Measuring Performance of Exchange Traded Funds*

Marlène Hassine
ETF Strategy
Lyxor Asset Management, Paris
marlene.hassine@lyxor.com

Thierry Roncalli
Research & Development
Lyxor Asset Management, Paris
thierry.roncalli@lyxor.com

February 2013

Abstract

Fund selection is an important issue for investors. This topic has spawned abundant academic literature. Nonetheless, most of the time, these works concern only active management, whereas many investors, such as institutional investors, prefer to invest in index funds. The tools developed in the case of active management are also not suitable for evaluating the performance of these index funds. This explains why information ratios are usually used to compare the performance of passive funds. However, we show that this measure is not pertinent, especially when the tracking error volatility of the index fund is small. The objective of an exchange traded fund (ETF) is precisely to offer an investment vehicle that presents a very low tracking error compared to its benchmark. In this paper, we propose a performance measure based on the value-at-risk framework, which is perfectly adapted to passive management and ETFs. Depending on three parameters (performance difference, tracking error volatility and liquidity spread), this efficiency measure is easy to compute and may help investors in their fund selection process. We provide some examples, and show how liquidity is more of an issue for institutional investors than retail investors.

Keywords: Passive management, index fund, ETF, information ratio, tracking error, liquidity, spread, value-at-risk.

JEL classification: G11.

1 Introduction

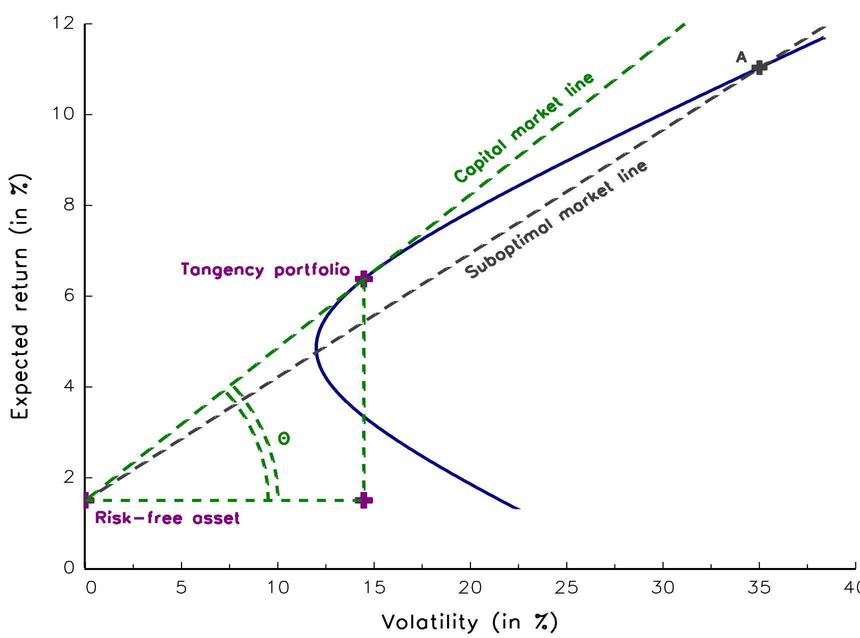
The market portfolio concept has a long history and dates back to the seminal work of Markowitz (1952). In that paper, Markowitz defines precisely what portfolio selection means: "*the investor does (or should) consider expected return a desirable thing and variance of return an undesirable thing*". Indeed, Markowitz shows that an efficient portfolio is a portfolio that maximizes the expected return for a given level of risk (corresponding to the variance of return). Markowitz concludes that there is not only one optimal portfolio, but a set of optimal portfolios called the efficient frontier (represented by the solid blue curve in Figure 1). By studying liquidity preference, Tobin (1958) shows that the efficient frontier becomes a

*We are profoundly grateful to Bou Ly Wu for his support with data management and the computation of liquidity spreads on limit order books. We would also like to thank Arnaud Llinas, Raphaël Dieterlen, Valérie Lalonde, François Millet and Matthieu Mouly for their helpful comments.

straight line in the presence of a risk-free asset. If we consider a combination of an optimized portfolio and the risk-free asset, we obtain a straight line (represented by the dashed black line in Figure 1). But one straight line dominates all the other straight line and the efficient frontier. It is called the Capital Market Line (CML), which corresponds to the green dashed line in Figure 1. In this case, optimal portfolios correspond to a combination of the risk-free asset and one particular efficient portfolio named the tangency portfolio. Sharpe (1964) summarizes the results of Markowitz and Tobin as follows: "*the process of investment choice can be broken down into two phases: first, the choice of a unique optimum combination of risky assets; and second, a separate choice concerning the allocation of funds between such a combination and a single riskless asset*". This two-step procedure is today known as the *Separation Theorem*.

One of the difficulties faced when computing the tangency portfolio is that of precisely defining the vector of expected returns of the risky assets and the corresponding covariance matrix of returns. In 1964, Sharpe developed the CAPM theory and highlighted the relationship between the risk premium of the asset (the difference between the expected return and the risk-free rate) and its beta (the systematic risk with respect to the tangency portfolio). Assuming that the market is at equilibrium, he showed that the prices of assets are such that the tangency portfolio is the market portfolio, which is composed of all risky assets in proportion to their market capitalization. That is why we use the terms, tangency portfolio and market portfolio indiscriminately nowadays.

Figure 1: Efficient frontier and the tangency portfolio



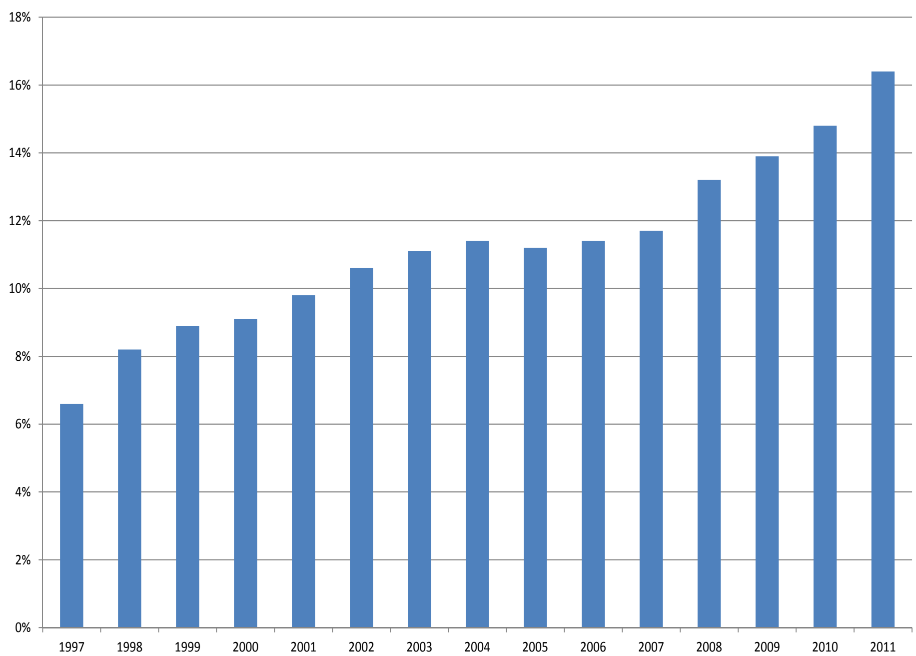
Another step forward was made by Jensen (1968), who measured the performance of 115 (equity) mutual funds using the alpha measure and concluded that:

> "*The evidence on mutual fund performance indicates not only that these 115 mutual funds were on average not able to predict security prices well enough to*

> *outperform a buy-the-market-and-hold policy, but also that there is very little evidence that any individual fund was able to do significantly better than that which we expected from mere random chance*".

The seminal work of Michael Jensen is the starting point of the development of passive management. Indeed, the first index fund was launched in 1971 by John McQuown at Wells Fargo Bank for the Samsonite luggage company (Bernstein, 1992). It took a long time for investors to accept the ideas of Markowitz, Sharpe and Jensen, but passive management represents now a large part of the asset management industry as shown in Figure 2.

Figure 2: Index funds' share in the equity mutual fund market



There is an abundance of literature that compares the performance of active managers. Indeed, academics have developed many econometrics tools in this domain to measure their alpha and their skill (Grinblatt and Titman, 1989; Blake *et al.*, 1993; Carhart, 1997; Barras *et al.*, 2010). Professionals have also created databases and ranking systems to evaluate mutual funds. For instance, Morningstar is certainly one of the best known scoring platforms. The literature is scarcer when it comes to measuring the performance of passive management. Generally, one uses the same quantitative tools as those developed for passive management to investigate the efficiency of an index fund. The aim of this paper is to provide more appropriate tools for investor looking to invest in index funds, and in exchange traded funds in particular.

The paper is organized as follows. Section 2 presents the efficiency measure, which is in fact a risk measure. In particular, we describe the reasoning behind this indicator and its rationale. Empirical results are reported in Section 3. In section 4, we consider different variations on the tracker efficiency measure. For example, we consider other definitions of the risk measure and the liquidity spread. Finally, in Section 5 we draw our conclusions

2 Measuring the efficiency of exchange traded funds

2.1 Performance or efficiency measurement?

2.1.1 Understanding fund picking in active management

Fund picking could be summarized as a two-step procedures:

- Defining a universe of funds, that are sensitive to certain risk factors;
- Picking one or more elements of this universe by combining quantitative and qualitative criteria.

The first step is crucial in order to define a universe of mutual funds, that are sufficiently homogeneous in terms of risk analysis. For instance, if we want to invest in equities, it could be useful to describe the universe more accurately. Does the investment concern global equities, regional equities or country equities? Does the investment definition relate to specific sectors, by focusing on or excluding some of them? What is the bias of the investment universe in terms of style analysis? The capacity of the investor to clearly define the borders of the universe is the most important part of the fund picking process, because the second step is relatively straightforward once the first step has been accurately completed. For example, investing in large capitalization equities located in the Eurozone that do not belong to the banking system, and are recognized as socially responsible, effectively reduces the universe to a small number of mutual funds.

Nevertheless, defining the investment universe is not always straightforward, for many reasons. First, the category may be very large with numerous mutual funds. Suppose we wanted to invest in sovereign bonds in the Eurozone. According to the Morningstar database there are more than 200 such funds available. Second, a category may not necessarily exist if the criteria are too specific[1]. Third, investment styles are heterogeneous, because they cannot be defined in a precise way. Take value and growth styles, for example. Everybody knows what they mean, but if we ask two people to classify the components of the S&P 500 index according to growth/value risk factors, we will get two different answers. This shows that investing in active management may have a significant cost, because thorough research is necessary to find accurate information. That is why, in the end, many investments are based on past performance.

2.1.2 Why is fund picking different with passive management?

With passive management, fund picking is more straightforward. Indeed, the fund universe is clearly defined once the benchmark has been chosen and the problem of performance attribution between alpha and beta components is not relevant. The investor's goal is also to identify the investment vehicle that tracks the chosen index most accurately. Rating systems such as Morningstar and Europerformance (see Box 1) are thus not suited to this selection challenge, because they are more concerned by absolute performance. For passive managers, absolute performance is meaningless. In an ideal world, they would like to buy and sell the investment at any moment and have exactly the same return as the index. Passive managers thus focus on factors other than absolute return:

[1] For example, an investor seeking exposure to investment grade sovereign bonds in the Eurozone, excluding German bonds (because the yield is too low) and Spanish bonds (because they are too risky), would struggle to find mutual funds that satisfy these criteria.

Box 1: The Europerformance/Edhec rating methodology

The ratings are constructed by combining three criteria. First, one measures extreme risks using a Cornish-Fisher value-at-risk at a 99% confidence level. If the VaR of the fund is too high, the fund is not rated. Second, the alpha of the fund is estimated. It is done in two steps. In a first step, the style analysis of Sharpe is considered to build the composite benchmark of the fund. When this benchmark is selected, one deduces alpha by a regression analysis:

$$R_t - r = \alpha + \sum_{j=1}^{n} \beta_j \left(R_t^{(j)} - r \right) + \varepsilon_t$$

where R_t is the return of the fund, $R_t^{(j)}$ is the return of the i^{th} selected benchmark and r is the risk-free rate. The alpha criterion allows us to distinguish funds rated below and above 3 stars. For example, the 4 and 5 stars correspond to funds which have a strictly positive alpha. The distinction between 4 and 5 stars is done using a gain frequency measure. It corresponds to the number of times in percent that the fund has delivered performance that was better than that of its benchmark. If the gain frequency measure is less than 50%, the fund is rated 4 stars, otherwise it is rated 5 stars. Finally, a super rating is attributed to funds rated 5 stars and which have a Hurst exponent H larger than $1/2$.

1. Management fees, other costs and additional revenues;
2. Tracking error volatility;
3. Liquidity;
4. Structuration risks.

The first factor is important because it impacts on the net performance of the investment vehicle. For instance, high management fees generally have a negative impact on the performance of mutual funds. Indeed, Elton *et al.* (1993) investigated the informational efficiency of mutual fund performance for the period 1964-1984 and concluded:

> "*There is statistically significant relationship between alpha and expenses and a relationship that can clearly be seen by examining the quintiles. Higher expenses are associated with poorer performance. Management does not increase performance by an amount sufficient that justify higher fees*".

The results are the same for index funds (Elton *et al.*, 2004). Nevertheless, management fees are not the only factor that impact directly on the return of an index fund. We can cite for example brokerage costs, dividend taxes, etc. A fund may also benefit from revenues derived from securities lending, fiscal optimization or index arbitrage[2].

The second factor measures the relative risk of the fund with respect to the index. It corresponds in general to the volatility of the tracking error, which is defined as the difference

[2] Index arbitrage concerns addition/deletion mechanisms of the index and corporate actions as subscription rights or optional dividend management (Dieterlen and Hereil, 2012).

between fund and index returns. It indicates how the tracker performance may deviate from the index performance. In the words of Harry Markowitz, it is 'an undesirable thing' for the investor. This performance difference comes from the fact that replication strategies cannot be perfect, because the invested portfolio can not track the index weights exactly all the time. We generally distinguish between three replication methods: full replication, optimized sampling and synthetic replication. In full replication, the fund manager invests in all index securities in the same proportion as the underlying index. But the portfolio weights can never be identical to the theoretical index weights, unless the fund manager buys the whole market[3]. Moreover, the portfolio weights are not static, as a result of subscriptions and redemptions, dividends, cash drag, etc.

The third factor is a key element for passive funds. The liquidity of the fund is inherently connected to the liquidity of the index. A S&P 500 index fund is certainly more liquid than an MSCI emerging markets index fund. But the investment universe is not the only factor in fund liquidity. These funds are mainly bought by institutional investors, especially in Europe. Liquidity is then crucial when they carry out subscriptions or redemptions. These decisions may have a big impact on the market. Moreover, the fund manager has little margin for discretionary decisions, because he has to follow the systematic index strategy. The index ignores more and less liquidity, but the index fund does not. Some recent events, like the March 2011 Tohoku earthquake in Japan or the Greek debt crisis, have shown how difficult it can be to manage the liquidity of an index fund in the face of such stress scenarios. Redemptions from Japanese equity index funds were difficult to execute in the days after the catastrophe in northern Japan on Friday March 11, 2011. In the same way, the decision of EuroMTS to keep Greek bonds in the EuroMTS Eurozone Govies index (ex CNO) after Greece was downgraded from investment grade in April 2010 caused problems in managing index funds that were benchmarked to this index. The problem is more complex with trackers, because they offer intra-day liquidity. And this liquidity has a cost, which can be measured by bid-ask spreads.

Another element that is important for investors is the operational structure of an ETF. We generally distinguish two main structures (physical and synthetic) used in the European markets. Even if the operational structure may be sometimes an important criteria for investors, it is more a go/no go check test than a statistic that can be taken into account in the design of an efficiency measure.

2.1.3 Portfolio optimization when there is a benchmark

We may think that the Markowitz approach is less relevant when there is a benchmark. However, it could easily be adapted by replacing the volatility of the portfolio by the volatility of the tracking error. Let b (or x) be the vector of weights of the benchmark (or the active portfolio). We note n the number of assets in the investment universe and R_i the return of

[3] Let us assume that the index is composed of two stocks. Let $N_{i,t}$ and $P_{i,t}$ be the number of outstanding shares and the price of the i^{th} stock at time t. The index weights are given by:

$$w_{i,t} = \frac{N_{i,t}P_{i,t}}{\sum_{i=1}^{2} N_{i,t}P_{i,t}}$$

For example, if $N_{1,t} = 136017$, $N_{2,t} = 87123$, $P_{1,t} = 110.54$, $N_{2,t} = 125.23$, we have $w_{1,t} = 57.95\%$ and $w_{2,t} = 42.05\%$. If the value of the fund is \$1000, full replication involves buying respectively 5 and 3 shares of the first and second stocks, whereas the cash represents \$71.61. The portfolio weights of the two stocks are then $w_{1,t} = 55.27\%$ and $w_{2,t} = 37.57\%$. At time $t+1$, we assume that there is a subscription of \$60. If the prices remain the same, we could buy another share of either the first stock or the second stock. In the two cases, we obtain portfolio weights that are different from the index weights.

the asset i. The tracking error of the active portfolio x with respect to the benchmark is the difference between the return of the portfolio $R(x)$ and the return of the benchmark $R(b)$:

$$\begin{aligned} e &= R(x) - R(b) \\ &= \sum_{i=1}^{n} x_i R_i - \sum_{i=1}^{n} b_i R_i \\ &= x^\top R - b^\top R \\ &= (x-b)^\top R \end{aligned}$$

The expected tracking error is defined as follows:

$$\begin{aligned} \mu(x \mid b) &= \mathbb{E}[e] \\ &= (x-b)^\top \mu \end{aligned}$$

whereas tracking error volatility is equal to[4]:

$$\begin{aligned} \sigma^2(x \mid b) &= \sigma^2(e) \\ &= (x-b)^\top \Sigma (x-b) \end{aligned}$$

The objective of the investor is to maximize the expected tracking error with a constraint on the tracking error volatility:

$$\begin{aligned} x^\star &= \arg\max (x-b)^\top \mu \\ &\text{u.c.} \quad \mathbf{1}^\top x = 1 \text{ and } \sigma(e) \leq \sigma^\star \end{aligned}$$

We gave an example of the tracking-error efficient frontier in Figure 3. It is a straight line when there is no restriction (Roll, 1992). If we impose some constraints, the efficient frontier is moved to the left and is no longer a straight line.

To compare the performance of different portfolios, a better measure than the Sharpe ratio is the information ratio, which is defined as follows (Grinold and Kahn, 2000):

$$\begin{aligned} \mathrm{IR}(x \mid b) &= \frac{\mu(x \mid b)}{\sigma(x \mid b)} \\ &= \frac{(x-b)^\top \mu}{\sqrt{(x-b)^\top \Sigma (x-b)}} \end{aligned}$$

If we consider a combination of the benchmark b and the active portfolio x, the composition of the portfolio is:

$$y = (1-\alpha) b + \alpha x$$

with $\alpha \geq 0$ the proportion of wealth invested in the portfolio x. We get:

$$\mu(y \mid b) = (y-b)^\top \mu = \alpha \mu(x \mid b)$$

and:

$$\sigma^2(y \mid b) = (y-b)^\top \Sigma (y-b) = \alpha^2 \sigma^2(x \mid b)$$

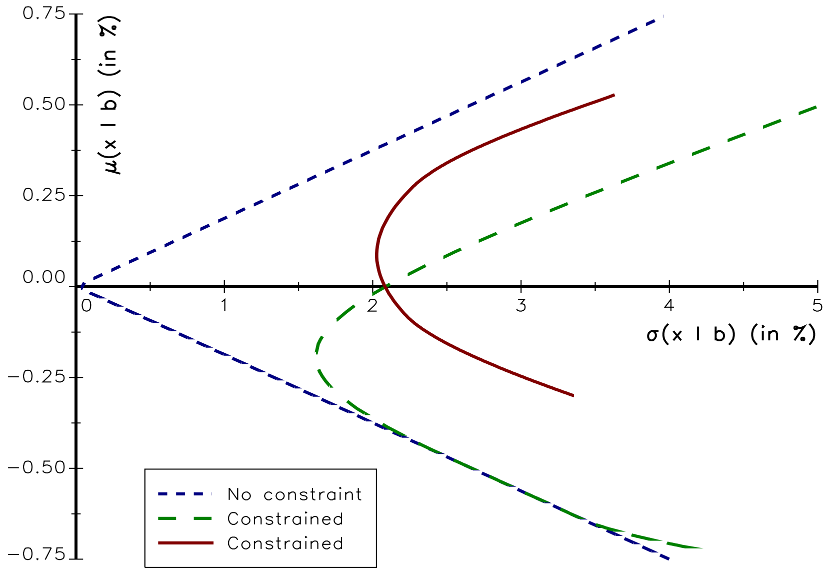
We deduce that:

$$\mu(y \mid b) = \mathrm{IR}(x \mid b) \cdot \sigma(y \mid b)$$

It is the equation of a linear function between the expected tracking error and the tracking error volatility of the portfolio y. It implies that the efficient frontier is a straight line:

[4] In IOSCO terminology, $\mu(x \mid b)$ corresponds to *Tracking Difference* (TD) whereas $\sigma(x \mid b)$ is called *Tracking Error* (TE).

Figure 3: Tracking-error efficient frontier



"*If the manager is measured solely in terms of excess-return performance, he or she should pick a point on the upper part of this efficient frontier. For instance, the manager may have a utility function that balances expected value added against tracking-error volatility. Note that because the efficient set consists of a straight line, the maximal Sharpe ratio is not a usable criterion for portfolio allocation*" (Jorion, 2003, page 172).

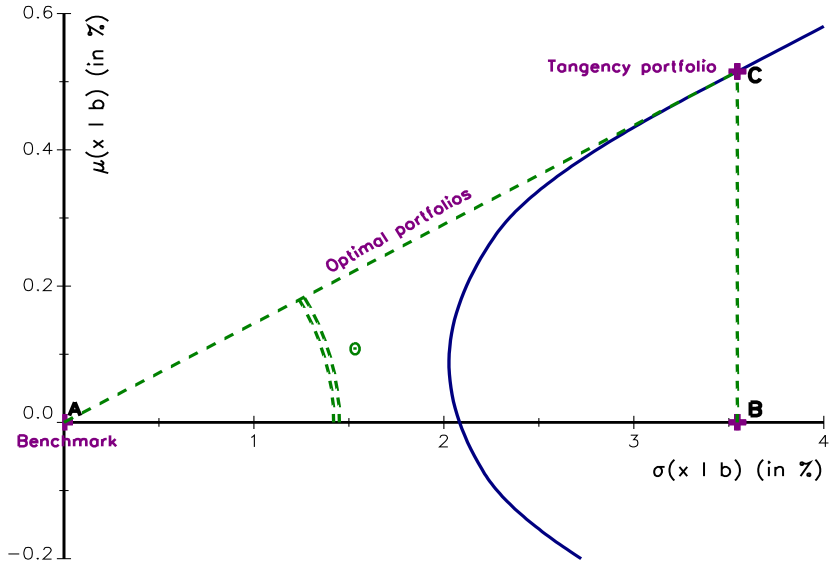
If we add some other constraints to the portfolio optimization problem, the efficient frontier is no more a straight line. In this case, one optimized portfolio dominates all the other portfolios. It is the portfolio that belongs to the efficient frontier and the straight line that is tangent to the efficient frontier. It is also the portfolio that maximizes the information ratio. An illustration is provided in Figure 4. In this case, we check that the tangency portfolio is the one that maximizes the tangent θ, which is the information ratio:

$$\begin{aligned}\tan\theta &= \frac{BC}{AB} \\ &= \frac{\mu(x \mid b)}{\sigma(x \mid b)} \\ &= \mathrm{IR}(x \mid b)\end{aligned}$$

2.1.4 The information ratio as a selection criteria for benchmarked funds

To understand why the information ratio is suitable for comparing benchmarked funds, we consider the example given in Figure 5. It is obvious that portfolio x_2 is preferable to portfolio x_1 because it has a better excess-return performance for the same tracking-error volatility. If we compare portfolios x_3 and x_2, they don't have the same tracking-error

Figure 4: The geometry of the information ratio



volatility. Nevertheless, we could build a portfolio x_4 by mixing benchmark b and portfolio x_2 such that x_4 has the same tracking-error volatility as x_3. According to the previous analysis, we get:

$$\left\{ \begin{array}{rcl} x_4 & = & (1-\alpha)\, b + \alpha x_2 \\ \alpha & = & \sigma\left(x_3 \mid b\right) / \sigma\left(x_2 \mid b\right) \end{array} \right.$$

In this case, portfolio x_4 is preferable to portfolio x_3 because it performs better. We can conclude that x_2 is preferable to x_3. The information ratio of x_4 is by its very nature equal to the information ratio of x_2. So, to select a portfolio, we simply compare their information ratios. We thus get the following result[5].

Proposition 1 *Let x and y be two portfolios benchmarked on the same index b. x is preferable to y if and only if the information ratio of x is greater than the information ratio of y:*

$$x \succ y \Leftrightarrow \mathrm{IR}\left(x \mid b\right) \geq \mathrm{IR}\left(y \mid b\right)$$

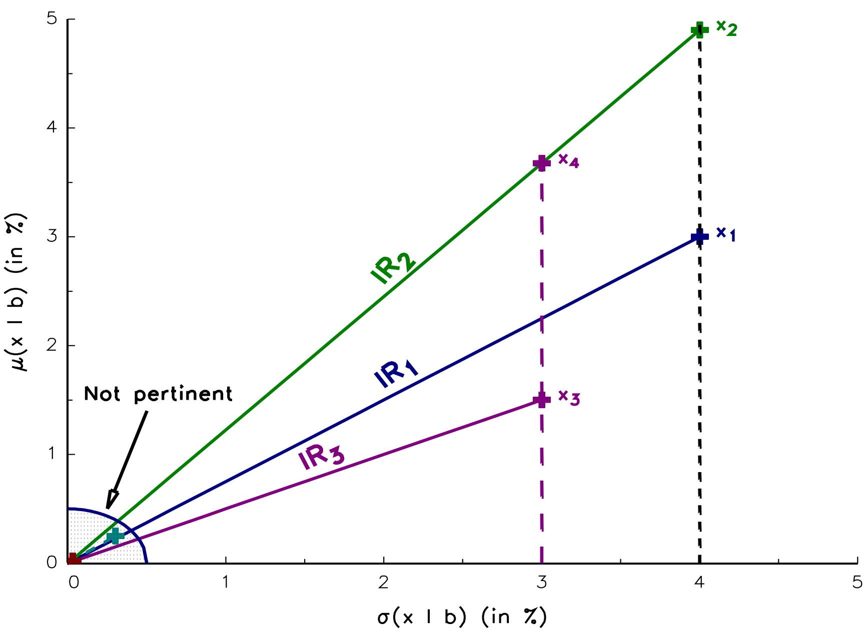
2.1.5 Pitfalls of the information ratio

This proposition is the core of fund selection processes for benchmarked funds. However, it is only valid for funds with significant tracking-error volatility. Indeed, to establish this proposition, we have made the assumption that we can replicate the benchmark exactly. In real life, we need to use a tracker to proxy the benchmark. Let y be the combination of tracker x_0 and portfolio x. We get:

$$y = (1-\alpha)\, x_0 + \alpha x$$

[5]We assume that the information ratio is positive.

Figure 5: Comparing benchmarked portfolios



In Appendix A.1, we show that:

$$\mathrm{IR}\left(y \mid b\right) = \frac{\left(1-\alpha\right)\mu\left(x_0 \mid b\right) + \alpha\mu\left(x \mid b\right)}{\sqrt{\left(1-\alpha\right)\sigma^2\left(x_0 \mid b\right) + \alpha\sigma^2\left(x \mid b\right) + \left(\alpha^2-\alpha\right)\sigma^2\left(x \mid x_0\right)}} \tag{1}$$

If $\mu\left(x_0 \mid b\right) = 0$ and $\sigma\left(x_0 \mid b\right) = 0$, we get the previous result:

$$\mathrm{IR}\left(y \mid b\right) = \mathrm{IR}\left(x \mid b\right)$$

In general, the tracker is such that[6] $\mu\left(x_0 \mid b\right) < 0$ and $\sigma\left(x_0 \mid b\right) > 0$. In this case, we obtain:

$$\mathrm{IR}\left(y \mid b\right) < \mathrm{IR}\left(x \mid b\right)$$

and these two information ratios are closed when the tracking-error volatility of the portfolio x is larger than that of the tracker x_0 – $\sigma\left(x_0 \mid b\right) \ll \sigma\left(x \mid b\right)$. In other cases the use of the information ratios may be not appropriate.

Let us illustrate this drawback with an example. We assume that $\mu\left(x_0 \mid b\right) = -5$ bps, $\mu\left(x_1 \mid b\right) = 45$ bps, $\sigma\left(x_1 \mid b\right) = 45$ bps, $\sigma\left(x_1 \mid x_0\right) = 25$ bps and $\sigma\left(x_0 \mid b\right) = 5$ bps. Portfolio x_1 and tracker x_0 are represented in Figure 6. Next we consider a second portfolio x_2 with the following characteristics $\mu\left(x_2 \mid b\right) = 15$ bps and $\sigma\left(x_2 \mid b\right) = 15$ bps. We get $\mathrm{IR}\left(x_1 \mid b\right) = 1.29$ and $\mathrm{IR}\left(x_2 \mid b\right) = 1.0$. Using Proposition 1, we can deduce that x_1 is better than x_2 because we could build the portfolio x_3 that dominates x_2:

$$x_3 \succ x_2 \Rightarrow x_1 \succ x_2$$

But the problem is that the information ratio of x_3 could not reached because the benchmark could only be replicated by tracker x_0. In this case, the combination of x_1 and x_0 gives x_4

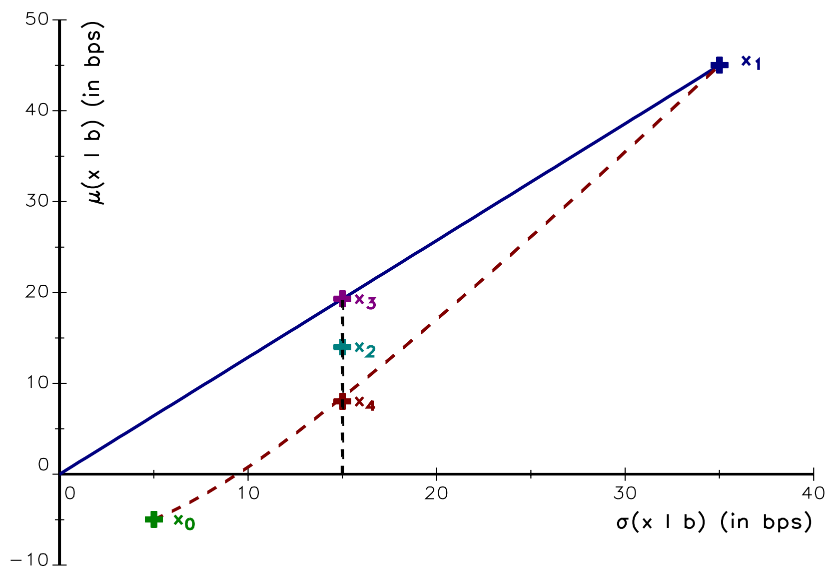
[6]Due to the management fees.

and we find that $x_2 \succ x_4$. In this example, if we target a tracking-error volatility of 15 bps, we conclude that x_2 dominates x_1 even if the information ratio of x_1 is larger. We conclude that:

$$\mathrm{IR}\,(x \mid b) > \mathrm{IR}\,(y \mid b) \nRightarrow x \succ y$$

This problem concerns benchmarked funds with low tracking-error volatility in particular. In this case, the information ratio cannot be used to select funds. This is especially true for trackers that aim to have the lowest tracking-error volatility. Moreover, the information ratio is not appropriate for measuring the relative performance of the tracker that presents the lowest tracking error. Indeed, it is generally a fund that has a small but negative excess-return performance. Because its information ratio is negative, this tracker is not efficient from the information ratio perspective. It means that no investors are interested in using it. This theoretical point of view is in contradiction with practice, because passive investors like to consider funds with the smallest tracking error volatility.

Figure 6: Information ratio with the tracker



Remark 1 *Another problem arises from the fact that the information ratio ignores the magnitude of tracking-error volatility. For instance, if we consider two trackers with the following characteristics:*

Tracker	$\mu\,(x_0 \mid b)$	$\sigma\,(x_0 \mid b)$	$\mathrm{IR}\,(x_0 \mid b)$
#1	0.02%	0.03%	0.66
#2	0.40%	0.50%	0.80

the second tracker is preferred to the first tracker if we use the information ratio as the selection criteria. However, the first tracker does a better replication job than the second one.

2.2 A comprehensive efficiency indicator for trackers

Let us consider a simple model with two periods. The investor buy the tracker x at time $t = 0$ and sells it at time $t = 1$. Note the corresponding tracking error e. The relative PnL of the investor with respect to the benchmark b is:

$$\Pi(x \mid b) = e - s(x \mid b)$$

where $s(x \mid b)$ is the bid-ask spread of the tracker. We define the loss $\mathcal{L}(x \mid b)$ of the investor as follows:

$$\mathcal{L}(x \mid b) = -\Pi(x \mid b)$$

The tracker efficiency measure is a risk measure applied to the loss function $\mathcal{L}(x \mid b)$ of the investor. We propose to use value-at-risk, which is today commonly accepted as a standard risk measure. In this case, the efficiency measure $\zeta_\alpha(x \mid b)$ is defined as follows[7]:

$$\zeta_\alpha(x \mid b) = -\{\inf \zeta : \Pr\{\mathcal{L}(x \mid b) \leq \zeta\} \geq \alpha\}$$

This means that the investor has a probability of $1 - \alpha$ of losing an amount greater than $-\zeta_\alpha(x \mid b)$. Let $\mathbf{F}$ be the probability distribution function of $\mathcal{L}(x \mid b)$. We get:

$$\zeta_\alpha(x \mid b) = -\mathbf{F}^{-1}(\alpha)$$

If we consider the tracking error model of the previous section and if we assume that asset returns are Gaussian, we obtain the closed-form formula shown in Appendix A.2. We can then derive the following definition:

Definition 1 *The efficiency measure $\zeta_\alpha(x \mid b)$ of the tracker x with respect to the benchmark b corresponds to:*

$$\zeta_\alpha(x \mid b) = \mu(x \mid b) - s(x \mid b) - \Phi^{-1}(\alpha)\,\sigma(x \mid b) \tag{2}$$

where $\mu(x \mid b)$ is the expected value of the tracking error, $s(x \mid b)$ is the bid-ask spread and $\sigma(x \mid b)$ is the volatility of the tracking error.

We illustrate the computation of the efficiency measure in Figure 7 when the confidence level α is set to 95%. We assume that $\mu(x \mid b) = 50$ bps, $\sigma(x \mid b) = 40$ bps and $s(x \mid b) = 20$ bps. The curve corresponds to the density function of the PnL of the investor. It is not centered on $\mu(x \mid b)$ because of the cost of selling the tracker. It explains that the mean of the P&L is 30 bps. According to the Gaussian hypothesis, there is a 50% probability that the investor's gain is greater than 30 bps, but there is also a 50% probability that investor's gain will be less than 30 bps. The probability of the investor making a loss is therefore equal to 22.66%. To compute the efficiency measure of the tracker, we take the quantile at 5%. We finally obtain $\zeta_\alpha(x \mid b) = -35.79$ bps.

In Figure 8, we show the impact of the different factors on the efficiency measure. The solid line corresponds to the previous parameter values whereby we modify one parameter value for each dashed line. For instance, in the first panel, we consider that $\mu(x \mid b)$ is equal to 70 bps. This tracker is preferred as the original tracker because it has a better excess-return performance, the same tracking error volatility and the same bid-ask spread. In the second panel, the second tracker has a wider spread[8]. In this case, we prefer the first tracker. For the third panel, we obtain the same conclusion if the second tracker has a greater tracking error volatility. We can then deduce the following result:

[7] We consider the opposite of the loss in order to obtain an ascending order: the bigger the efficiency measure, the better the tracker.

[8] It is equal to 40 bps.

Figure 7: Computing the efficiency measure

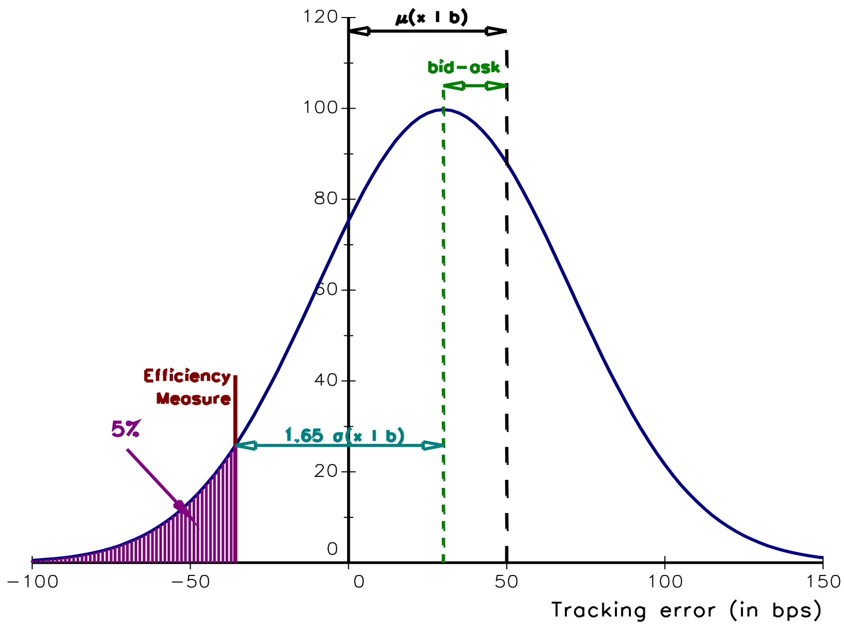
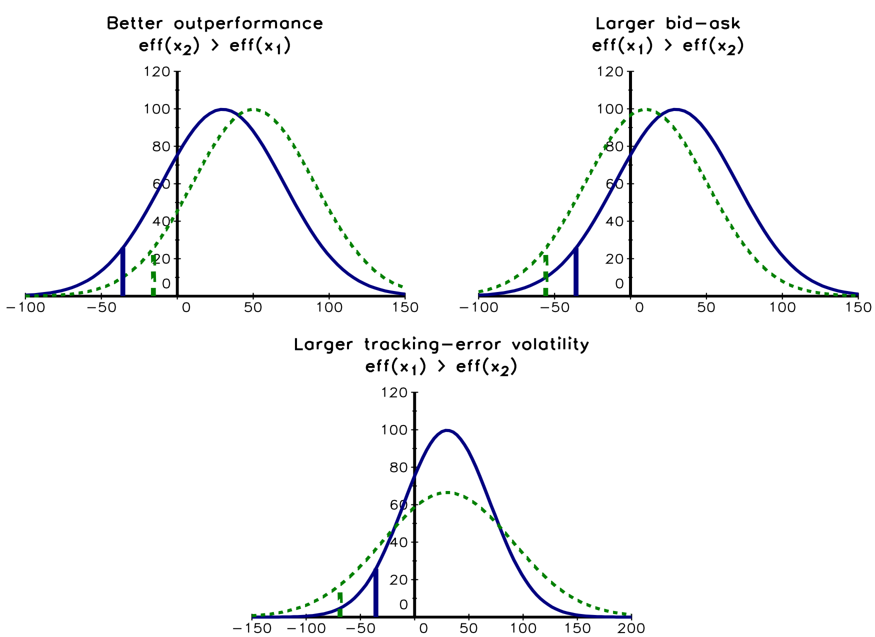


Figure 8: The effets of the different factors on the efficiency measure



Proposition 2 *Let x and y be two trackers benchmarked to the same index b. x is preferable to y if and only if the efficiency measure of x is larger than the efficiency measure of y:*

$$x \succ y \Leftrightarrow \zeta_\alpha (x \mid b) \geq \zeta_\alpha (y \mid b)$$

Example 1 *We consider two trackers x and y with the following characteristics: $\mu (x \mid b) = 40$ bps, $s (x \mid b) = 20$ bps, $\sigma (x \mid b) = 30$ bps, $\mu (x \mid b) = 30$ bps, $s (x \mid b) = 15$ bps, $\sigma (x \mid b) = 20$ bps. We have $\Phi^{-1} (\alpha) = 1.65$ at the confidence level 95%. It follows that:*

$$\begin{aligned} \zeta_\alpha (x \mid b) &= 40 - 20 - 1.65 \times 30 \\ &= -29.5 \end{aligned}$$

and:

$$\begin{aligned} \zeta_\alpha (y \mid b) &= 30 - 15 - 1.65 \times 20 \\ &= -18.0 \end{aligned}$$

Because $\zeta_\alpha (y \mid b) > \zeta_\alpha (x \mid b)$, we can deduce that tracker y is more efficient than tracker x. A loss of more than 29.5 bps has a 5% probability of occurring for tracker x. In the case of tracker y, this loss is only 18 bps for the same level of probability.

3 Empirical results

In this paragraph, we apply the efficiency measure $\zeta_\alpha (x \mid b)$ to the European ETF market. The confidence level is set at 95%. Let $[t_0, T]$ be the study period with n trading dates. We compute $\zeta_\alpha (x \mid b)$ as follows:

$$\zeta_\alpha (x \mid b) = \hat{\mu} (x \mid b) - \hat{s} (x \mid b) - 1.65 \cdot \hat{\sigma} (x \mid b)$$

where $\hat{\mu} (x \mid b)$, $\hat{s} (x \mid b)$ and $\hat{\sigma} (x \mid b)$ are the estimated statistics for the given study period. We have:

$$\begin{aligned} \hat{\mu} (x \mid b) &= \left(\frac{V_T (x)}{V_{t_0} (x)} \right)^{1/(T-t_0)} - \left(\frac{V_T (b)}{V_{t_0} (b)} \right)^{1/(T-t_0)} \\ \hat{s} (x \mid b) &= \frac{1}{n} \sum_{t=t_0}^{T} s_t (x \mid b) \\ \bar{\mu} (x \mid b) &= \frac{1}{n-1} \sum_{t=t_0+1}^{T} R_t (x) - \frac{1}{n-1} \sum_{t=t_0+1}^{T} R_t (b) \\ \hat{\sigma} (x \mid b) &= \sqrt{\frac{260}{n-1} \sum_{t=t_0+1}^{T} \left(R_t (x) - R_t (b) - \bar{\mu} (x \mid b) \right)^2} \end{aligned}$$

where $V_t (x)$ (or $V_t (b)$) is the net asset value of the tracker (or benchmark) at time t, $s_t (x \mid b)$ is the bid-ask spread[9] at time t and $R_t (x)$ (or $R_t (b)$) is the daily return of the tracker (or the benchmark) defined as follows:

$$R_t (x) = \frac{V_t (x)}{V_{t-1} (x)} - 1$$

[9] More precisely, we compute the best spread of the first limit order for each listing place and each trading day t. $s_t (x \mid b)$ is then the weighted average by considering the daily volume of the different listing places.

3.1 An application to European ETFs

The study period begins in November, 30th 2011 and ends in November, 30th 2012. We consider three benchmarks: the Eurostoxx 50 index, the S&P 500 index and the MSCI World index. To compute the statistic $\zeta_\alpha(x \mid b)$ properly, we need to make some statistical adjustments:

- Some ETFs distribute dividends. In this case, we have to rebuild the net asset value by incorporating these dividends in order to compute performance that can be compared to the performance of the benchmark.
- Dividends also influence the tracking-error volatility. That is why we correct the daily performance of the tracker every time a dividend is paid.
- Some ETFs have poor liquidity meaning, that we cannot observe a spread for each trading date. To compute $\hat{s}(x \mid b)$, we decide to exclude dates with zero trading volume.
- ETFs may be listed in different currencies. We therefore adopt the point of view of a European investor and consider the Euro as the default currency.

Table 1: Computation of the efficiency measure (Eurostoxx 50)

Tracker	$\hat{\mu}(x)$	$\hat{\mu}(x \mid b)$	$\hat{s}(x \mid b)$	$\hat{\sigma}(x)$	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$
Amundi	15.29	62.56	9.84	22.33	11.97	32.98
db X-trackers	15.33	65.97	12.27	22.86	7.31	41.64
iShares (DE)	15.05	37.88	7.96	21.62	56.54	−63.38
iShares	15.25	58.46	10.39	21.92	19.62	15.70
Lyxor	15.30	63.51	8.48	22.01	14.89	30.47
Source	14.90	23.51	15.38	22.23	7.25	−3.83
Index	14.67			22.09		

Table 2: Computation of the efficiency measure (S&P 500)

Tracker	$\hat{\mu}(x)$	$\hat{\mu}(x \mid b)$	$\hat{s}(x \mid b)$	$\hat{\sigma}(x)$	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$
Amundi	19.49	9.19	16.97	12.59	3.14	−12.97
Credit Suisse	19.57	16.99	18.23	12.50	4.63	−8.88
db X-trackers	19.56	16.04	18.26	12.79	4.65	−9.90
HSBC	19.68	28.20	20.58	12.68	3.45	1.92
iShares	19.34	−6.10	7.45	12.56	4.90	−21.63
Lyxor	19.60	19.87	13.56	12.59	0.98	4.69
Source	19.34	−5.30	17.81	12.87	1.78	−26.04
UBS	19.40	0.02	41.13	12.85	0.59	−42.09
Index	19.40			12.76		

Results[10] are reported in Tables 1, 2 and 3. According to the efficiency measure $\zeta_\alpha(x \mid b)$, the best tracker for the Eurostoxx 50 index is db X-trackers, followed by Amundi and Lyxor. We notice that $\zeta_\alpha(x \mid b)$ assigns a positive value to most of the trackers, due probably to

[10] All the statistics are expressed in bps except $\hat{\mu}(x)$ and $\hat{\sigma}(x)$ which are measured in %.

fiscal optimization of dividends. These results are different for the S&P 500 and MSCI World indices. Indeed, $\zeta_\alpha(x \mid b)$ is generally negative. It is due to lower performance and a wider spread, even if ETFs tracking the S&P 500 and MSCI World indices have less tracking-error volatility[11] than for the Eurostoxx 50 index. At the end, Lyxor is ranked top, followed by HSBC and Credit Suisse for the S&P 500 index, and Commerzbank and Amundi for the MSCI World index.

Table 3: Computation of the efficiency measure (MSCI World)

Tracker	$\hat{\mu}(x)$	$\hat{\mu}(x \mid b)$	$\hat{s}(x \mid b)$	$\hat{\sigma}(x)$	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$
Amundi	17.40	−20.75	23.91	10.32	3.67	−50.72
Commerzbank	17.42	−18.25	18.72	10.83	3.56	−42.85
db X-trackers	17.36	−24.46	13.20	10.30	25.18	−79.20
iShares	17.18	−42.08	13.75	10.14	50.80	−139.65
Lyxor	17.37	−23.09	11.93	10.09	1.55	−37.58
Source	17.08	−52.64	24.83	10.31	1.69	−80.25
UBS	17.36	−23.98	31.25	10.39	14.51	−79.16
Index	17.60			10.18		

Table 4: Computation of the efficiency measure (MSCI EM)

Tracker	$\hat{\mu}(x)$	$\hat{\mu}(x \mid b)$	$\hat{s}(x \mid b)$	$\hat{\sigma}(x)$	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$
Credit Suisse	13.20	−205.79	30.05	13.18	150.68	−484.46
db X-trackers	14.13	−112.34	15.85	13.07	12.83	−149.35
iShares	14.18	−107.56	17.90	13.07	160.21	−389.80
Lyxor	14.45	−80.01	20.72	13.07	14.87	−125.28
Source	14.16	−109.20	50.30	13.22	3.96	−166.02
Index	15.25			13.12		

3.2 Different benchmarks

One difficulty arises when ETF providers do not choose the same benchmark to give access to an asset class. For instance, if we consider Japanese equities, we notice that some providers have chosen to replicate the Topix index, whereas other providers have opted for the MSCI Japan index. In this case, results depend on the reference index. Nevertheless, although the two benchmarks are highly correlated, they may be coherent and the choice of the reference index has a limited impact on the ranking. In this situation, we use principal component analysis to build a reference index, as suggested by Amenc and Martellini (2002). The idea is to extract the best possible one-dimensional summary of the set of these different benchmarks.

For instance, in our example, the reference index is composed by 49.50% of the Topix index and 50.50% of the MSCI Japan index[12], and we obtain the results given in Table 5. For Japanese equities, iShares appears then the best tracker.

[11] Nevertheless, we note some divergence in the case of MSCI World trackers. This is even more so when we consider MSCI EM trackers (see Table 4.

[12] Methodological details to compute these weights are provided in Appendix A.3.

Table 5: Computation of the efficiency measure (Japanese equities)

Tracker	$\hat{\mu}(x)$	$\hat{\mu}(x \mid b)$	$\hat{s}(x \mid b)$	$\hat{\sigma}(x)$	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$
db X-trackers	−0.72	−35.87	15.82	16.58	256.73	−475.29
iShares	−0.76	−40.64	19.98	15.97	66.84	−170.90
Lyxor	−1.29	−92.96	14.98	16.10	62.81	−211.58
Source	−0.67	−31.40	35.11	16.58	65.30	−174.25
Index	−0.36			15.84		

4 Variations on the tracker efficiency measure

4.1 Tracking error volatility or semi-volatility?

One could argue that tracking error volatility is not a reliable measure of risk, because it depends on both positive and negative tracking errors. That is why different authors suggest using semi-variance instead of variance (Markowitz, 1959; Fishburn, 1977; Harlow, 1991; Estrada, 2007). Semi-variance is a special case of lower partial moments (see Appendix A.4). In Table 7, we report the values taken by the semi-volatility $\hat{\sigma}_-(x \mid b)$ of the tracking error for the Eurostoxx 50 trackers. We consider two values for the threshold. When $\tau = \mu(x \mid b)$, the semi-volatility $\hat{\sigma}_-(x \mid b)$ corresponds to the square root of the below-mean semi-variance. When $\tau = 0$, the semi-volatility $\hat{\sigma}_-(x \mid b)$ is the square root of below-target semi-variance (Nawrocki, 1999). By definition, semi-volatility is lower than volatility. We verify this property in Table 7. If the distribution of tracking errors is perfectly symmetric around the mean, the ratio between volatility and semi-volatility is equal to $\sqrt{2}$. That is why we compute the modified efficiency measure as follows:

$$\zeta_\alpha^\star(x \mid b) = \hat{\mu}(x \mid b) - \hat{s}(x \mid b) - 1.65 \cdot \sqrt{2} \cdot \hat{\sigma}_-(x \mid b)$$

We notice that the ranking changes slightly in comparison to the previous one. Amundi is thus the best tracker, followed by db X-trackers and Lyxor.

Table 6: Estimated risk of Eurostoxx 50 trackers

			$\tau = \mu(x \mid b)$		$\tau = 0$	
Tracker	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$	$\hat{\sigma}_-(x \mid b)$	$\zeta_\alpha^\star(x \mid b)$	$\hat{\sigma}_-(x \mid b)$	$\zeta_\alpha^\star(x \mid b)$
Amundi	11.97	32.98	3.43	44.72	1.05	50.28
db X-trackers	7.31	41.64	4.11	44.12	2.86	47.04
iShares (DE)	56.54	−63.38	40.38	−64.32	40.01	−63.45
iShares	19.62	15.70	11.86	20.38	11.00	22.40
Lyxor	14.89	30.47	5.74	41.63	4.28	45.05
Source	7.25	−3.83	1.88	3.74	1.06	5.66

Looking at the results in Table 7, the ranking of S&P 500 trackers is unchanged: Lyxor remains the best tracker, followed by HSBC and Credit Suisse. We conclude that the choice of semi-variance over volatility does not fundamentally alter the results. But the efficiency measure based on semi-variance has an important drawback. It is less sensitive to tracking errors, which is however an important criteria for investors. Another important point is that semi-variance is just another way of measuring the skewness of a distribution:

Table 7: Estimated risk of S&P 500 trackers

			$\tau = \mu(x \mid b)$		$\tau = 0$	
Tracker	$\hat{\sigma}(x \mid b)$	$\zeta_\alpha(x \mid b)$	$\hat{\sigma}_-(x \mid b)$	$\zeta^\star_\alpha(x \mid b)$	$\hat{\sigma}_-(x \mid b)$	$\zeta^\star_\alpha(x \mid b)$
Amundi	3.14	−12.97	1.22	−10.64	0.84	−9.74
Credit Suisse	4.63	−8.88	3.01	−8.26	2.53	−7.13
db X-trackers	4.65	−9.90	3.39	−10.13	2.91	−9.01
HSBC	3.45	1.92	2.01	2.92	1.25	4.71
iShares	4.90	−21.63	3.26	−21.15	3.34	−21.34
Lyxor	0.98	4.69	0.68	4.73	0.28	5.67
Source	1.78	−26.04	1.10	−25.67	1.23	−25.97
UBS	0.59	−42.09	0.37	−41.98	0.35	−41.93

> "*By taking the variance and dividing it by the below-mean semi-variance, a measure of skewness resulted. If the distribution is normally distributed then the semi-variance should be one-half of the variance [...] If the ratio is not equal to 2, then there is evidence that the distribution is skewed or asymmetric*" (Nawrocki, 1999, page 11).

If we compute this ratio for the previous trackers, we get a range between 1.9 and 14.8 for the Eurostoxx 50 index and between 1.9 and 6.6 for the S&P 500 index. This clearly demonstrates that tracking errors exhibit high skewness. This result is confirmed by the Kernel analysis of empirical density (see Figures 9, 10 and 11). In this context, we think that replacing volatility by semi-volatility is not a good method. It would be better to adopt another risk measure that takes the skewness of tracking errors into account.

Figure 9: Estimated density of daily tracking errors (Eurostoxx 50)

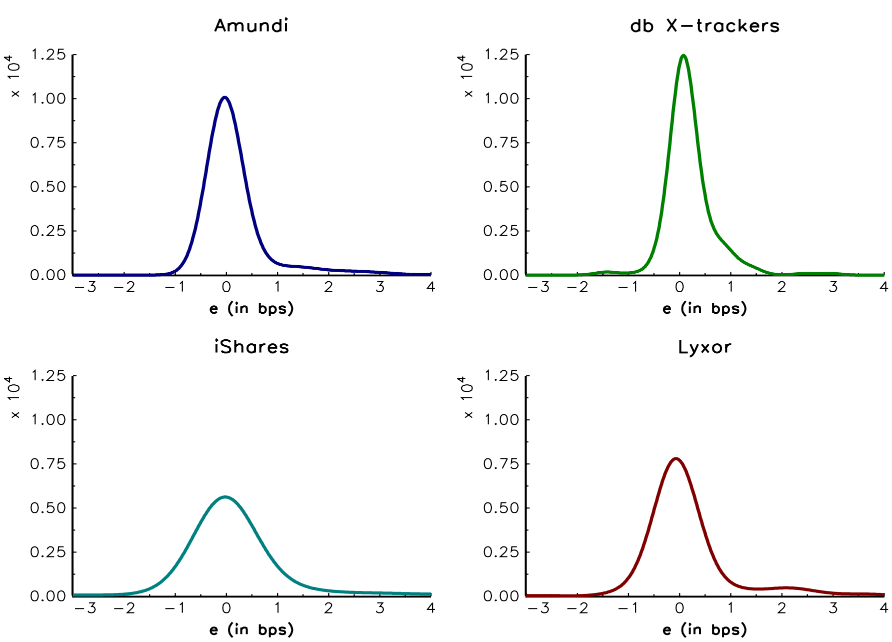


Figure 10: Estimated density of daily tracking errors (S&P 500)

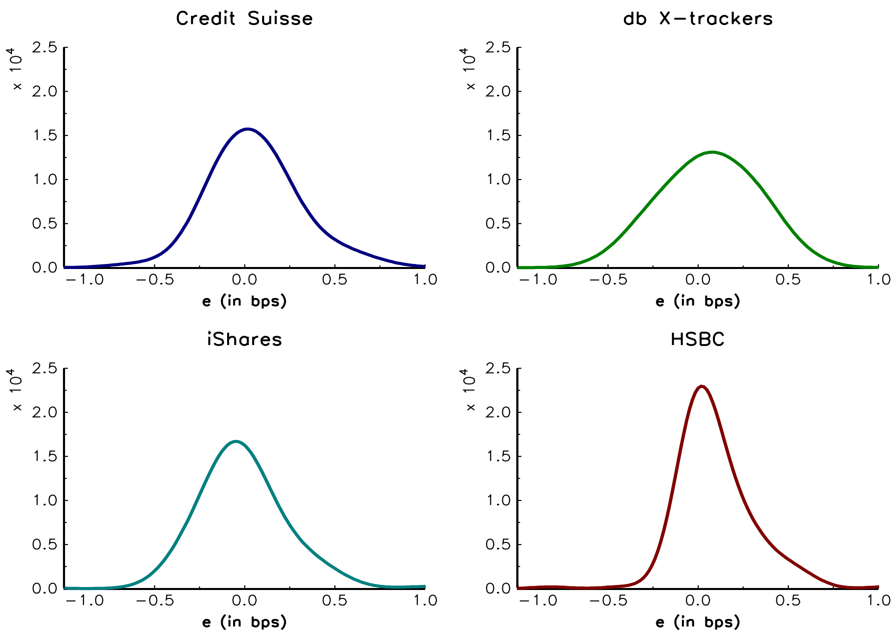
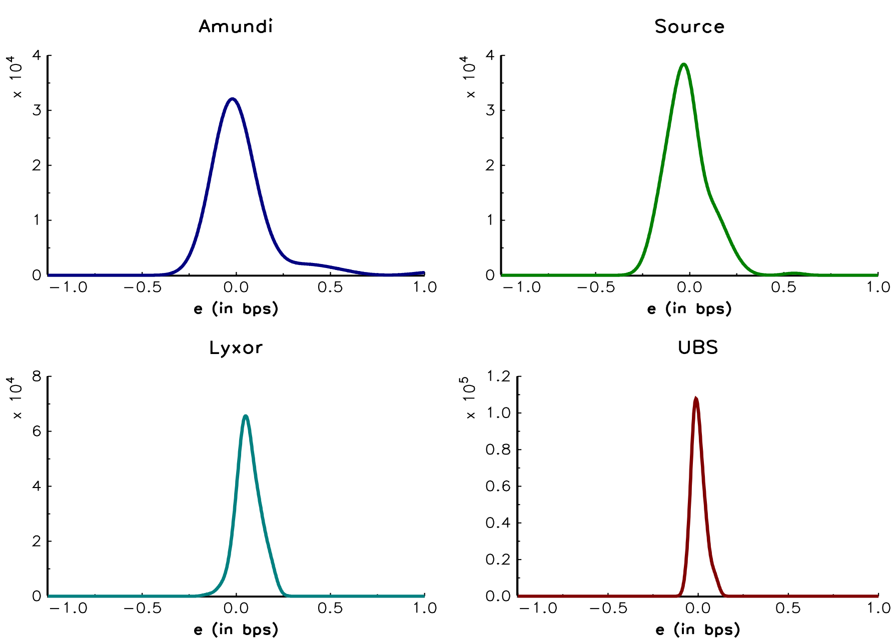


Figure 11: Estimated density of daily tracking errors (S&P 500)



4.2 Choosing another risk measure

The efficiency measure $\zeta_\alpha(x \mid b)$ is based on the assumption that tracking errors are normally distributed. In this paragraph, we explore non-gaussian risk measures. For instance, we can replace gaussian value-at-risk by historical value-at-risk or the expected shortfall. One of the difficulties with these two risk measures is that they are generally computed on a daily basis, but the conversion to an annual measure is not straightforward. In the previous gaussian value-at-risk, we adopt the square root rule: annual volatility of tracking errors is obtained by considering the product of the daily volatility of tracking errors and the factor $\sqrt{260}$. This rule is universally accepted and has been used in fund reports for many years. Unfortunately, the square root of time is appropriate when returns are normally distributed, but may be not so in other cases (McNeil and Frey, 2000). Because there is no solution to this problem, professionals still use the square root rule in the non-gaussian case, but only when returns are centered. Following this common practice, we suggest defining the efficient indicator as follows:

$$\zeta_\alpha^\star(x \mid b) = \mu(x \mid b) - s(x \mid b) - \mathbf{F}_0^{-1}(\alpha)$$

where $\mathbf{F}_0$ is the probability distribution of centered tracking errors.

In Tables 8 and 9, we compute the efficiency indicator $\zeta_\alpha(x \mid b)$ with five risk measures: $\sigma(x \mid b)$ corresponds to volatility, $\sigma_-(x \mid b)$ is below-mean semi-volatility, $\mathrm{VaR}_\alpha(x \mid b)$ is historical value-at-risk, $\mathrm{ES}_\alpha(x \mid b)$ corresponds to the historical expected shortfall, whereas $\mathrm{CF}_\alpha(x \mid b)$ uses the Cornish-Fisher expansion to compute value-at-risk. In the case of Eurostoxx 50 trackers, we observe some minor differences. Indeed, Lyxor turns out to be the best tracker if we are looking at historical value-at-risk. The rank of db X-trackers varies between 1 and 3 depending on the risk measures. For S&P 500 trackers, we observe a good consistency across the various rankings.

Table 8: Efficiency measures of Eurostoxx 50 trackers

Tracker	$\sigma(x \mid b)$	$\sigma_-(x \mid b)$	$\mathrm{VaR}_\alpha(x \mid b)$	$\mathrm{ES}_\alpha(x \mid b)$	$\mathrm{CF}_\alpha(x \mid b)$
Amundi	32.98	44.72	47.75	47.25	62.54
db X-trackers	41.64	44.12	45.90	41.02	47.15
iShares (DE)	−63.38	−64.32	−59.40	−102.51	−62.72
iShares	15.70	20.38	34.77	12.21	27.93
Lyxor	30.47	41.63	47.97	38.26	61.48
Source	−3.83	3.74	5.00	4.63	19.17

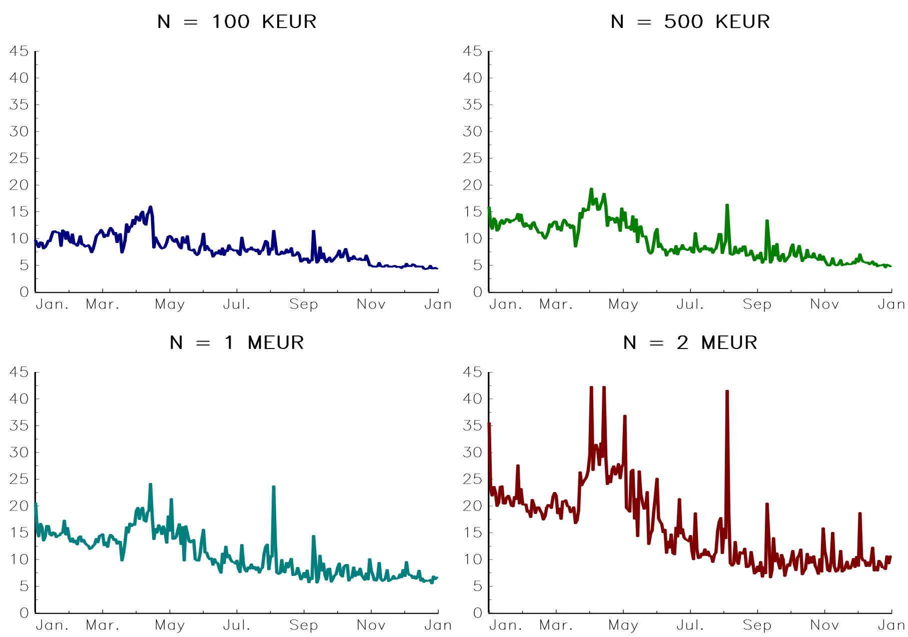
Table 9: Efficiency measures of S&P 500 trackers

Tracker	$\sigma(x \mid b)$	$\sigma_-(x \mid b)$	$\mathrm{VaR}_\alpha(x \mid b)$	$\mathrm{ES}_\alpha(x \mid b)$	$\mathrm{CF}_\alpha(x \mid b)$
Amundi	−12.97	−10.64	−10.24	−10.55	−8.04
Credit Suisse	−8.88	−8.26	−6.60	−10.37	−7.35
db X-trackers	−9.90	−10.13	−8.88	−11.77	−10.02
HSBC	1.92	2.92	4.24	2.23	3.73
iShares	−21.63	−21.15	−19.10	−23.59	−20.23
Lyxor	4.69	4.73	4.88	4.25	4.71
Source	−26.04	−25.67	−25.50	−26.00	−25.37
UBS	−42.09	−41.98	−41.93	−42.09	−41.95

4.3 Taking the liquidity risk into account

The spread measure $s(x \mid b)$ used previously corresponds to the daily average of the first limit order spreads. This measure may be pertinent for a retail investor, but is not for an institutional investor. Institutional investors buy or sell a notional N, that can not generally be executed via the best first limit orders. That is why we consider another spread measure $s_N(x \mid b)$ corresponding to intraday spreads weighted by the duration between two ticks for a given notional[13]. To illustrate the use of this spread, we are going to look at the Eurostoxx 50 trackers.

Figure 12: Evolution of the spread of the Amundi tracker



In Figure 12, we shown the change in the spread $s_N(x \mid b)$ for the Amundi Eurostoxx 50 tracker and different values of the notional N. We notice that this spread changes with respect to market liquidity. For each tracker and each notional value, we have shown the corresponding boxplot in Figure 13. The boxplot indicates the minimum value, the quartile range, the median and the last decile. We check that:

$$N_1 \geq N_2 \Rightarrow s_{N_1}(x \mid b) \geq s_{N_2}(x \mid b)$$

We notice also that for some trackers, the spread increases greatly in line with volume. In Figure 14, we have drawn some scatterplots. It is interesting to note that when liquidity is not an issue, the trackers are more or less equivalent in terms of spread[14]. This is not the case when there are some liquidity problems (see Figure 15).

[13]Computational details are provided in Appendix A.5.

[14]If we consider two trackers x and y, we have:

$$s_N(x \mid b) \simeq s_N(y \mid b)$$

if $s_N(x \mid b) \simeq 0$ and $s_N(x \mid b) \simeq 0$.

Figure 13: Boxplot of ETF spreads

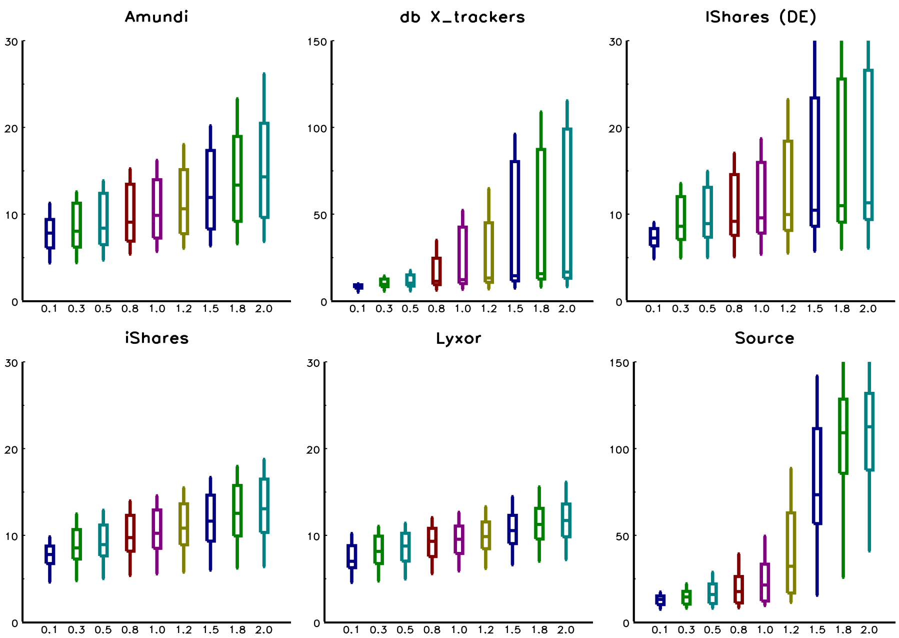


Figure 14: Scatterplot of ETF spreads with a volume of 1 MEUR

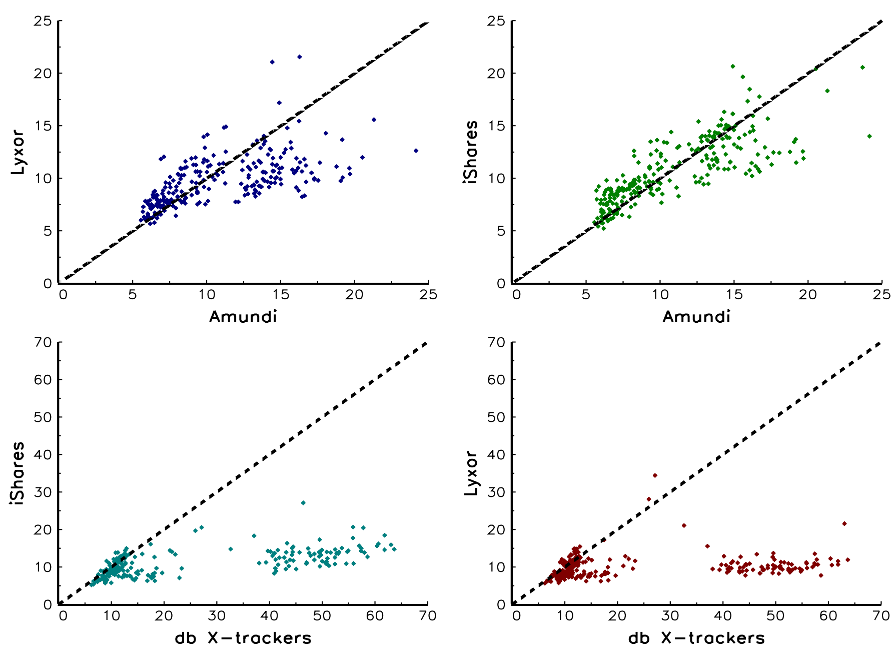
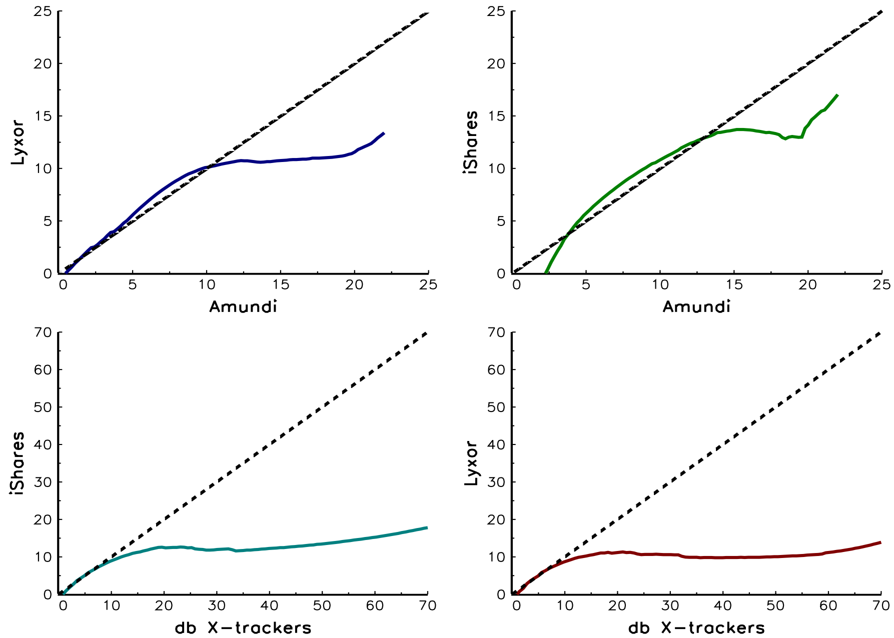


Figure 15: Quantile regression of ETF spreads with a volume of 1 MEUR



Let us now compute the efficiency indicator $\zeta_{\alpha}^{\star}(x \mid b)$ by considering this new spread measure. The results are shown in Table 10. We consider two calculations: one based on the median spread and another one based on the quantile at the 95% confidence level. We notice that this new measure has a big impact on the efficiency indicator. For instance, if we consider a notional of 2 MEUR, we obtain a different ranking.

Table 10: Impact of the liquidity on the efficiency measure (Eurostoxx 50)

Tracker	100 KEUR		1 MEUR		2 MEUR	
	50%	95%	50%	95%	50%	95%
Amundi	35.01	30.19	32.96	25.45	28.52	15.00
db X-trackers	45.48	43.28	41.65	−2.39	37.29	−66.68
iShares (DE)	−62.66	−65.05	−65.00	−77.04	−66.74	−98.73
iShares	18.30	15.72	15.85	10.33	13.04	5.31
Lyxor	31.93	27.89	29.39	24.98	27.23	19.97
Source	−1.46	−8.80	−9.80	−64.29	−101.00	−193.50

Remark 2 *In our definition of the efficiency indicator, we assume that there is one trade each year. If there are m trades per year, the performance measure becomes:*

$$\zeta_{\alpha}(x \mid b) = \mu(x \mid b) - m \cdot s(x \mid b) - \Phi^{-1}(\alpha)\,\sigma(x \mid b)$$

This formula highlights the importance of liquidity for active managers. For instance, a highly active manager will only be interested in the spread measure because:

$$\lim_{m \to \infty} \zeta_{\alpha}(x \mid b) = -m \cdot s(x \mid b)$$

5 Conclusion

In this paper, we have developed a performance measure to compare passive management, in particular tracker investment vehicles such as exchange traded funds. This measure is very different from the ones used to assess the performance of active management. It is a value-at-risk measure based on three parameters: the performance difference between the fund and the index, the volatility of the tracking error and the liquidity spread. This simple measure may be easily implemented by investors or rating agencies in relation to the fund picking process.

This paper also highlights the role of liquidity spread in measuring the efficiency of an ETF. It is particular true for institutional investors, who may subscribe or redeem large investment amounts. The liquidity spread is also the most important parameter for active managers who use ETFs to implement their convictions in line with their tactical asset allocation.

References

[1] Amenc N. and Martellini L. (2002), The Brave New World of Hedge Fund Indices, *EDHEC-Risk Working paper.*

[2] Barras L., Scaillet O. and Wermers R. (2010), False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas, *Journal of Finance*, 65(1), pp. 179-216.

[3] Bawa V.S. (1975), Optimal Rules for Ordering Uncertain Prospects, *Journal of Financial Economics*, 2(1), pp. 95-121.

[4] Beasley J.E., Meade N. and Chang T.-J. (2003), An Evolutionary Heuristic for the Index Tracking Problem, *European Journal of Operational Research*, 148(3), pp. 621-643.

[5] Blake C.R., Elton E.J. and Gruber M.J. (1993), The Performance of Bond Mutual Funds, *Journal of Business*, 66(3), pp. 371-403.

[6] Bernstein P.L. (1992), *Capital Ideas: The Improbable Origins of Modern Wall Street*, Free Press.

[7] Bernstein P.L. (2007), *Capital Ideas Evolving*, John Wiley & Sons.

[8] Carhart M.M. (1997), On Persistence in Mutual Fund Performance, *Journal of Finance*, 52(1), pp. 57-82.

[9] Dieterlen R. and Hereil P. (2012), Index Arbitrage Explained, *Journal of Indexes Europe*, 2(6), pp. 18-23.

[10] Elton E.J., Gruber M.J., Das S. and Hlavka M. (1993), Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, *Review of Financial Studies*, 6(1), pp. 1-22.

[11] Elton E.J., Gruber M.J. and Busse J.A. (2004), Are Investors Rational? Choices among Index Funds, *Journal of Finance*, 59(1), pp. 261-288.

[12] Estrada J. (2007), Mean-semivariance Behavior: Downside Risk and Capital Asset Pricing, *International Review of Economics and Finance*, 16(2), pp. 169-185.

[13] Fishburn P.C. (1977), Mean-Risk Analysis with Risk Associated with Below-Target Returns, *American Economic Review*, 67(1), pp. 116-126.

[14] Gastineau G.L. (2002), Equity Index Funds Have Lost Their Way, *Journal of Portfolio Management*, 28(2), pp. 55-64.

[15] Gastineau G.L. (2004), The Benchmark Index ETF Performance Problem, *Journal of Portfolio Management*, 30(2), pp. 96-103.

[16] Grinblatt M. and Titman S. (1989), Portfolio Performance Evaluation: Old Issues and New Insights, *Review of Financial Studies*, 2(3), pp. 393-421.

[17] Grinold R.C. and Kahn R.N. (2000), *Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk*, Second edition, McGraw-Hill.

[18] HARLOW W.V. (1991), Asset Allocation in a Downside-Risk Framework, *Financial Analysts Journal*, 47(5), pp. 28-40.

[19] JENSEN M.C. (1968), The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance*, 23(2), pp. 389-416.

[20] JORION P. (2003), Portfolio Optimization with Tracking-Error Constraints, *Financial Analysts Journal*, 59(5), pp. 70-82.

[21] MARKOWITZ H. (1952), Portfolio Selection, *Journal of Finance*, 7(1), pp. 77-91.

[22] MARKOWITZ H. (1959), *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons.

[23] MCNEIL A.J. and FREY R. (2000), Estimation of Tail-related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach, *Journal of Empirical Finance*, 7(3-4), pp. 271-300.

[24] NAWROCKI D.N. (1999), A Brief History of Downside Risk Measures, *Journal of Investing*, 8(3), pp. 9-25.

[25] POPE P.F. and YADAV P.K. (1994), Discovering Errors in Tracking Error, *Journal of Portfolio Management*, 20(2), pp. 27-32.

[26] PRIGENT J.-L. (2007), *Portfolio Optimization and Performance Analysis*, Chapman & Hall.

[27] ROLL R. (1992), A Mean/Variance Analysis of Tracking Error, *Journal of Portfolio Management*, 18(4), pp. 13-22.

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

[29] TOBIN J. (1958), Liquidity Preference as Behavior Towards Risk, *Review of Economic Studies*, 25(2), pp. 65-86.

A Technical appendix

A.1 Proof of the equation (1)

We have:

$$\begin{aligned}\sigma^2(x \mid b) &= (x-b)^\top \Sigma (x-b) \\ &= x^\top \Sigma x + b^\top \Sigma b - 2x^\top \Sigma b \\ &= \sigma^2(x) + \sigma^2(b) - 2\rho(x,b)\,\sigma(x)\,\sigma(b)\end{aligned}$$

We deduce that the correlation between the portfolio x and the benchmark b is:

$$\rho(x,b) = \frac{\sigma^2(x) + \sigma^2(b) - \sigma^2(x \mid b)}{2\sigma(x)\,\sigma(b)}$$

It follows that:

$$\begin{aligned}2(x_0-b)^\top \Sigma (x-b) &= 2\left(x_0^\top \Sigma x - x_0^\top \Sigma b - x^\top \Sigma b + b^\top \Sigma b\right) \\ &= \sigma^2(x_0) + \sigma^2(x) - \sigma^2(x \mid x_0) - \\ &\quad \sigma^2(x_0) - \sigma^2(b) + \sigma^2(x_0 \mid b) - \\ &\quad \sigma^2(x) - \sigma^2(b) + 2\sigma^2(x \mid b) + \sigma^2(b) \\ &= \sigma^2(x_0 \mid b) + \sigma^2(x \mid b) - \sigma^2(x \mid x_0)\end{aligned}$$

The weights of the portfolio y are given by:

$$y = (1-\alpha)\,x_0 + \alpha x$$

We deduce that:

$$y - b = (1-\alpha)(x_0 - b) + \alpha(x-b)$$

It follows that:

$$\mu(y \mid b) = (1-\alpha)\,\mu(x_0 \mid b) + \alpha\mu(x \mid b)$$

and:

$$\begin{aligned}\sigma^2(y \mid b) &= (y-b)^\top \Sigma (y-b) \\ &= (1-\alpha)^2 (x_0-b)^\top \Sigma (x_0-b) + \\ &\quad \alpha^2 (x-b)^\top \Sigma (x-b) + \\ &\quad 2\alpha(1-\alpha)(x_0-b)^\top \Sigma (x-b) \\ &= (1-\alpha)^2 \sigma^2(x_0 \mid b) + \alpha^2\sigma^2(x \mid b) + \\ &\quad \alpha(1-\alpha)\left(\sigma^2(x_0 \mid b) + \sigma^2(x \mid b) - \sigma^2(x \mid x_0)\right) \\ &= (1-\alpha)\,\sigma^2(x_0 \mid b) + \alpha\sigma^2(x \mid b) + \left(\alpha^2-\alpha\right)\sigma^2(x \mid x_0)\end{aligned}$$

We deduce that:

$$\begin{aligned}\mathrm{IR}(y \mid b) &= \frac{\mu(y \mid b)}{\sigma(y \mid b)} \\ &= \frac{(1-\alpha)\,\mu(x_0 \mid b) + \alpha\mu(x \mid b)}{\sqrt{(1-\alpha)\,\sigma^2(x_0 \mid b) + \alpha\sigma^2(x \mid b) + (\alpha^2-\alpha)\,\sigma^2(x \mid x_0)}}\end{aligned}$$

A.2 Derivation of the efficiency measure $\zeta_\alpha(x \mid b)$

The probability distribution function of the tracking error e is Gaussian with:

$$e \sim \mathcal{N}\left(\mu(x \mid b), \sigma^2(x \mid b)\right)$$

We have:

$$\begin{aligned}
& \Pr\{\mathcal{L}(x \mid b) \leq \zeta\} = \alpha \\
\Leftrightarrow\quad & \Pr\{e - s \leq \zeta\} = \alpha \\
\Leftrightarrow\quad & \Pr\{e \leq s + \zeta\} = \alpha \\
\Leftrightarrow\quad & \Pr\left\{\frac{e - \mu(x \mid b)}{\sigma(x \mid b)} \leq \frac{s + \zeta - \mu(x \mid b)}{\sigma(x \mid b)}\right\} = \alpha \\
\Leftrightarrow\quad & \Phi\left(\frac{s + \zeta - \mu(x \mid b)}{\sigma(x \mid b)}\right) = \alpha
\end{aligned}$$

It follows that:

$$\frac{s + \zeta - \mu(x \mid b)}{\sigma(x \mid b)} = \Phi^{-1}(\alpha)$$

or:

$$\zeta = s - \mu(x \mid b) + \Phi^{-1}(\alpha)\,\sigma(x \mid b)$$

We deduce that:

$$\begin{aligned}
\zeta_\alpha(x \mid b) &= -\zeta \\
&= \mu(x \mid b) - s - \Phi^{-1}(\alpha)\,\sigma(x \mid b)
\end{aligned}$$

A.3 Building a benchmark from a set of competing indices

Let $\{b_1, \ldots, b_m\}$ be a set of competing indices. The underlying idea is to compute a benchmark b representative of these indices. Let Ω be the $m \times m$ covariance matrix of the returns $R_t(b_j)$. By computing the eigendecomposition of Ω, we get:

$$\Omega = V \Lambda V^\top \tag{3}$$

where V is the matrix of the eigenvectors and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ is a diagonal matrix of eigenvalues[15]. Because the decomposition (3) corresponds to the principal component analysis (PCA), we may extract the main representative factor defined by:

$$F_t = \sum_{j=1}^{m} V_{j,1} R_t(b_j)$$

As noticed by Amenc and Martellini (2002), this first component of the PCA maximizes the representativeness of the set of competing indices. The reference index is then a weighted portfolio of the competing indices:

$$b = \sum_{j=1}^{m} w_j b_j$$

where the weights are proportional to the loading coefficients:

$$w_j = \frac{V_{j,1}}{\sum_{k=1}^{m} V_{k,1}}$$

[15] We assume that $\lambda_1 \geq \ldots \geq \lambda_m$.

A.4 Lower partial moments

Let X be a random variable with distribution $\mathbf{F}$. The mathematical expectation of X is[16]:

$$\begin{aligned}\mu &= \mathbb{E}[X] \\ &= \int_{-\infty}^{\infty} x \,\mathrm{d}\mathbf{F}(x) \\ &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x\end{aligned}$$

whereas the centered moment of order n is:

$$\begin{aligned}\mu_n(X) &= \mathbb{E}[(X-\mu)^n] \\ &= \int_{-\infty}^{\infty} (x-\mu)^n f(x)\,\mathrm{d}x\end{aligned}$$

The variance of X corresponds to the second moment: $\sigma^2(X) = \mu_2(X)$. $\sigma(X)$ is called standard deviation, but it is well known as volatility in finance, where X represents the return on an asset.

The lower partial moment (LPM) is obtained by (Bawa, 1975):

$$\begin{aligned}\mathrm{LPM}_n(X;\tau) &= \mathbb{E}[\max(0,\tau-X)^n] \\ &= \int_{-\infty}^{\tau} (\tau-x)^n f(x)\,\mathrm{d}x\end{aligned}$$

where τ is a threshold. If τ is equal to the mean μ, we get:

$$\begin{aligned}\mathrm{LPM}_n(X;\mu) &= \mathbb{E}[\max(0,\mu-X)^n] \\ &= \int_{-\infty}^{\mu} (\mu-x)^n f(x)\,\mathrm{d}x\end{aligned}$$

Semi-variance is then defined as the second lower partial moment:

$$\begin{aligned}\mathrm{SV}(x) &= \mathrm{LPM}_2(X;\mu) \\ &= \mathbb{E}\left[\max(0,\mu-X)^2\right] \\ &= \int_{-\infty}^{\mu} (\mu-x)^2 f(x)\,\mathrm{d}x\end{aligned}$$

If the distribution of X is symmetric around the mean, we get:

$$\begin{aligned}\int_{-\infty}^{\infty} (\mu-x)^2 f(x)\,\mathrm{d}x &= \int_{-\infty}^{\mu} (\mu-x)^2 f(x)\,\mathrm{d}x + \int_{\mu}^{\infty} (\mu-x)^2 f(x)\,\mathrm{d}x \\ &= 2\int_{-\infty}^{\mu} (\mu-x)^2 f(x)\,\mathrm{d}x\end{aligned}$$

We deduce that semi-variance is half of the variance.

[16] We note $f(x)$ the associated density function.

A.5 Computation of the spread $s_N(x \mid b)$

A.5.1 Analytical expression

We define the daily spread $s_N(x \mid b)$ as a weighted average of intraday spreads:

$$s_N(x \mid b) = \frac{\sum_{j=\text{open}}^{\text{close}} s_j (t_{j+1} - t_j)}{\sum_{j=\text{open}}^{\text{close}} (t_{j+1} - t_j)}$$

where s_j is the spread of the j^{th} tick and $t_{j+1} - t_j$ the elapsed time between two consecutive ticks:

$$s_j = c_j \frac{\left(\bar{P}_j^+ - \bar{P}_j^-\right)}{\bar{P}_j^0}$$

We have also:

$$\bar{P}_j^\bullet = \frac{\sum_{k=1}^K \bar{Q}_{j,k}^\bullet P_{j,k}^\bullet}{\sum_{k=1}^K \bar{Q}_{j,k}^\bullet}$$

where $P_{j,k}^+$ (resp. $P_{j,k}^-$) is the ask (or bid) price at t_j for the k^{th} limit order. The average mid price $\bar{P}_j^0$ corresponds to:

$$\bar{P}_j^0 = \frac{\bar{P}_j^+ + \bar{P}_j^-}{2}$$

The quantity $\bar{Q}_{j,k}^+$ and $\bar{Q}_{j,k}^-$ are defined as follows:

$$\bar{Q}_{j,k}^\bullet = \max\left(0, \min\left(Q_{j,k}^\bullet, Q_j^\star - \sum_{l=1}^{k-1} Q_{j,l}^\bullet\right)\right)$$

Here, $Q_{j,k}^+$ and $Q_{j,k}^-$ are the ask and bid volumes of the k^{th} limit order. The reference quantity $Q_j^\star$ is the ratio between the trading notional N and the mid price:

$$Q_j^\star = \frac{N}{\bar{P}_j^0}$$

Sometimes it may appear that the trading volume on the order book is lower than the notional N. That is why the factor c_j may be greater than one:

$$c_j = \max\left(1, \frac{Q_j^\star}{\min\left(\sum_{k=1}^K Q_{j,k}^+, \sum_{k=1}^K Q_{j,k}^-\right)}\right)$$

For instance, if we wish to execute an order of 2 MEUR and there is only a trading volume of 1 MEUR, we multiply the spread by two.

Remark 3 *For each trading day, we compute the daily spread for the different listing places using the previous formulas and we take the best spread.*

A.5.2 Example

Let us illustrate the spread calculation with the order book given in Table 11. For $Q^\star = 1000$, we get the results shown in Table 12. We deduce that the mid price $\bar{P}_j^0$ is 85.98 whereas the spread s_j is equal to 20.12 bps. It corresponds to a notional $N = Q^\star \times \bar{P}_j^0$ of 85 981 €.

Table 11: Limit order book

k	Buy orders $Q^-_{j,k}$	Buy orders $P^-_{j,k}$	Sell orders $Q^+_{j,k}$	Sell orders $P^+_{j,k}$
1	900	85.90	600	86.05
2	200	85.85	300	86.06
3	57	85.82	400	86.20
4	18	85.75	213	86.21
5	117	85.74	73	86.22
6	1000	85.73	200	86.23
7	3000	85.72	1500	86.25

Table 12: Computing the spread for $Q^\star = 1000$

k	Buy orders $\bar{Q}^-_{j,k}$	Buy orders $P^-_{j,k}$	Sell orders $\bar{Q}^+_{j,k}$	Sell orders $P^+_{j,k}$
1	900	85.90	600	86.05
2	100	85.85	300	86.06
3	0	85.82	100	86.20
$\sum_{k=1}^{K} \bar{Q}^\bullet_{j,k}$	1000		1000	
$\bar{P}^\bullet_j$		85.89		86.07

If we wish to execute an order for 100 KEUR, it implies that $Q^\star > 1000$. Given a notional N, it is then possible to find the optimal value of $Q^\star$:

$$Q^\star = \inf \left\{ Q \in \mathbb{N} : Q\bar{P}^0_j \geq N \right\}$$

We may solve this nonlinear inequality by using a bisection algorithm. For instance, if N is equal to 100 KEUR, we obtain $Q^\star = 1163$. In this case, the spread is equal to 23.24 bps. Sometimes, there are not enough orders in the book. For instance, if $N = 500$ KEUR, the mid price is 85.972 meaning that $Q^\star = 5816$. The coefficient c is then larger than 1 and we obtain a spread equal to 87.81 bps.

Table 13: Computing the spread for a given notional N

k	$N = 100$ KEUR				$N = 500$ KEUR			
	Buy orders		Sell orders		Buy orders		Sell orders	
	$\bar{Q}^-_{j,k}$	$P^-_{j,k}$	$\bar{Q}^+_{j,k}$	$P^+_{j,k}$	$\bar{Q}^-_{j,k}$	$P^-_{j,k}$	$\bar{Q}^+_{j,k}$	$P^+_{j,k}$
1	900	85.90	600	86.05	900	85.90	600	86.05
2	100	85.85	300	86.06	200	85.85	300	86.06
3	57	85.82	263	86.20	57	85.82	400	86.20
4	6	85.75	0	86.21	18	85.75	213	86.21
5	0	85.74	0	86.22	117	85.74	73	86.21
6	0	85.73	0	86.23	1000	85.73	200	86.23
7	0	85.72	0	86.25	3000	85.72	1500	86.25
$\sum_{k=1}^{K} \bar{Q}^\bullet_{j,k}$	1163		1163		5292		3286	
$\bar{P}^\bullet_j$		85.89		86.09		85.76		86.19

B Description of the data

In the following table, for each universe we show the name of the tracker, its description and the corresponding Bloomberg code.

Universe	Name	Description	BBG code
Eurostoxx 50	Amundi	AMUNDI ETF DJE 50	C50 FP
	db X-trackers	DBX-TRACKERS EURO STXX 50	XESX GY
	iShares (DE)	ISHARES EURO STOXX 50 DE	SX5EEX GY
	iShares	ISHARES EURO STOXX 50	EUN2 GY
	Lyxor	LYXOR ETF Euro Stoxx 50	MSE FP
	Source	EURO STOXX 50 Source ETF	SDJE50 GY
	Eurostoxx 50	ESTX 50 € NRt	SX5T
S&P 500	Amundi	AMUNDI ETF S&P 500	500 FP
	Credit Suisse	CSETF ON S&P 500	CSSPX SW
	db X-trackers	DB X-TRACKERS S&P 500 ETF	D5BM GY
	HSBC	HSBC S&P 500 ETF	HSPD LN
	iShares	ISHARES S&P 500 Index Fund	IUSA LN
	Lyxor	LYXOR ETF S&P 500	SP5 FP
	Source	S&P 500 SOURCE ETF	SPXS LN
	UBS	UBS S&P 500 INDEX-A	S5USAS SW
	S&P 500	S&P 500 Net TR	SPTR500N
MSCI World	Amundi	AMUNDI ETF MSCI WORLD	CW8 FP
	Commerzbank	CCOMSTAGE ETF MSCI WORLD-I	CBNDDUWI GY
	db X-trackers	DB X-TRACKERS MSCI WORLD TRN	XMWO GY
	iShares	ISHARES MSCI World	IWRD LN
	Lyxor	LYXOR ETF MSCI World	WLD FP
	Source	MSCI-WORLD SOURCE ETF	SMSWLD GY
	UBS	UBS-ETF MSCI WORLD	WRDCHA SW
	MSCI World	MSCI Daily TR Net World USD	NDDUWI
MSCI EM	Credit Suisse	CS ETF ON MSCI EMERGING MRKT	CSEM SW
	db X-trackers	DB X-TRACKERS EMERG MARKET	XMEM GY
	iShares	iShares MSCI Emerging Markets (EUR)	IEMM NA
	Lyxor	LYXOR ETF MSCI Emerging Markets	LEM FP
	Source	MSCI EMERG MARKET SOURCE ETF	MXFS LN
	MSCI EM	MSCI Daily TR Net Emerging Markets	NDUEEGF
Japanese Equities	db X-trackers	DB X-TRACKERS MSCI JAPAN TRN	XMJP GY
	iShares	ISHARES MSCI JAPAN FUND	IJPN LN
	Lyxor	LYXOR ETF Japan (TOPIX)	JPN FP
	Source	MSCI JAPAN SOURCE ETF	SMSJPN GY
	Topix	TOPIX TR Index	TPXDDVD
	MSCI Japan	MSCI Daily TR Gross Japan USD	GDDUJN