFs from the sample do not necessarily have normally distributed excess returns and that, therefore, a more elaborated tracking error taking into account the 3rd and 4th moments of the distribution is relevant. I will test the normality of excess returns for the 30 ETFs for the year 2012. SAS Studio offers two ways to check the normality of a probability distribution. The first way is the use of the Shapiro-Wilk test. The objective of this test is to provide a test statistic to assess the normal distribution of the ETFs (Shapiro & Wilk, 1965). The test statistic is equal to:

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_i\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \; where\ a\ is\ a\ constant$$

Here, the null hypothesis is the normality of excess return. If the above W statistic is below a given threshold, normality is rejected.

The second way used by SAS Studio to test normality is the use of goodness-of-fit tests based on the empirical distribution function (EDF). The EDF goodness-of-fit tests measure the differences between $F_n(x)$ the empirical distribution function and $F(x)$ the distribution function. The EDF statistics give more powerful tests for the null hypothesis than chi-square tests (Stephens, 1974). Here are the 3 EDF statistics used by SAS Studio to test normality.

Kolmogorov-Smirnov statistic: $D = sup_x |F_n(x) - F(x)\,|$

The Anderson-Darling statistic: $A^2 = n\int_{-\infty}^{+\infty}\big(F_n(x) - F(x)\big)^2\,\big[F(x)\big(1 - F(x)\big)\big]^{-1} dF(x)$

The Cramér-von Mises statistic: $W^2 = n\int_{-\infty}^{+\infty}\big(F_n(x) - F(x)\big)^2 dF(x)$

The Kolmogorov-Smirnov statistic is a supremum statistic while the Anderson-Darling statistic and the Cramér-von Mises statistic are both quadratic statistics.

An application of these tests to the first ETF from the sample, the SPDR S&P 500 ETF, shows that its excess returns over the benchmark index are not normally distributed. As it can be inferred from Table 3, the four tests return a p value smaller than 1%.

Table 3 : Tests for Normality (SPDR S&P 500 ETF)				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.899229	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.099434	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.620522	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4.04704	**Pr > A-Sq**	<0.0050

It is well known that, the smaller the p-value, the stronger the evidences against the null hypothesis. Therefore, there are rather strong evidences against the null hypothesis and the latter can be rejected with a risk of less than 1%. This implies that the excess returns are not normally distributed and that the skewness and the excess kurtosis of the distribution are statistically significantly different from 0. Hence, the standard deviation is not sufficient to measure the volatility of the tracking difference and the skewness and the kurtosis must be taken into account as well. This justifies the use of a more elaborated tracking error.

These results can be verified with a graphic analysis. The Q-Q plot is a statistical procedure that plots the quantiles of the studied distribution against the quantiles of the normal distribution. If the quantiles do not match, the studied variable is not distributed according to a normal distribution. The Q-Q plot for the distribution of excess returns for the SPDR S&P 500 ETF, as presented in Figure 2, shows further evidences against the null hypothesis of normality.

Figure 2: Q-Q plot for excess returns of the SPDR S&P 500

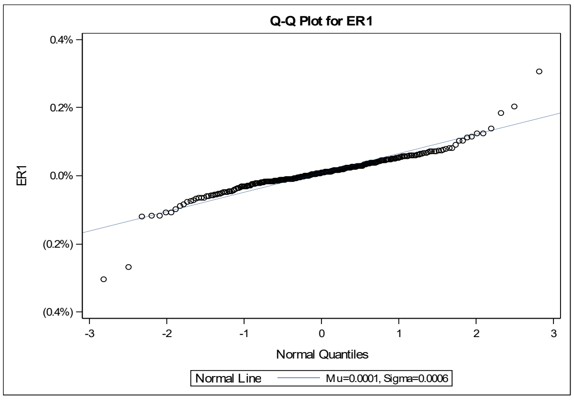


Table 4: Results of the normality tests for excess returns (2012)		
ETF	**Benchmark**	**Normality Test (H0: Excess returns are normally distributed)**
SPDR S&P 500 ETF	S&P 500	Rejection
iShares Core S&P 500	S&P 500	Rejection
Vanguard 500 ETF	S&P 500	Rejection
Horizons S&P 500 ETF	S&P 500	Rejection
iShares Russell 3000	Russell 3000	Rejection
SPDR Russell 3000 ETF	Russell 3000	Rejection
iShares Russell 2000	Russell 2000	Non rejection
iShares Nasdaq Biotechnology	NASDAQ Biotechnology	Rejection
PowerShares QQQ ETF	NASDAQ-100	Non rejection
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average	Rejection
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX	Rejection
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX	Non rejection
iShares S&P 100	S&P 100 INDEX	Rejection
iShares Russell 1000	RUSSELL 1000 INDEX	Non rejection
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX	Rejection
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX	Rejection
iShares PHLX Semiconductor	PHLX Semiconductor	Rejection
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH	Rejection
db x-trackers - DAX UCITS ETF (DR)	DAX	Rejection
Lyxor DAX (DR) UCITS ETF	DAX	Non rejection
ComStage - DAX TR UCITS ETF	DAX	Rejection
iShares US Financial Services	Dow Jones U.S. Financials Index	Non rejection
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX	Rejection
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite	Non rejection
Vanguard REIT ETF	MSCI US REIT INDEX	Non rejection
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50	Rejection
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50	Rejection
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50	Rejection
SSGA SPDR AEX EUR	AEX	Rejection
ISHARES AEX EUR	AEX	Rejection

Table 4 displays the results of the normality tests for each ETF from the sample. The tests were carried out using SAS Studio. They show that most ETFs do not have normally distributed excess returns. Therefore, a modified tracking error taking into account the skewness and the kurtosis of the distribution would be more relevant than the traditional tracking error since the latter assumes normality.

3.3. The new performance measure

In the literature review, I have identified some drawbacks for the measures that are currently used to assess the performance of ETFs. Therefore, in this section, I will develop a new performance measure. This new measure will later be applied on the sample and its quality will be assessed. The first step toward the creation of a new measure of performance will be the development of a more elaborated tracking error taking into account the skewness and the kurtosis of the distribution of excess returns. This step is justified by the results obtained in the previous subsection. Indeed, I have shown that the excess returns of ETFs over their benchmarks were not normally distributed. Therefore, the standard deviation alone cannot be used to measure the volatility of the tracking difference. The modified tracking error will be the tracking error multiplied by a number. This number will involve the quantile of the transformed distribution obtained with the Cornish-Fisher expansion to take into account the skewness and the kurtosis of the distribution.

The problem that arises when considering its incorporation in the measure is the role it will play. Will it act like a penalty? To answer this question, it is necessary to determine whether the measure will have to be maximized or minimized. Obviously, the rational investor will aim to minimize the modified tracking error[4]. That is because it is a measure of uncertainty.

It is well known that risk-averse investors have a preference for a positive skewness and an aversion for a high kurtosis. That is because a positive skewness can be interpreted as a higher probability of having higher than average returns as the distribution is skewed to the right. The kurtosis represents the probability of extreme returns, positive or negative. A distribution with a high kurtosis will have fat tails. The excess kurtosis used in the Cornish-Fisher expansion is equal to the kurtosis minus 3, which is the kurtosis of a normal distribution. Therefore, it is obvious that a risk-averse investor will require a higher risk premium for assets whose returns have a high kurtosis. Thus, it can be inferred that rational investors desire a positive skewness and a low kurtosis.

Cavenaile and Lejeune (2012) have shown that the quantity $z_{CF,\alpha}$ is an increasing function with respect to the skewness and a decreasing function with respect to the kurtosis for the confidence levels above 95.84%. The quantity $z_{CF,\alpha}$ is the α-quantile of the new distribution obtained with the Cornish-Fisher expansion. This means that risk-averse investors will look

[4] However, there could be situations when an investor would wish for a high tracking error. Indeed, if the tracking difference is slightly below 0, a high tracking error would translate into a higher probability for high excess returns.

for a high $z_{CF,\alpha}$ because it would mean a positive skewness and a low kurtosis. One could think that to minimize the MVaR, the investor should require the lowest $z_{CF,\alpha}$ as possible. It is not the case because $z_{CF,\alpha}$ is negative and, therefore, investors want the highest $z_{CF,\alpha}$ (the closest to zero).

The previous remark from Cavenaile and Lejeune (2012) suggests that the Cornish-Fisher expansion is applicable only if some conditions are met. The first one is that the confidence level should be above 95.84%. The second one is that, for every level of confidence above 95.84%, there is a minimum skewness required for the MVaR consistency. In order not to be too limited, I have chosen to work with the 96% confidence level because it is the one just above the 95.84% minimum level. This level of confidence allows me to consider distributions of excess returns with skewness down to -3.13. Taking these considerations into account, I can apply the Cornish-Fisher expansion to find a modified VaR or ES that is consistent with the preferences of a risk-averse investor.

The main problem in Roncalli's information ratio and ETF efficiency indicator is that they assume the normality of excess returns. In order to correct this problem, the modified expected shortfall can be used. Indeed, an involvement of the 96%-quantile of the Cornish fisher distribution based on the expected shortfall framework will be able to take into account the 3rd and 4th moments of the distribution.

The modified tracking error that I propose has the following equation:

$$MTE = \sigma(x|b) * \frac{ES_{CF,\alpha}}{ES_{\alpha}} \tag{10}$$

Because the VaR lacks some interesting properties such as subadditivity and convexity, its optimization is complex (Uryasev, 2000). The expected shortfall however possesses these characteristics. Moreover, as the expected shortfall is higher than the VaR by definition, minimizing the ES is equal to minimizing the VaR. It can be concluded from there that the risk-averse investor, who wants the quantity $z_{CF,\alpha}$ to be the closest to 0, will also want the quantity $ES_{CF,\alpha}$ close to 0. The modified tracking error characterized by Equation (10) based on the expected shortfall framework will therefore be a more consistent measure of risk and can even be minimized. These are the reasons why I have chosen the expected shortfall over the VaR framework. The second component of the new modified tracking error will be a ratio of the Cornish-Fisher expected shortfall to the expected shortfall under the normality assumption. This modified tracking error will take into account the skewness and the kurtosis of the distribution of the excess returns.

For example, if I want to rank the four following ETFs tracking the Euro Stoxx 50 for 2015 according to their tracking error, the ranking changes according to the tracking error that is used as it is shown in Table 5. Therefore, the modified tracking error does bring a contribution to the measurement of the performance of an ETF and gives a ranking different from the one obtained with the tracking error.

Table 5: Tracking error and modified tracking error

Ranking	1	2	3	4
Tracking error	Lyxor	Deka	Amundi	ComStage
Modified tracking error	Deka	Lyxor	Amundi	ComStage

The incorporation of the Cornish-Fisher expansion therefore allows me to propose a more elaborated tracking error that takes into account the skewness and the kurtosis of the distribution. This modified tracking error is justified by the non-normality of excess returns.

The second step toward the elaboration of the new performance measure will be to find a way to take into account the performance of the benchmark in order to be able to compare ETFs tracking different benchmarks. Intuitively, one can think that the importance of the benchmark performance is justified because the higher the returns of the benchmark over the risk-free rate of interest, the less important the tracking error. Moreover, an ETF tracking an index with very low excess returns over the risk-free rate of interest will need a higher tracking error because a high tracking error will increase the probability of having high returns for the ETF.

In order to take into account the performance of the benchmark, I use the performance measure proposed by Hübner (2012) and already introduced in this thesis. As stated before, the quantity $\frac{\sigma_P - \sigma_B}{\sigma_P}$ is the normalized idiosyncratic risk and is linked to the tracking error by Equation (4). Equation (3) can be rewritten as follows:

$$\pi_P = (1 - \iota_P)\alpha_P - \iota_P(R_B - R_f) \tag{11}$$

Merging Equations (11) and (4) allows me to introduce the tracking error in the performance measure.

$$=> \pi_P = (1 - (1 - \frac{\sqrt{\sigma_P^2 - \sigma_\varepsilon^2}}{\sigma_P}))\alpha_P - (1 - \frac{\sqrt{\sigma_P^2 - \sigma_\varepsilon^2}}{\sigma_P})(R_B - R_f)$$

$$=> \pi_P = \frac{\sqrt{\sigma_P^2 - \sigma_\varepsilon^2}}{\sigma_P}\alpha_P - (R_B - R_f)(1 - \frac{\sqrt{\sigma_P^2 - \sigma_\varepsilon^2}}{\sigma_P}) \quad (12)$$

This performance measure has an advantage over the information ratio because it allows me to take into account the excess return of the benchmark.

Now that I have a performance measure that takes into account the performance of the benchmark, I combine it with the modified tracking error in order to find the final measure of performance. If I use the same notation as in Hassine and Roncalli (2013), α_P is equivalent to $\mu(x|b)$ and measures the expected excess return of an ETF over its benchmark in the case of passive portfolio management. Replacing α_P by $\mu(x|b)$ in Equation (12) and replacing the tracking error by the modified tracking error from Equation (4) yields the new performance measure (π^A). The final performance measure for an ETF taking into account the 3rd and 4th moments of the distribution of excess return as well as the performance of the benchmark is the following:

$$\pi^A = \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\mu(x|b) - \left(R_B - R_f\right)\left(1 - \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\right) \quad (13)$$

Where

σ_P: Standard deviation of the returns of the ETF

$\mu(x|b)$: Expected excess return of the ETF over the benchmark

$R_B - R_f$: Excess returns of the benchmark over the risk-free rate of interest

MTE: Modified tracking error $\left(MTE = \sigma(x|b) * \frac{ES_{CF,\alpha}}{ES_\alpha}\right)$

3.4. Testing the measure

In order to test the new performance measure, I will assess its robustness in the measure of the persistence of the performance. In order to do so, I will study the performance persistence of the ETFs from one period to another. I will try to show that the measure is good at selecting the funds managed by managers with persistent superior skills. Therefore, I will test the stability of the ranking provided by the new performance measure and I will try to find dependencies between performances in two different periods. Indeed, the performance measure will be efficient if there is a positive correlation between the ranking for one period and the ranking for the next period, assuming that the funds from the sample are persistent in their performance. There are two kinds of methods that can be used to test the persistence in the performance. It can be done either with parametric methods, such as linear regressions, or with non-parametric methods. The latter is more interesting because it can be applied to more cases since it does not require any assumption regarding the distribution of the variables. I will use all these methods to determine whether the performance measure is persistent in its ranking. I will first consider four periods, one from each of the years 2012, 2013, 2014, and 2015. Then, I will consider two periods selected from two years between from 2012 and 2015 to test the performance persistence on the entire period of study. This will be done in a similar way as in Kuo and Mateus (2008). Finally, I will run the same tests on the information ratio to verify the performance persistence of the ETFs from the sample. If the results of the tests carried out using the information ratio provide evidences of persistence for the funds from the sample, it would mean that these funds are persistent in their performance. Therefore, the ability of the new performance measure to detect this persistence would assess its quality.

3.4.1. Parametric tests

Parametric tests are tests that rely on assumptions regarding the distribution of the data. Therefore, when performing parametric tests, one must always check the initial assumptions.

Regression analysis

The goal of a regression analysis is to try to find whether there exists a linear relationship between the variables from one period and the variables from the next period. Ferruz, Sarto and Vargas (2004) use the regression analysis to find statistical significance of the relationship between the performance in a given period and the performance in the next period. Their goal was to apply the regression analysis to find whether the performance in a period was a good indicator of the performance in the next period. They applied it to Spanish investment funds and they proposed the following model:

$$P_{p(t+1)} = \alpha_p + \beta_p P_{p(t)} + \varepsilon_p$$

Where $P_{p(t+1)}$ and $P_{p(t)}$ are the performances of the portfolio p for the period t and t+1. Therefore, if one can find significant value for β, then it can be concluded that there is a linear relationship between the performances in the two periods and that there is persistence.

3.4.2. Non-parametric tests

Non-parametric tests are tests that do not rely on any assumption regarding the distribution of the data. Here, in order to test the hypothesis of the absence of persistence, I will build contingency tables and run some non-parametric statistical tests as in Babalos, Caporale, Kostakis and Philippas (2008). The contingency tables will be based on the relative performances of the ETFs. ETFs with performances above the median will be categorized as winners (W) and ETFs with performances below the median will be referred to as losers (L). The idea will be to see whether winners in a period tend to stay winners in the next period. If so, it can be concluded that there is persistence in the ranking obtained with the new performance measure. The contingency tables will be constructed as follows:

Period 1/ Period 2	Winners	Losers
Winners	WW (“Hot hand”)	WL
Losers	LW	LL (“Cold hand”)

Previous studies have shown that it is possible to build contingency tables with more dimensions using the deciles instead of the median (see Carhart, 1997). Once the contingency tables are built, I will perform the following non-parametric tests to see whether there are statistically significantly more ETFs in the WW and LL cases.

Malkiel’s Z statistic

Malkiel (1995) developed a test statistic for repeat winners. If there is no persistence, the probability of a winning fund to win in the next period should be 1/2 because the performances from both periods are independent from each other. Therefore, one will find evidences of persistence if he can reject the null hypothesis that the probability p of an ETF to win two times in a row is equal to 1/2. The random variable Y representing the number of WW will obviously be distributed according to a binomial distribution B(n,p). The De Moivre–Laplace theorem tells us that when n, the number of observations, is sufficiently large, the random variable $Z = \frac{(Y-np)}{\sqrt{np(1-p)}}$ can be approximatively distributed according to a standard normal distribution. This is the reason why I chose a sample of 30 ETFs. Indeed, in practice, we can approximate a binomial law with a normal law if we have at least 30 observations. The Malkiel's test statistic is:

$$Z = \frac{(Y - np)}{\sqrt{np(1-p)}} \sim N(0,1)$$

Where Y is the number of persistent winners (WW) and n is the total number of observation (WW+WL). This test statistic can be applied to the 2x2 contingency table:

$$Z = \frac{WW - 0.5 * (WW + WL)}{\sqrt{0.5 * 0.5 * (WW + WL)}}$$

Cross-Product Ratio or Odds Ratio

The cross-product ratio is defined by Brown and Goetzmann (1995) as the odds ratio of the number of repeat performers to the number of funds that do not repeat. The odds ratio is therefore equal to:

$$OR = \frac{WW * LL}{WL * LW} = \frac{Repeat\ performers}{Do\ not\ repeat}$$

Here, the null hypothesis will be that the performance in the first period is uncorrelated to the performance in the second period. That would mean an odds ratio of one. Therefore, the null hypothesis will be rejected when the odds ratio is significantly different from one. The test statistic that is used is the following.

$$Z = \frac{\ln(OR)}{\sigma_{\ln(OR)}} \sim N(0,1) \quad where\ \sigma_{\ln(OR)} = \sqrt{\frac{1}{WW} + \frac{1}{WL} + \frac{1}{LW} + \frac{1}{LL}}$$

Again, this test statistic can be used only if the sample is sufficiently large (≥30).

Chi square test

Kahn and Rudd (1995) proposed to use a chi square test in order to test independency between the performances in two sub-periods. Here, the null hypothesis will be that there is independency. Therefore, the objective will be to find statistical evidences against this null hypothesis to prove that there is persistence. The test statistic used is the following:

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim {\chi^2}_{(r-1)(s-1)}$$

Where O_{ij} is the observed frequency of the ith row and the jth column. E_{ij} is the expected frequency of the ith row and the jth column. This test statistic is distributed according to a chi square distribution with (r-1)(s-1) degrees of freedom.

For a 2x2 contingency table, the chi square test statistic is equal to:

$$\chi^2 = \frac{(WW - D1)^2}{D1} + \frac{(WL - D2)^2}{D2} + \frac{(LW - D3)^2}{D3} + \frac{(LL - D4)^2}{D4}$$

Where

$$D1 = \frac{(WW + WL)(WW + LW)}{N}$$

$$D2 = \frac{(WW + WL)(WL + LL)}{N}$$

$$D3 = \frac{(LW + LL)(WW + LW)}{N}$$

$$D4 = \frac{(LW + LL)(WL + LL)}{N}$$

Spearman's correlation coefficient

Spearman's rank correlation coefficient is different from the methods that are described above since it does not require building contingency tables. This test is a rank order test. It will be used to determine whether there exists a positive correlation between the ranking in period one and the ranking in period two. There will be a positive correlation if the coefficient is close to one. This would imply that there is performance persistence over the two periods. Spearman's rank correlation coefficient differs from the Pearson product-moment correlation coefficient since it works on ranked data (Kuo & Mateus, 2008).

If R1 represents the rank of one ETF in the first period and R2 represents the rank of the same ETF in the second period, the Spearman rank correlation coefficient is defined as follows:

$$r_s = 1 - \frac{6}{n(n-1)^2} \sum_{i=1}^{n} (R1 - R2)^2$$

And the test statistic used to measure the statistical significance of the Spearman rank correlation is the following: $Z = r_s\sqrt{n-1}$[8] $\sim N(0,1)$

4. Results

4.1. An example: Application on ETFs tracking the S&P 500 in 2015

Now that the new performance measure and the methodology have been presented, I will test it on actual ETFs and present the detailed computations. In this subsection, I will apply the new performance measure to the ETFs from the sample tracking the S&P 500 in order to rank them from best to worst. The ETFs that will be used to test this measure are the SPDR S&P 500 ETF, the iShares Core S&P 500, the Vanguard 500 ETF and the Horizons S&P 500 ETF. The period of study is the year 2015. The data for these ETFs was retrieved from Yahoo Finance.

The first step to build the new performance measure is to compute the excess returns over the benchmark for all ETFs. In order for the returns to be comparable, I first need to express them in daily returns. The daily returns were computed using the following common formula:

$$R_t = \frac{Adusted\ closing\ price_t - Adjusted\ closing\ price_{t-1}}{Adjusted\ closing\ price_{t-1}}$$

In order to use the new performance measure, I need to know some information about the distribution of the excess returns. The mean, standard deviation, skewness and excess kurtosis for the excess returns of the SPDR S&P 500 ETF are presented in Table 6.

Table 6: Descriptive statistics SPDR S&P 500	
Mean	7.72683E-05
Standard Deviation	0.000450559
Skewness	-0.542803512
Excess Kurtosis	6.235394864

From the table above, it can be inferred that the tracking error of this ETF is 0.000451. This is the first element that is required to compute the modified tracking error. The next element will be the component of the expected shortfall, which is given by the following formula:

$$CVaR_{1-\alpha} = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}}e^{-\frac{z_\alpha^2}{2}}\ where\ z_\alpha = -1.7507\ for\ a\ level\ of\ confidence\ (1-\alpha) = 96\%$$

In the above formula, the level of confidence (1-α) is 96%. As explained earlier, this level of confidence allows me to consider skewness down to -3.13. Thus, the distribution of excess returns for the SPDR S&P 500 ETF fulfils the required conditions for the Cornish-Fisher expansion to be applied. The formula used to get the component of the expected shortfall modified using the Cornish-Fisher expansion is the formula from Maillard (2012):

$$MCVaR_{1-\alpha} = CVaR\left[1 + z_\alpha \frac{S}{6} + (1 - 2z_\alpha^2)\frac{S^2}{36} + (-1 + z_\alpha^2)\frac{K}{24}\right]$$

An application of the above formula gives the expected shortfall components needed in order to build the modified tracking error. The results are presented in Table 7:

Table 7: Modified tracking error SPDR S&P 500	
$ES_{CF,\alpha}$	3.560859804
$ES_{,\alpha}$	2.154344351
$\sigma(x\|b)$	0.000450559
MTE	0.000744716

Now that the modified tracking error is computed, I have to take into account the performance of the benchmark. In order to do that, I need the risk-free rate of interest for the period covered. The risk-free investment does not really exist but in practice, the 3-month US Treasury Bill[5] is usually chosen as risk-free rate of interest as in Hübner (2007). The annual risk-free rate can be found using the following formula where $R_{f12}$is the annual risk-free rate and R_{f3} the 3-month risk free rate:

$$R_{f12} = (1 + R_{f3})^4 - 1$$

Therefore, the benchmark generates an excess performance over the risk-free rate of:

$$R_b - R_f = -0.00807$$

The final step will be to input all the information above in the new performance measure:

$$\pi^A = \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\mu(x|b) - (R_B - R_f)\left(1 - \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\right)$$

The results and ranking are presented in Table 8:

Table 8: Ranking and performance of ETFs tracking the S&P 500				
Ranking	1	2	3	4
Rank	Horizons	SPDR	iShares	Vanguard
π^A	0.2261%	0.0101%	0.0095%	0.0092%

[5] The data for the 3-month US Treasury Bill was retrieved from the website of the U.S. Department of the Treasury. https://www.treasury.gov

The new performance measure introduced above allows us to provide a ranking for ETFs. Through this example, I have shown that this measure takes into account the 3^{rd} and 4^{th} moments of the excess returns distribution. Moreover, the performance of the benchmark over the risk-free rate also brings a contribution. Although it might not be clear in this example since the EFTs all track the same benchmark. Therefore, the next step of the analysis of this performance measure will be to test ETFs tracking different benchmarks.

4.2. Results for the sample

4.2.1. Results

In this subsection, I apply the new performance measure to the sample that has been presented above. Then, the ETFs will be ranked accordingly and this ranking will be compared with the ranking obtained with the information ratio to try to see the contribution that the new performance measure brings. Finally, a rolling window analysis will be presented and I will study the evolution of the performance measure when using 1 month, 3 month or 6 month rolling windows for a few ETFs.

I have already provided a detailed example on the way to compute the performance measure for the SPDR S&P 500 ETF. Table 9 below lists the ETFs from the sample and their performance in 2015 based on the new performance measure developed in this thesis. Their information ratio, tracking error and modified tracking error are also presented.

Table 9: Results (2015)

ETF	Benchmark	New Performance Measure	Information Ratio	Tracking Error	Modified Tracking Error
SPDR S&P 500 ETF	S&P 500	0.0101%	17.1495%	0.0451%	0.0745%
iShares Core S&P 500	S&P 500	0.0095%	16.1909%	0.0494%	0.0601%
Vanguard 500 ETF	S&P 500	0.0092%	17.9302%	0.0455%	0.0506%
Horizons S&P 500 ETF	S&P 500	0.2261%	13.0263%	0.5933%	0.5923%
iShares Russell 3000	Russell 3000	0.0088%	16.7042%	0.0428%	0.0437%
SPDR Russell 3000 ETF	Russell 3000	0.2497%	1.5818%	0.3783%	0.4707%
iShares Russell 2000	Russell 2000	0.0241%	6.5909%	0.0771%	0.0843%
iShares Nasdaq Biotechnology	NASDAQ Biotechnology	-0.0590%	0.8698%	0.1283%	0.1836%
PowerShares QQQ ETF	NASDAQ-100	-0.0072%	6.7290%	0.0548%	0.0575%
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average	0.0112%	21.1131%	0.0440%	0.0398%
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX	0.0102%	9.1489%	0.0544%	0.0497%
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX	0.0241%	9.4959%	0.0587%	0.0925%
iShares S&P 100	S&P 100 INDEX	0.0077%	17.4465%	0.0462%	0.0525%
iShares Russell 1000	RUSSELL 1000 INDEX	0.0088%	17.0274%	0.0437%	0.0470%
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX	0.4220%	1.7736%	0.4926%	0.7253%
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX	0.0259%	5.1340%	0.1331%	0.1711%
iShares PHLX Semiconductor	PHLX Semiconductor	0.0149%	6.8374%	0.0793%	0.1050%
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH	-0.4166%	0.6858%	0.2350%	0.3752%
db x-trackers - DAX UCITS ETF (DR)	DAX	-0.0116%	-1.0003%	0.0759%	0.0761%
Lyxor DAX (DR) UCITS ETF	DAX	-0.0123%	-2.1936%	0.0736%	0.0757%
ComStage - DAX TR UCITS ETF	DAX	-0.0141%	-0.6666%	0.0759%	0.0848%
iShares US Financial Services	Dow Jones U.S. Financials Index	0.0825%	2.3948%	0.3077%	0.3222%
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX	-0.0125%	0.2898%	0.0642%	0.0848%
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite	-0.1371%	1.9158%	0.1836%	0.2330%
Vanguard REIT ETF	MSCI US REIT INDEX	0.0921%	14.7690%	0.1027%	0.2772%
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50	0.0080%	11.8024%	0.1067%	0.1161%
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50	0.0004%	12.8931%	0.0826%	0.1772%
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50	0.0058%	10.4553%	0.1335%	0.1567%
SSGA SPDR AEX EUR	AEX	0.0014%	14.4946%	0.0666%	0.0850%
ISHARES AEX EUR	AEX	-0.0344%	10.5027%	0.0751%	0.1926%

From the above table, it is easy to create a ranking for ETFs according to their respective performances.

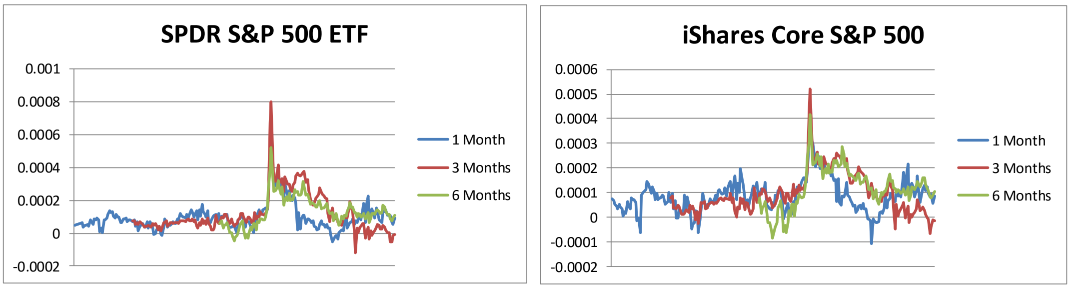
Table 10: Rankings obtained with new performance measure and information ratio (2015)					
ETF	**Benchmark**	**New Performance Measure**	**Rank**	**Information Ratio**	**Rank**
SPDR S&P 500 ETF	S&P 500	0.0101%	12	17.1495%	4
iShares Core S&P 500	S&P 500	0.0095%	13	16.1909%	7
Vanguard 500 ETF	S&P 500	0.0092%	14	17.9302%	2
Horizons S&P 500 ETF	S&P 500	0.2261%	3	13.0263%	10
iShares Russell 3000	Russell 3000	0.0088%	16	16.7042%	6
SPDR Russell 3000 ETF	Russell 3000	0.2497%	2	1.5818%	24
iShares Russell 2000	Russell 2000	0.0241%	8	6.5909%	19
iShares Nasdaq Biotechnology	NASDAQ Biotechnology	-0.0590%	28	0.8698%	25
PowerShares QQQ ETF	NASDAQ-100	-0.0072%	22	6.7290%	18
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average	0.0112%	10	21.1131%	1
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX	0.0102%	11	9.1489%	16
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX	0.0241%	7	9.4959%	15
iShares S&P 100	S&P 100 INDEX	0.0077%	18	17.4465%	3
iShares Russell 1000	RUSSELL 1000 INDEX	0.0088%	15	17.0274%	5
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX	0.4220%	1	1.7736%	23
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX	0.0259%	6	5.1340%	20
iShares PHLX Semiconductor	PHLX Semiconductor	0.0149%	9	6.8374%	17
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH	-0.4166%	30	0.6858%	26
db x-trackers - DAX UCITS ETF (DR)	DAX	-0.0116%	23	-1.0003%	29
Lyxor DAX (DR) UCITS ETF	DAX	-0.0123%	24	-2.1936%	30
ComStage - DAX TR UCITS ETF	DAX	-0.0141%	26	-0.6666%	28
iShares US Financial Services	Dow Jones U.S. Financials Index	0.0825%	5	2.3948%	21
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX	-0.0125%	25	0.2898%	27
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite	-0.1371%	29	1.9158%	22
Vanguard REIT ETF	MSCI US REIT INDEX	0.0921%	4	14.7690%	8
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50	0.0080%	17	11.8024%	12
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50	0.0004%	21	12.8931%	11
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50	0.0058%	19	10.4553%	14
SSGA SPDR AEX EUR	AEX	0.0014%	20	14.4946%	9
ISHARES AEX EUR	AEX	-0.0344%	27	10.5027%	13

The ranking from Table 10 highlights the contribution of the return of the benchmark over the risk-free rate because the ETFs that are tracking the same benchmark are usually not far from each other in terms of performance. Indeed, we can see that the ETFs tracking the S&P 500, the DAX index and the Euro Stoxx 50 are close to each other. Moreover, this new performance measure gives a ranking that is completely different from the ranking obtained with the information ratio.

4.2.2. Rolling window analysis

In order to obtain the previous results, I considered the performance measure for a period of one year. However, it is possible to apply it to shorter periods. In this subsection, I will consider the performance measure over 1 month, 3 month and 6 month periods. The risk free rate will be the 3 month US Treasury bill compounded for 1 month, 3 months and 6 months. The mean, the standard deviation, the skewness and the kurtosis of excess returns will be computed in the same way as for the annual performance measure. This will allow me to perform a moving analysis with different rolling windows. The goal will be to compare these different windows in order to see whether the results vary depending on the width of the window. Moreover, this analysis will be a first step to assess the stability of the performance measure over time. A moving analysis of all the ETFs in the sample would be too heavy. Therefore, we will consider only a few ETFs for the year 2015. That is because the objective here is not to determine whether the performance measure is persistent and effective, this will be dealt with in the next section, but rather to graphically analyze the behavior of the measure when different rolling windows are considered. Figure 3 shows the performance of the SPDR S&P 500 ETF and the iShares Core S&P 500.

Figure 3: Rolling window analysis S&P 500



The first thing that can be inferred from these graphs is that the performance measure seems to evolve in the same direction whatever the width of the window. Indeed, the 1 month window, the 3 month window and the 6 month window all behave in the same way in 2015. This short analysis could be the first step toward the assessment of the stability and the efficiency of the new performance measure.

The second thing that can be inferred from these graphs is that the performances of the benchmarks appear to have a relatively big impact on the measure. That is because the graphs evolve in a similar way for both ETFs. This result is coherent with the ranking for 2015 presented earlier. Indeed, it was shown that the ETFs tracking the same benchmark were usually ranked not far from each other. In order to further analyze this result, I will consider two others ETFs tracking the Euro Stoxx 50 and two tracking the DAX index.

Figure 4: Rolling window analysis Euro Stoxx 50

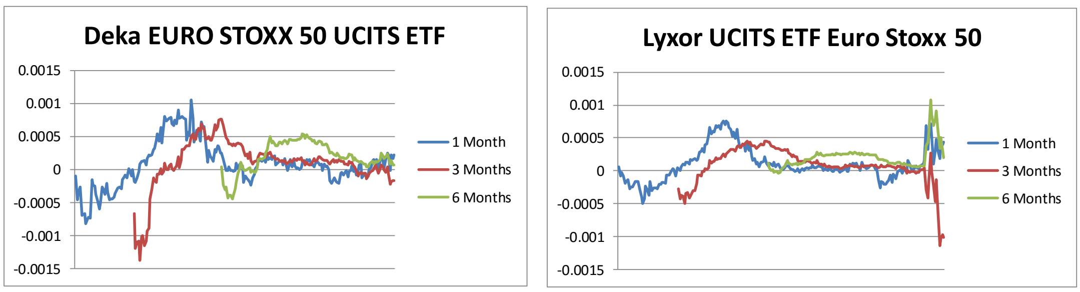
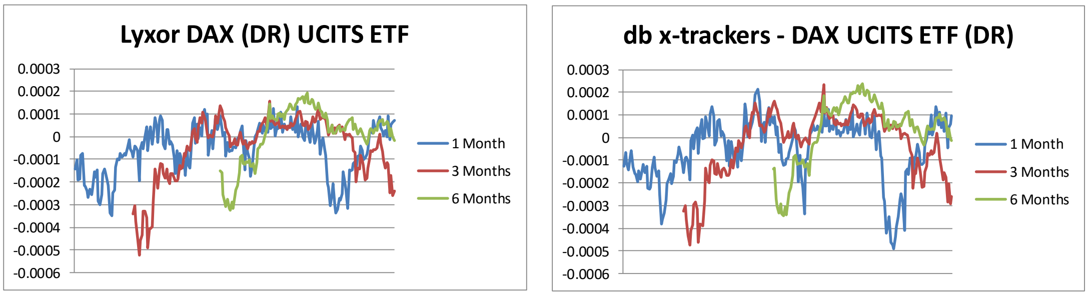


Figure 5: Rolling window analysis DAX



Figures 4 and 5 give further evidence of the importance of the benchmark in the new performance measure. Moreover, we can still observe a similarity between the results from different rolling windows. As for the ETFs tracking the S&P 500, the measure evolves in the same way whatever the width of the window. This last result is important as it would mean that the measure is flexible and can be used to calculate the performance of ETFs for time periods that are longer or shorter than one year.

5. Results - The efficiency of the measure

In order to assess the efficiency of the new performance measure, I test its robustness in the measurement of performance persistence. In order to do that, I considered the performances of ETFs from the sample during four periods (2012, 2013, 2014 and 2015). In this section, I will present the results obtained from the statistical tests described in Section 3. The results from those tests will indicate whether there is persistence from one period to the next. As stated above, the objective will be to find relationships and correlations between the performances or the rankings for the different periods. A positive correlation between the performances in two successive periods would indicate that the measure is efficient for this period and can highlight the skills of the superior managers, assuming that the ETFs are persistent in their performance. In the following analysis, the performances of the ETFs for 2012, 2013, 2014 and 2015 will be given by the variables Y1, Y2, Y3 and Y4. A summary of the annual performances for the ETFs in the sample is presented in Table 11.

Table 11: Results

ETF	Benchmark	2012	2013	2014	2015
SPDR S&P 500 ETF	S&P 500	-0.1033%	-0.1687%	-0.0141%	0.0101%
iShares Core S&P 500	S&P 500	-0.2786%	-0.2969%	-0.0250%	0.0095%
Vanguard 500 ETF	S&P 500	-0.0550%	-0.9454%	-0.4248%	0.0092%
Horizons S&P 500 ETF	S&P 500	-5.0384%	-7.6824%	-2.4348%	0.2261%
iShares Russell 3000	Russell 3000	-0.0327%	-0.7892%	-0.0075%	0.0088%
SPDR Russell 3000 ETF	Russell 3000	-1.9192%	-10.2855%	-2.0075%	0.2497%
iShares Russell 2000	Russell 2000	-0.0793%	-0.1734%	-0.0037%	0.0241%
iShares Nasdaq Biotechnology	NASDAQ Biotechnology	-0.6967%	-0.4313%	-0.0670%	-0.0590%
PowerShares QQQ ETF	NASDAQ-100	-0.0214%	-0.0955%	-0.0283%	-0.0072%
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average	-0.0219%	-0.1062%	-0.0108%	0.0112%
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX	-0.0802%	-0.3507%	-0.0278%	0.0102%
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX	-0.0326%	-0.1534%	-0.0204%	0.0241%
iShares S&P 100	S&P 100 INDEX	-0.0269%	-0.1073%	-0.0166%	0.0077%
iShares Russell 1000	RUSSELL 1000 INDEX	-0.0420%	-1.0028%	-0.0182%	0.0088%
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX	-6.1323%	-16.1322%	-3.7212%	0.4220%
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX	-2.7610%	-1.1351%	-0.3106%	0.0259%
iShares PHLX Semiconductor	PHLX Semiconductor	-0.0424%	-0.0862%	-0.0750%	0.0149%
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH	-0.4899%	-5.7528%	-2.3849%	-0.4166%
db x-trackers - DAX UCITS ETF (DR)	DAX	-0.0701%	-0.0495%	-0.0084%	-0.0116%
Lyxor DAX (DR) UCITS ETF	DAX	-0.0525%	-0.0505%	-0.0081%	-0.0123%
ComStage - DAX TR UCITS ETF	DAX	-0.0614%	-0.0515%	-0.0072%	-0.0141%
iShares US Financial Services	Dow Jones U.S. Financials Index	-0.6849%	-0.8334%	-0.3711%	0.0825%
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX	-6.9515%	-0.1152%	-7.2470%	-0.0125%
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite	-0.5017%	-5.1455%	-0.2115%	-0.1371%
Vanguard REIT ETF	MSCI US REIT INDEX	-0.0562%	0.0238%	-0.3959%	0.0921%
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50	-0.2955%	-0.0459%	0.0058%	0.0080%
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50	-0.0254%	-0.0311%	0.0054%	0.0004%
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50	-0.0310%	-0.0462%	0.0012%	0.0058%
SSGA SPDR AEX EUR	AEX	-0.1232%	-0.3474%	-0.0525%	0.0014%
ISHARES AEX EUR	AEX	-0.3560%	-0.1195%	-0.0357%	-0.0344%

5.1. One-year periods

5.1.1. Parametric test

Regression analysis

Here, the objective is to check whether there is a linear relationship between the performance of an ETF in the first period (Y1) and its performance in the second period (Y2). The correlation coefficient for the two variables is computed in Table 12 using SAS Studio.

Table 12: Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	Y2	**Y1**
Y2 Y2	1.00000	0.60455 0.0004
Y1 Y1	0.60455 0.0004	1.00000

The first line indicates the Pearson correlation coefficient between both variables and the second line returns the p value from a hypothesis test where the null hypothesis states that the correlation coefficient is equal to zero and that the two variables are uncorrelated. Since the p value is smaller than 1%, the null hypothesis can be rejected with a risk of less than 1%. This means that there exists a positive linear relationship between Y1 and Y2. The analysis of variance presented in Table 13 also returns a p value of 0.004 meaning that, when the variable Y2 is added, it improves the quality of prediction of the variable Y1.

Table 13: Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	0.01454	0.01454	16.13	0.0004
Error	28	0.02525	0.00090173		
Corrected Total	29	0.03979			

Estimates for the parameters are given in the following table:

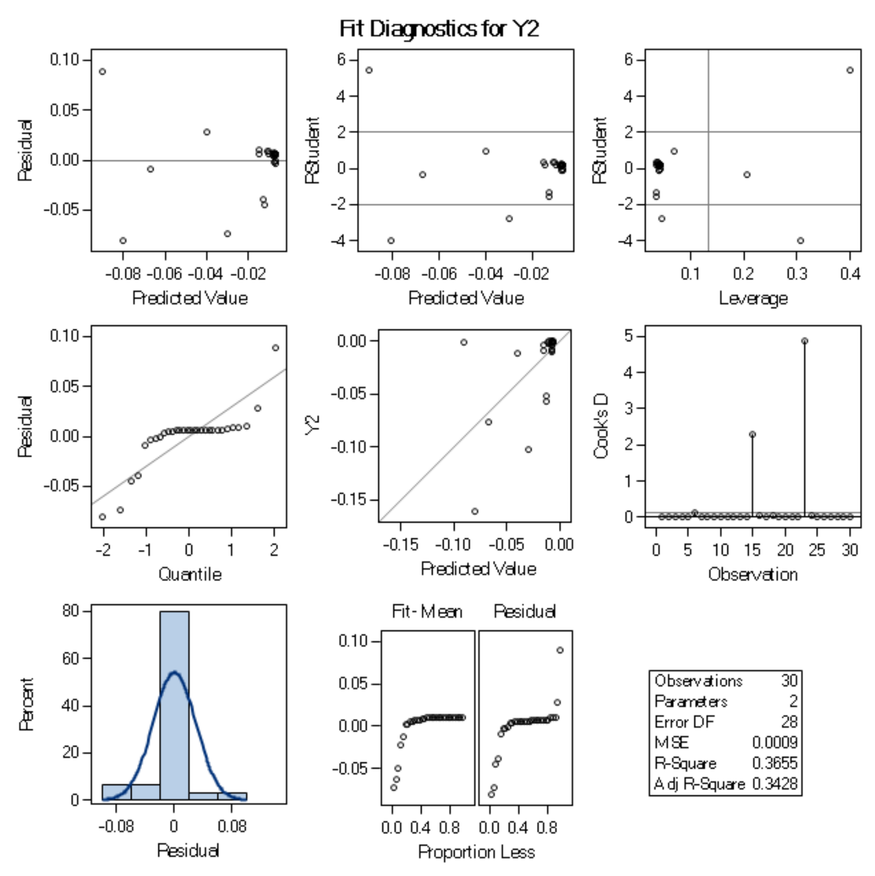
Table 14: Parameter Estimates								
Variable	**Label**	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**95% Confidence Limits**	
Intercept	Intercept	1	-0.00662	0.00612	-1.08	0.2880	-0.01915	0.00590
Y1	Y1	1	1.20590	0.30027	4.02	0.0004	0.59081	1.82098

Here, it cannot be proven that the alpha, the intercept parameter, is significantly different from 0. However, since the p value for the slope parameter is smaller than 1%, there exists a significant linear relationship between Y1 and Y2. The linear model has the following equation:

$$Y2 = 1.20590 * Y1$$

Moreover, the analysis of the residuals presented in Table 15 gives the results for the R-square and the adjusted R-square. The R-square is an indication of the fitness of the model. It measures the percentage of the variability in the dependent variable explained by the model. The adjusted R-square follows the same principle but gives a penalty if we use a too large number of parameters. Therefore, this model explains 34.28% of the variability in Y2.

Table 15: Analysis of the residuals



This regression analysis of Y1 and Y2 shows that there is a positive relationship between the performances of ETFs in 2012 and their performances in 2013. This would mean that the new measure of performance gives a rather stable ranking from 2012 to 2013.

The last thing that needs to be done is to check the initial assumption. Indeed, a regression analysis assumes that the residuals are normally distributed. Therefore, in order to check that, I will run the normality tests that I have developed in a previous section. The results are presented in Table 16.

Table 16: Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	**W**	0.712007	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.29421	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.872909	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	4.202961	**Pr > A-Sq**	<0.0050

All the tests reject the hypothesis of normality with a level of confidence above 99%. Since the required conditions to perform a regression analysis are not met, it cannot be rigorously concluded that the measure gives a persistent ranking from 2012 to 2013 based on this analysis. This invalids the previous results.

Table 17 lists the results for the regression analysis over the period of study.

Table 17: Regression Analysis

Period	Pearson Correlation Coefficient	P value	Analysis of Variance (test statistic)	P value	β	P value	adj. R-square	Test for normality assumption
2012-2013	0.60455	0.0004	16.13	0.0004	1.2059	0.0004	34.28%	Rejection
2013-2014	0.48932	0.0061	8.81	0.0061	0.20437	0.0061	21.23%	Rejection
2014-2015	-0.20787	0.2703	1.26	0.2703	-0.0176	0.2703	0.90%	Rejection

According to the above table, the new performance measure is efficient in the measurement of the performance persistence in the first two periods of study. However, since the normality assumption is violated for every period, the regression analysis cannot be used.

5.1.2. Non-parametric tests

As a result, I have to rely on non-parametric tests to test the performance persistence.

In order to be able to use the non-parametric tests described in the previous section, I build the winners/losers contingency table for each period.

Table 18: Winners/losers contingency tables

2013 2012	W	L
W	11	4
L	4	11

2014 2013	W	L
W	10	5
L	5	10

2015 2014	W	L
W	7	8
L	8	7

A first look at the contingency tables allows us to see that there seems to be performance persistence for the first two periods.

Malkiel's Z statistic

Table 19 illustrates the results for the Malkiel's test.

Table 19: Malkiel's test		
Period	**Malkiel's test statistic**	**P value**
Formula	$Z = \frac{WW - 0.5 * (WW + WL)}{\sqrt{0.5 * 0.5 * (WW + WL)}}$	
2012-2013	1.8074	0.0354
2013-2014	1.2910	0.0984
2014-2015	-0.2582	0.6019

Since non-parametric tests do not rely on any assumption regarding the distribution of the data, the normality of the residuals is not required. From this table, it can be concluded that there is persistence of the performance from 2012 to 2013. Indeed, the null hypothesis can be rejected with a risk of less than 10%. Therefore, I reject the fact that the performance in 2013 is independent from the performance in 2012. However, it cannot be rejected for the other two periods.

Cross-Product Ratio or Odds Ratio

The cross-product ratio test gives the following results:

Table 20: Cross-Product Ratio or Odds Ratio				
Period	**Odd Ratio**	**σlog(OR)**	**Test statistic**	**P value**
Formula	$OR = \frac{WW * LL}{WL * LW}$	$\sigma_{\ln(OR)} = \sqrt{\frac{1}{WW} + \frac{1}{WL} + \frac{1}{LW} + \frac{1}{LL}}$	$Z = \frac{\ln(OR)}{\sigma_{\ln(OR)}}$	
2012-2013	7.5625	0.8257	2.4502	0.0071
2013-2014	4	0.7746	1.7897	0.0368
2014-2015	0.7656	0.7319	-0.3649	0.6424

The results obtained with the cross-product ratio are somehow different from those obtained with Malkiel's test. Since the null hypothesis is the absence of persistence, it can be inferred that the performance in 2013 is correlated with the performance in 2012, the same goes for the 2013-2014 period. According to this test, there is persistence of the performance in the first two periods.

Chi square test

The chi square test statistics and p values for the different periods are presented in the tables below.

Table 21 : Chi square test results 2012-2013			
Statistic	**DF**	**Value**	**Prob**
Chi-Square	1	6.5333	0.0106
Likelihood Ratio Chi-Square	1	6.7939	0.0091
Continuity Adj. Chi-Square	1	4.8000	0.0285
Mantel-Haenszel Chi-Square	1	6.3156	0.0120

Table 22 : Chi square test results 2013-2014			
Statistic	**DF**	**Value**	**Prob**
Chi-Square	1	3.3333	0.0679
Likelihood Ratio Chi-Square	1	3.3980	0.0653
Continuity Adj. Chi-Square	1	2.1333	0.1441
Mantel-Haenszel Chi-Square	1	3.2222	0.0726

Table 23 : Chi square test results 2014-2015			
Statistic	**DF**	**Value**	**Prob**
Chi-Square	1	0.1333	0.7150
Likelihood Ratio Chi-Square	1	0.1334	0.7149
Continuity Adj. Chi-Square	1	0.0000	1.0000
Mantel-Haenszel Chi-Square	1	0.1289	0.7196

The above tests give the same results as Malkiel's test since the null hypothesis is rejected for the first period (2012-2013). However, there is not enough evidence against the null hypothesis of independency for the last 2 periods.

Spearman's correlation coefficient

Spearman's rank correlation coefficient does not rely on contingency tables. Here, under the null hypothesis, the ranks from the first period are uncorrelated with the ranks from the second period. Therefore, if there is not enough evidence against the null hypothesis, the hypothesis of the absence of persistence cannot be rejected. The results of Spearman's rank correlation test are presented in Tables 24, 25 and 26.

Table 24: Spearman's correlation coefficient 2012-2013		
Spearman Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	Y1	**Y2**
Y1 Y1	1.00000	0.58487 0.0007
Y2 Y2	0.58487 0.0007	1.00000

Table 25: Spearman's correlation coefficient 2013-2014		
Spearman Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	Y2	**Y3**
Y2 Y2	1.00000	0.62269 0.0002
Y3 Y3	0.62269 0.0002	1.00000

Table 26: Spearman's correlation coefficient 2014-2015		
Spearman Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	Y3	**Y4**
Y3 Y3	1.00000	-0.24271 0.1962
Y4 Y4	-0.24271 0.1962	1.00000

This test provides strong evidences against the null hypothesis in the two first cases since the p values are below 1%. Therefore, the null hypothesis can be rejected with a risk of less than 1%.

To sum up, the non-parametric tests that have been performed allow me to conclude that, from 2012 to 2013, there is performance persistence when the new performance measure is used. The results for the second period were less powerful since the tests did not all give the same results. However, the results are closer from persistence than from independency. Finally, there seems to be no persistence from 2014 to 2015 with the new performance measure. However, since the results are the non-rejection of the null hypothesis, it cannot be concluded that this null hypothesis is true. Table 27 summarizes the results for the tests of persistence.

Here, I considered four one-year periods and I have studied the robustness in the measure of performance persistence from one period to another. In the next subsection, I will consider two periods of two years and perform the same kind of analysis as it was done in Kuo and Mateus (2008).

Table 27: Summary Table for the tests of persistence

Period	Malkiel's test statistic	P value	Conclusion	Cross-Product Ratio's test statistic	P value	Conclusion	Chi Square statistic	P value	Conclusion	Spearman's Rho	P value	Conclusion
2012-2013	1.8074	0.0354	Persistence	2.4502	0.0071	Persistence	6.5333	0.0106	Persistence	0.58487	0.0007	Persistence
2013-2014	1.2910	0.0984	Cannot conclude	1.7897	0.0368	Persistence	3.3333	0.0679	Cannot conclude	0.62269	0.0002	Persistence
2014-2015	-0.2582	0.6019	Cannot conclude	-0.3649	0.6424	Cannot conclude	0.1333	0.715	Cannot conclude	-0.24271	0.1962	Cannot conclude

5.2. Two-year periods

I will now do the same analysis for 2 periods of 24 months. The objective is to be able to assess the efficiency of the performance measure over the entire period of study (4 years). The results for the two 24 month periods are presented below.

Table 28: Results for two-year periods			
ETF	**Benchmark**	**Period 1**	**Period 2**
SPDR S&P 500 ETF	S&P 500	-0.3394%	-0.0218%
iShares Core S&P 500	S&P 500	-0.8385%	-0.0173%
Vanguard 500 ETF	S&P 500	-0.9078%	-0.2507%
Horizons S&P 500 ETF	S&P 500	-16.2961%	-2.2853%
iShares Russell 3000	Russell 3000	-0.6371%	-0.0042%
SPDR Russell 3000 ETF	Russell 3000	-10.0321%	-1.6208%
iShares Russell 2000	Russell 2000	-0.2707%	0.0118%
iShares Nasdaq Biotechnology	NASDAQ Biotechnology	-1.5382%	-0.2056%
PowerShares QQQ ETF	NASDAQ-100	-0.1106%	-0.0405%
SPDR Dow Jones Industrial Average ETF	Dow Jones Industrial Average	-0.1399%	0.0010%
SPDR S&P MidCap 400 ETF	S&P MID CAP 400 INDEX	-0.4144%	-0.0068%
iShares Core S&P Mid-Cap	S&P MID CAP 400 INDEX	-0.1806%	-0.0126%
iShares S&P 100	S&P 100 INDEX	-0.1347%	-0.0115%
iShares Russell 1000	RUSSELL 1000 INDEX	-0.8145%	-0.0094%
SPDR Russell 1000 ETF	RUSSELL 1000 INDEX	-22.4222%	-3.5703%
Vanguard Russell 1000 ETF	RUSSELL 1000 INDEX	-8.4597%	-0.2125%
iShares PHLX Semiconductor	PHLX Semiconductor	-0.3448%	-0.0677%
SPDR Morgan Stanley Technology ETF	MORGAN STANLEY TECH	-4.4919%	-2.6581%
db x-trackers - DAX UCITS ETF (DR)	DAX	-0.1659%	-0.0275%
Lyxor DAX (DR) UCITS ETF	DAX	-0.1320%	-0.0270%
ComStage - DAX TR UCITS ETF	DAX	-11.7545%	-0.0308%
iShares US Financial Services	Dow Jones U.S. Financials Index	-1.8182%	-0.3378%
First Trust NYSE Arca Biotech ETF	NYSE ARCA BIOTECH INDEX	-15.3374%	-8.5208%
Fidelity Nasdaq Composite Tr Stk ETF	NASDAQ Composite	-4.2310%	-0.4595%
Vanguard REIT ETF	MSCI US REIT INDEX	-0.0490%	-0.5929%
Deka EURO STOXX 50 UCITS ETF	Euro Stoxx 50	-0.5538%	-0.0126%
Lyxor UCITS ETF Euro Stoxx 50	Euro Stoxx 50	-0.0881%	-0.0301%
iShares Core EURO STOXX 50 UCITS ETF	Euro Stoxx 50	-0.1159%	-0.0269%
SSGA SPDR AEX EUR	AEX	-0.4642%	-0.0312%
ISHARES AEX EUR	AEX	-0.9757%	-0.0996%

5.2.1. Parametric test

Regression analysis

I perform a regression analysis in the same ways as I did in the previous subsection. Table 29 shows the results of this analysis.

Table 29: Regression Analysis								
Period	**Pearson Correlation Coefficient**	**P value**	**Analysis of Variance (test statistic)**	**P value**	**β**	**P value**	**adj. R-square**	**Test for normality assumption**
2012/2013 - 2014/2015	0.7088	<.0001	28.27	<.0001	0.2079	<.0001	48.46%	Rejection

From the regression analysis, it can be deduced that there exists a linear relationship between the two variables representing the performances of ETFs in the two periods of study. Indeed, the Pearson correlation coefficient is close to one and the slope parameter is significantly different from zero. The adjusted R-square is even better than before. All these elements would indicate that the performances in the two periods are positively correlated. However, since the normality assumption is rejected, those results cannot be used to test the performance persistence. Therefore, I will rely on non-parametric tests.

5.2.2. Non-parametric tests

In the case of one-year periods, I have performed four non-parametric tests. I will do the same tests for two-year periods in order to evaluate the performance persistence with the new performance measure. Firstly, I build a winners/losers contingency table.

Table 30: Winners/losers contingency table

2014/2015 2012/2013	W	L
W	11	4
L	4	11

Secondly, I apply the well-known statistical tests that were already used in the previous subsection. The non-parametric tests all indicate that there is persistence of the performance from the first period (2012-2013) to the second (2014-2015). The results are presented in Table 31.

Table 31: Summary Table

Period	Malkiel's test statistic	P value	Conclusion	Cross-Product Ratio's test statistic	P value	Conclusion	Chi Square statistic	P value	Conclusion	Spearman's Rho	P value	Conclusion
2012/2013 - 2014/2015	1.8074	0.0354	Persistence	2.4502	0.0071	Persistence	6.5333	0.0106	Persistence	0.5746	0.0009	Persistence

5.3. Interpretations

To sum up, I have shown that the new performance measure is rather robust in the measure of performance persistence from one period to the next in the case of two-year periods. This means that winners tend to stay winners and losers tend to stay losers. These results indicate stability in the ranking and show evidences of the quality of the measure. Moreover, the previous analysis of four one-year periods reveals that the performance provided by the new measure is persistent from 2012 to 2013 and most likely persistent from 2013 to 2014. The above results indicate that the new performance measure for ETFs seems to be of a good quality. The results from the last two subsections assess the ability of the performance measure to identify ETFs managed by managers with superior skills. The results show that, in most cases, the new performance measure is able to identify persistent winners and losers.

These results are valid if the ETFs from the sample are persistent in their performances. In order to verify that, I run the same tests of persistence for a widely used performance measure, the information ratio. Tables 32 and 33 show that the ETFs from the sample are persistent when their performance is measured with the information ratio. This result supports the conclusions from the previous analyses. It means that the ETFs from the sample are persistent in their performances and therefore a good performance measure will need to be able to show this persistence. Since the results from the tests of persistence with the new performance measure are also positive, it confirms that the new performance measure is efficient. Indeed, it is able to provide a persistent ranking from one period to the next when the funds from the sample are persistent in their performances.

Table 32: Regression Analysis Information Ratio

Period	Pearson Correlation Coefficient	P value	Analysis of Variance (test statistic)	P value	β	P value	adj. R-square	Test for normality assumption
2012-2013	0.93285	<.0001	187.72	<.0001	0.89415	<.0001	86.56%	Non Rejection
2013-2014	0.83562	<.0001	64.79	<.0001	0.9512	<.0001	68.75%	Rejection
2014-2015	0.92256	<.0001	160.06	<.0001	0.8872	<.0001	84.58%	Non Rejection
Period	**Pearson Correlation Coefficient**	**P value**	**Analysis of Variance (test statistic)**	**P value**	**β**	**P value**	**adj. R-square**	**Test for normality assumption**
2012/2013 - 2014/2015	0.86391	<.0001	82.38	<.0001	0.9670	<.0001	73.73%	Rejection

Table 33: Non-parametric tests Information Ratio

Period	Malkiel's test statistic	P value	Conclusion	Cross-Product Ratio's test statistic	P value	Conclusion	Chi Square statistic	P value	Conclusion	Spearman's Rho	P value	Conclusion
2012-2013	3.3566	0.0004	Persistence	3.6056	0.0002	Persistence	22.5333	<.0001	Persistence	0.92659	<.0001	Persistence
2013-2014	2.8402	0.0023	Persistence	3.4851	0.0002	Persistence	16.1333	<.0001	Persistence	0.81535	<.0001	Persistence
2014-2015	3.8730	0.0001	Persistence	∞	/	Persistence	30	<.0001	Persistence	0.9079	<.0001	Persistence
Period	**Malkiel's test statistic**	**P value**	**Conclusion**	**Cross-Product Ratio's test statistic**	**P value**	**Conclusion**	**Chi Square statistic**	**P value**	**Conclusion**	**Spearman's Rho**	**P value**	**Conclusion**
2012/2013 - 2014/2015	3.3566	0.0004	Persistence	3.6056	0.0002	Persistence	22.5333	<.0001	Persistence	0.8389	<.0001	Persistence

A problem that arises from the results of the information ratio is that its tests of persistence give better results than those observed with the new performance measure. This could be interpreted as a better ability of identifying persistent winners, which is a rather unexpected result. Since the performance measure that I propose in this thesis takes into account more parameters and make fewer assumptions than the information ratio, it should be more adequate in identifying funds managed by superior managers.

This result can partly be explained by two elements. Firstly, the information ratios of the ETFs from the sample are mostly positive, which is not the case for most ETFs. Indeed, Roncalli (2014) states that the tracking difference of ETFs should be negative. Since the problems with the information ratio usually arise when the tracking difference is negative, it is normal that it gives rather good results when the tracking difference is positive for the majority of the funds. In that particular case, the information ratio works. Secondly, the information ratio is widely used by investors to measure performance and to assess the skills of a portfolio manager. As a result, portfolio managers can try to manipulate it to inflate their performance.

Goetzmann, Ingersoll, Spiegel, and Welch (2007) show that it is possible to inflate many widely used performance measures without increasing the skills of a portfolio manager. According to them, the Sharpe ratio is not manipulation-proof. That is because it is possible for manager to use derivative products to manipulate the distribution of the returns. Those strategies would involve payoffs that have very low volatility (Spurgin, 2001) but that generate a return distribution with a negative skewness (Goetzmann, Ingersoll, Spiegel, & Welch, 2004). Since the Sharpe ratio assumes normality, it will not take into account the negative skewness and it can therefore be manipulated. Indeed, a manager could use this strategy to reduce its volatility and its skewness, resulting in an improved Sharpe ratio. Goetzmann et al. (2007) show that the information ratio can be manipulated in the same way as the Sharpe ratio. They define the following information ratio:

$$IR = \frac{\bar{x} - \bar{x}_b}{\sqrt{Var(\tilde{x} - \tilde{x}_b)}}$$

$\bar{x}$ and $\bar{x}_b$ are the excess returns on the portfolio and on the benchmark respectively. They combine the excess returns by adding them together and consider the arbitrage portfolio that is a combination of $\tilde{x}$ and $\tilde{x}_b$. The information ratio of this portfolio with respect to the benchmark would be equal to the Sharpe ratio of $\tilde{x}$. Therefore, the information ratio can be manipulated in the same way. A manipulation of the information ratio will then cause the distribution of excess returns to have a negative skewness. The manager can indeed try to act on the distribution of excess returns to increase the mean and reduce the standard deviation. This will probably affect the skewness and the kurtosis as well but since the information ratio does not take them into account, it will only be interpreted as an increased performance. Therefore, the funds obtained with the information ratio should have a lower skewness on average than the funds selected with the new performance measure.

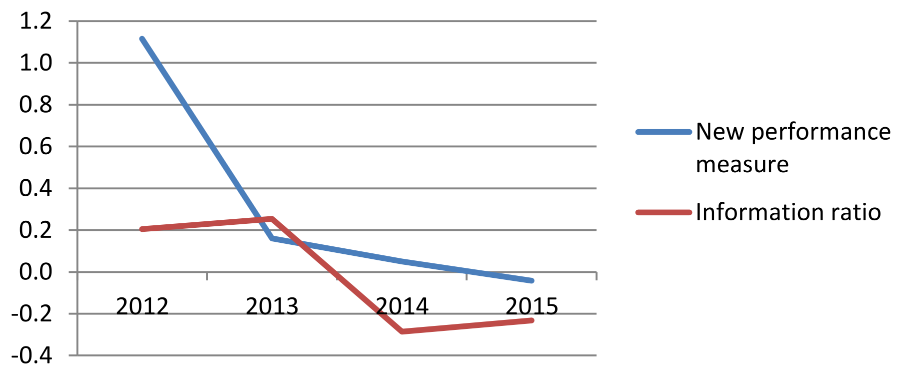
Spurgin (2001) presents various ways to increase the Sharpe ratio without increasing the actual performance of the portfolio. Since I have just explained that the information ratio could be manipulated in the same ways as the Sharpe ratio, the ways of “gaming” the Sharpe ratio presented in Spurgin (2001) are also applicable to the information ratio. He states that options change the return distribution to produce skewed, kurtotic or leptokurtotic return distributions. He proposes a new approach involving derivatives to get rid of the extreme returns. This extreme value swap, as he calls it, takes away the return from the highest monthly return and adds it to the lowest return. By doing so, the manager smooths the distribution of returns and he is able to reduce the standard deviation without affecting the total return.

If an investor wishes to invest in passive funds, he can build a portfolio composed of the persistent winners over the period of study. Therefore, I will test the superiority of the information ratio by constructing two portfolios. The first one will be composed of the persistent winner funds based on their performances with the new performance measure. The second portfolio will be composed of the persistent winners based on their performances with the information ratio. Every year, I will rebalance the portfolio so that it is only composed of the ETFs that are winners in the current period and winners in the next. Table 34 displays the average descriptive statistics of the excess returns of the ETFs in both portfolios. It can be easily seen that the ETFs selected by the new performance measure have, on average, lower standard deviation, higher skewness and lower kurtosis. This would indicate that, although the new performance measure identifies fewer persistent ETFs, it has a better selection power. Indeed, it selects funds that have better descriptive statistics than the information ratio.

Table 34: Descriptive statistics of excess returns of the persistent winners				
	2012	2013	2014	2015
Portfolio 1 (New performance measure)				
Mean	0.01%	0.01%	0.01%	0.01%
Average Standard Deviation	0.09%	0.07%	0.07%	0.05%
Average Skewness	1.1157	0.1609	0.0499	-0.0421
Average Excess Kurtosis	14.9647	1.8183	0.7830	2.3961
Portfolio 2 (Information ratio)				
Mean	0.01%	0.01%	0.01%	0.01%
Average Standard Deviation	0.10%	0.07%	0.06%	0.10%
Average Skewness	0.2052	0.2542	-0.2863	-0.2324
Average Excess Kurtosis	4.5621	7.3011	5.8357	6.5386

I will now focus on the skewness and compare the average skewness of the ETFs in both portfolios. If the skewness is significantly higher for the ETFs selected by the new performance measure, it would mean that the manipulation of the information ratio could help explain its better results at the tests of persistence. As already mentioned, performance manipulation can cause the distribution of returns to have a negative skewness (Goetzmann et al., 2004). Therefore, the objective is to try to demonstrate that the information ratio selects, as persistent winners, funds that are not necessarily well managed since their skewness suffer from performance manipulations.

Figure 6: Comparison of average skewness



It can be inferred from Figure 6 that the portfolio composed with the persistent winners identified with the new performance measure provides better results in terms of skewness. This difference could be explained by the fact that the information ratio was subject to manipulations. However, one must stay cautious when analyzing such results. Indeed, the skewness is not significantly different, except for the first year. Therefore, the manipulation of the information ratio can only explain a part of its better results. Nevertheless, this analysis shows that, despite less good results at the tests of persistence, the new performance measure that I developed in this thesis is of relatively good quality. Moreover, this measure should be more manipulation-proof than the information ratio. There are two kinds of manipulation that are presented in Goetzmann et al. (2007), the static and dynamic manipulations. The static manipulation consists in employing derivative strategies characterized by asymmetric payoffs to increase the performance measures that do not take into account the skewness of the distribution (Alcock, Glascock, & Steiner, 2013). Such a strategy will not work with the new performance measure since it penalizes a negative skewness. In that sense, it can be deduced that the performance measure that I propose in this thesis is protected against static performance manipulations.

6. Possible extensions

The performance measure that I propose in this thesis is not perfect and can be improved. In this section, I propose a few lines of thought to improve it. The first one is the problem of the too large modified tracking error. When applying the Cornish-Fisher expansion, it can happen that the kurtosis of the excess returns is extremely high. If it is the case, the ratio of the two expected shortfalls, one from the Cornish-Fisher expansion and one under the normality assumption, will also be high. This could lead to a modified tracking error greater than the standard deviation of the returns of the ETF. From Equation (13), it can be deduced that it is impossible as the square root of a negative number does not exist. Therefore, a modification of this ratio could be considered in order to deal with this problem.

The second extension is to take into account liquidity. Since ETFs are traded on an intraday basis, they have an advantage over mutual funds. The liquidity of an ETF should be taken into account when building a performance measure because it represents one of their biggest advantages. Hassine and Roncalli (2013) integrated a liquidity component in their ETF efficiency indicator. They chose the bid-ask spread as measure of liquidity or rather illiquidity. The smaller the bid-ask spread, the more liquid the fund. Therefore, they subtracted the bid-ask spread from the tracking difference in their performance measure.

A possible extension to the new performance measure could be to subtract the bid-ask spread as in Hassine and Roncalli (2013). Such a performance measure would then take into account the relative performance of the benchmark, the skewness and kurtosis of the distribution, and the liquidity. Such a measure would be very efficient as it would be able to assess a manager's ability to provide investors with an ETF that is liquid and that presents the lowest risk with respect to its benchmark. This would however make the measure even harder to compute.

7. Conclusion

In this thesis, I propose a new measure of performance for ETFs. The construction of the measure is based on aspects that are often omitted when studying the performance of passively managed funds. The first aspect is the non-normality of excess returns of the ETFs with respect to their benchmarks. The second aspect is the lack of consideration for the relative performance of the benchmark with respect to the risk-free rate of interest.

I take into account these aspects when building the new performance measure. Firstly, I create a modified tracking error based on the expected shortfall framework and on the Cornish-Fisher expansion. Secondly, I consider the research done by Hübner (2012) to take into account the performance of the benchmark. Merging the two frameworks, I propose a new performance measure for ETFs.

$$\pi^A = \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\mu(x|b) - \left(R_B - R_f\right)\left(1 - \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\right)$$

The results obtained show that the excess returns of ETFs with respect to their benchmarks are not normally distributed. Hence, the involvement of a more elaborated tracking error taking into account the 3rd and 4th moments of the distribution is justified. These results show that the information ratio and the ETF efficiency indicator proposed by Hassine and Roncalli (2013) are not applicable since they assume a normal distribution.

The graphic analysis of the performance measure using different rolling windows highlights the importance of the performance of the benchmark. Moreover, this analysis shows that the performance measure evolves in the same direction independently of the width of the window. This suggests that the new measure is stable over time and appears to be persistent in its ranking of the performance.

Finally, I assess the robustness of the measure in the measurement of the performance persistence. The results obtained show that the new performance measure is efficient since it has a good ability to produce a stable ranking over different time periods. Moreover, the results from the information ratio show that the funds are persistent in their performance. This strengthens the idea that the new performance measure is efficient. The information ratio seems to show better results than the performance measure introduced in this thesis but this could be explained by the choice of the sample or by performance manipulations. Indeed, the fact that the funds selected by the new performance measure have on average a better skewness (or other descriptive statistics) than the funds selected by the information ratio could be the consequence of performance manipulations. The information ratio is indeed not manipulation proof (Ingersoll et al., 2007).

This thesis has various theoretical implications since it questions the relevance of the currently used performance measures for passive management. The results imply that measures such as the information ratio and the ETF efficiency indicator might not be appropriate to all ETFs. However, the performance measure that I developed through this thesis seems suitable for ETFs. To answer the question raised in the introduction, it is possible to build a new performance measure for ETFs that would take into account parameters that are neglected by currently used performance measures.

I do not claim the performance measure developed in this thesis is better than any of the measures that have been proposed in the literature so far. However, it offers a new way to look at the performance of ETFs. That is because it takes into account parameters that are usually not considered when studying the performance of passively managed funds.

As stated in the previous section, this new performance measure presents some drawbacks. Firstly, when considering excess returns with high kurtosis, the measure might not work since it is only applicable when the square of the modified tracking error is smaller than the variance of the fund. Secondly, the results shown in this thesis are only valid for the sample that has been tested. In the future, further research can be carried out with larger samples to assess the efficiency of the measure on more ETFs and to consider more ETFs with negative information ratios. Lastly, the new performance measure does not take into account the liquidity, which is one of the main advantages of an ETF. Therefore, a further improvement of this new performance measure could be to include a liquidity parameter such as the bid-ask spread.

To sum up, the performance measure that I propose in this thesis is not perfect and presents some flaws. However, I have shown that it is efficient for the sample that I consider. Finally, this new performance measure offers a new look at the way of measuring the performance of ETFs.

8. References

Abner, D., (2013). *Visual Guide to ETFs*. Bloomberg Press.

Alcock, J., Glascock, J., & Steiner, E. (2013). Manipulation in U.S. REIT investment performance evaluation: Empirical evidence. *The Journal of Real Estate Finance and Economics, 47*, 434-465.

Avellaneda, M., & Zhang, S. (2010). Path-dependence of Leveraged ETF Returns. *SIAM J. Financial Mathematics, 1*, 586–603.

Babalos, V., Caporale, M., Kostakis, A., & Philippas, N. (2008). Testing for persistence in mutual fund performance and ex post verification problem: Evidence from the Greek market. *European Journal of Finance, 14*, 735-753.

Barro, D., & Canestrelli, E. (2012). Downside risk in multiperiod tracking error models. University Ca' Foscari of Venice, Dept. of Economics Research Paper Series, 17, 263-283.

Beasley, J., Meade, N., & Chang, T.-J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research, 148*, 621-643.

Ben-David, I., Franzoni, F., & Moussawi, R. (2012). ETFs, arbitrage, and contagion. *Swiss Finance Institute, 53*, 11-66.

Bohdalová, M. (2007). A comparison of Value-at-Risk methods for measurement of the financial risk. *E-Leader*. Retrieved from https://www.researchgate.net/profile/Maria_Bohdalova/publication/238796608_A_comparison_of_Value-at-Risk_methods_for_measurement_of_the_financial_risk_1/links/541817570cf25ebee988047e.pdf

Brown, S., & Goetzmann, W. (1995). Performance persistence. *The Journal of Finance, 50*, 679-698.

Carhart, M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52, 57-82.

Cavenaile, L., & Lejeune, T. (2012). A note on the use of modified Value-at-Risk. *The Journal of Alternative Investments*, 79-83.

Cogneau, P., & Hübner, G. (2009). The (more than) 100 ways to measure portfolio performance. Part 1: Standardized risk-adjusted measures. *Journal of Performance Measurement, 13,* 56-71.

Cogneau, P., & Hübner, G. (2009). The (more than) 100 ways to measure portfolio performance. Part 2: Special measures and comparison. *Journal of Performance Measurement, 14*, 56-69.

Cornish, E., & Fisher, R. (1937). Moments and cumulants in the specification of distributions. *Revue de l'Institut International de Statistique, 5*, 307-320.

Deville, L. (2008). Exchange traded funds: History, trading and research. In C. Zopounidis, M. Doumpos, & P. Pardalos (Eds.), *Handbook of Financial Engineering* (pp. 1-37). Springer.

Fabozzi, F., Neave, E., & Zhou, G. (2012). *Financial Economics*. New Jersey: Wiley.

Ferruz, L., Sarto, J., & Vargas, M. (2004). Parametric and non-parametric analysis of performance persistence in Spanish investment funds. *Frontiers in Finance and Economics, 1,* 85-100.

Fuller, R., Han, B., & Tung, Y. (2010). Thinking about indices and "passive" versus active management. *The Journal of Portfolio Management, 36*, 35-47.

Goetzmann, W., Ingersoll, J., Spiegel, M. I., & Welch, I. (2002). *Sharpening Sharpe ratios*. National bureau of economic research.

Goetzmann, W., Ingersoll, J., Spiegel, M., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies, 20,* 1503-1546.

Grinold, R. (1989). The fundamental law of active management. *Journal of Portfolio Management, 15*, 30-37.

Grinold, R., & Kahn, R. (2000). *Active portfolio management: A Quantitative approach for providing superior returns and controlling risk* (2nd ed.). McGraw-Hill.

Hassine, M., & Roncalli, T. (2013) Measuring the performance of exchange traded funds. *Journal of Investing, 4*, 57-85.

Hübner, G. (2007). How Do Performance Measures Perform? *Journal of Portfolio Management, 33*, 64-74.

Hübner, G. (2012). *Active performance in a global portfolio*. Unpublished document, HEC – The University of Liège, Liège.

Jensen, M. (1968). The performance of mutual funds in the period 1945-64. *Journal of Finance, 23*, 389-416.

Jeurissen, R., & van den Berg, J. (2005). *Index tracking using a hybrid genetic algorithm*. Paper presented at the ICSC Congress on Computational Intelligence Methods and Applications, Istanbul, Turkey.

Kahn, R., & Rudd, A. (1995). Does historical performance predict future performance? *Financial Analysts Journal, 51*, 43-52.

Kuo, T.-W., & Mateus, C. (2008, April). *The performance and persistence of exchange-traded funds: Evidence for iShares MSCI country-specific ETFs*. Paper presented at Swiss Society for Financial Market Research 11th Conference, Zürich, Switzerland.

Maillard, D. (2012). A user's guide to the Cornish Fisher expansion. Retrieved from: http://ssrn.com/abstract=1997178

Malkiel, B. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance, 50*, 549-572.

Malkiel, B. (2003). Passive investment strategies and efficient markets. *European Financial Management, 9*, 1-10.

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

Reilly, F., & Brown, C. (2012). *Analysis of investment & management of portfolio* (10th ed.). Mason, OH: South-Western Cengage Learning.

Rockafellar, R., & Uryasev, S. (2000). Optimization of conditional Value-at-Risk. *Journal of Risk, 2*, 21-41.

Roll, R. (1992). A mean/variance analysis of tracking error. *Journal of portfolio management, 18*, 13-22.

Roncalli, T. (2014, February). *Improving the efficiency of the European ETF market: Implications for regulators, providers, exchanges and investors.* Paper presented at Working Group on Risk, Paris, France.

Rudd, A. (1986). Portfolio management. *Journal of Accounting, Auditing & Finance, 1*, 242-252.

Russell, B. (2013). ETFs vs. mutual funds: Which is better? *U.S. News Digital Weekly, 5*, 1-1.

Schizas, P. (2014). Active ETFs and their performance vis-à-vis passive ETFs, mutual funds, and hedge funds. *The Journal of Wealth Management, 17*, 84-98.

Shankar, G. (2007). Active versus passive index management: A performance comparison of the S&P and Russell indexes. *The Journal of Investing, 16*, 85-95.

Shapiro, S., & Wilk, M. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika, 52*, 591-611.

Sharpe, W. (1991). The arithmetic of active management. *Financial Analysts Journal, 21*, 49-58.

Sharpe, W. (1994). The Sharpe Ratio. *Journal of Portfolio Management, 21*, 49-58.

Spurgin, R. B. (2001). How to game your Sharpe ratio. *The Journal of Alternative Investments, 4*, 38-46.

Stephens, M. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association, 69*, 730-737.

Treynor, J. (1965). How to rate management of invested funds. *Harvard Business Review, 43*, 63-75.

Uryasev, S. (2000). Conditional Value-at-Risk: Optimization algorithms and applications. *Financial Engineering News, 14*, 1-5.

A study of the performance of exchange traded funds

Exchange traded funds (ETFs) are collective investment vehicles that have known a growing interest over the past years. Yet, only a few studies were dedicated to the measurement of their performance. In this thesis, I examine one of the most widely used performance measure for passive management, the information ratio. I analyze its weaknesses and assumptions, and justify the need for a new performance measure that will be applicable to ETFs. The information ratio does not work well when the tracking difference is negative and it does not take into account the magnitude of the tracking error (Roncalli, 2014). Moreover, it assumes that the excess returns are normally distributed.

I select a sample of 30 ETFs and show that their excess returns are not normally distributed. Therefore, I develop a new performance measure that takes into account the skewness and the kurtosis of the distribution. Moreover, I consider the work of Hübner (2005) to take into account the relative performance of the benchmark. The new performance measure that I develop in this thesis has the following equation:

$$\pi^A = \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\mu(x|b) - (R_B - R_f)\left(1 - \frac{\sqrt{\sigma_P^2 - MTE^2}}{\sigma_P}\right)$$

I apply this new performance measure on the ETFs from the sample and analyze the results. As expected, the ranking obtained seems to show a positive correlation between performances of ETFs tracking the same benchmark. Moreover, a rolling window analysis highlights the stability of the measure when using windows of different widths.

In order to assess the quality of the measure, I first test its robustness in the measurement of performance persistence. I show that, using well known statistical tests, the measure is relatively robust in measuring performance persistence since the results indicate persistence for the sample. The same tests are then performed on the information ratio. Since the results also show persistence of the sample, it means that the new performance measure is good at identifying persistent winners when the sample is composed of funds that are persistent in their performance.

However, the results of the tests of persistence indicate slightly better results for the information ratio. I explore two hypotheses to explain this result. Firstly, it can be due to the characteristics of the ETFs from the sample. Secondly, it can be explained by performance manipulations on the information ratio.