Exchange Traded Funds: Toward a Tailored Selection Approach*

Maxime Bonelli†

Working Paper: April 2015

Abstract

The Exchange Traded Fund market is growing, and passive investors need to have reasonable grounds for choosing this investment vehicle as well as robust selection criteria. The author motivates the assertion that a tracker performance must be assessed relative to its benchmark index and not be based on absolute returns nor fund characteristics, such as expense ratio. Furthermore, the different indicators of tracking quality used in the industry, as well as some variations, are reviewed. Using examples of three major equity indices, the author argues that Exchange Traded Fund selection cannot be based on a single performance measure assessing quality common to all investors due to their various specific investment purposes and the divergent ETFs rankings according to the different indicators. Therefore, the author highlights that Exchange Traded Fund selection should be carried in an investor-specific framework founded on statistical measurements related to tracking quality relevant for each investor.

Keywords: Passive investment, ETF, tracking error, tracking difference.
JEL classifications: G00, G11.

*The author would like to thank Jean-René Giraud for valuable comments. Financial support from Koris International is gratefully acknowledged.

†Research & Development at Koris International and PhD student at Inria Sophia Antipolis Méditerranée Research Center. **Email**: maxime.bonelli@koris-intl.com, **Address**: Koris International, 200 avenue de Roumanille, Espace Saint Philippe, 06410 Biot, France.

1 Introduction

The asset management industry is becoming increasingly segmented between passive versus active management. Passive investor strategy consists of investing in index trackers instead of actively managed funds. The seminal academic studies for this approach are the market portfolio concept of Markowitz (1952) and the Capital Asset Pricing Model (CAPM) of Sharpe (1964). Furthermore, the efficient market hypothesis developed in Fama (1965) also supports this direction. From passive investors' standpoint, selecting the proper index tracker is the central question especially because no standard selection methodology has emerged in the industry. Still, many passive managers argue that fees and absolute performance are the primary selection criteria. However, for passive investors, absolute performance has less interest, as they would like to have a delta-one exposure to an index, i.e., trade an asset having the same return as its benchmark. Thus, the selection criteria are different from those used by active managers and must necessarily include an analysis of the fund with respect to its index.

In practice, the main passive investment vehicles are index mutual funds and Exchange Traded Funds (ETFs). The latter have been used since the launch of the SPDR in 1993, whereas the first index mutual fund was introduced earlier around 1975. ETFs represent a true alternative to index mutual funds due to their flexibility. For managers, they offer in particular the possibility to be purchased in smaller sizes and continuously during the trading day[1]. Moreover, there exist ETFs relative to benchmarks, where there are no index mutual funds or futures contracts. Several studies discuss index mutual funds performance (see Frino and Gallagher, 2001; Gastineau, 2002) and their comparison with Exchange Traded Funds (see Gastineau, 2004; Kostovetsky, 2003; Svetina and Wahal, 2010). Regarding the assessment of ETF performance, Hassine and Roncalli (2013) recently introduced an ETF efficiency measure based on the value-at-risk framework. Amenc et al. (2012) analyze the relative risk of alternative indices using high quantiles of the distribution of return differences.

In this paper, we discuss the different indicators of tracking quality used in the ETF industry. Using trackers of three major equity indices of different market segments, we show that rankings of ETFs differ according to the different indicators. Therefore, we motivate that a single performance measure cannot be used in Exchange-Traded Funds selection by all passive investors. The main reason supporting this view is the plurality of the latter: some invest for short-term trading or hedging purposes, some for long-term investments, some want to be dynamic, while others want to buy and hold, and the behavior seriously influences the analysis that ought to be made on the ETF candidates. Thus, ETFs selection should be performed evaluating various indicators relevant to each investor depending on his/her specific characteristics and constraints. In the next section, we discuss statistics of interest and illustrate our motivation. Additionally, we show that ETF features, such as expense ratio (administration and management fees known ex-ante), and the replication model (physical or synthetic), are often not indicative of the tracking quality and thus cannot be considered as robust selection criteria. We add that trading analysis is a problem that should not be neglected and remains open.

2 Exchange Traded Fund selection

The first step of the ETF selection procedure is to select the tracked index: knowing the universe covered, understanding how it is defined and rebalanced, how it deals with dividend distribution, and the associated tax implications are essential components. On this last point, investors should be aware

[1]For a description of how ETFs work, see Gastineau (2001) and Fuhr (2001).

of the three main types of benchmark indices in terms of dividend treatment. The first type is *Price Index*, i.e., ex-dividends, that denotes indices strictly taking the price of their underlyings without making any hypothesis of dividends/coupons reinvestment or withholding tax. *Net Total Return* indicates indices including a given percentage of dividends: the index provider makes an assumption of reinvestment of its underlyings dividends/coupons, for which it makes an extra assumption of withholding tax. Finally, *Gross Total Return* indices include 100% of dividends without taking into account any withholding tax. Of course, using a *Net Total Return* index as benchmark will allow the ETF manager to benefit from tax optimization techniques; that is why investors should pay close attention to the dividends treatment policy of their benchmark. The development of "smart beta" indices offering alternative weighted schemes (different from the standard market-capitalization weighted indices) renders the index selection task even more challenging. See Amenc et al. (2012) and Hsu et al. (2011) on the topic of smart beta. In sum, the choice of the index will be highly responsible for the absolute return the investor will obtain. Hence, as mentioned earlier, the analysis of the performance of a tracker in itself must be performed with respect to its benchmark index: an ETF based on an index with low returns will logically generate low absolute returns. However, it can have a better tracking quality than another ETF based on a different index with greater absolute returns.

Quantitative measures, such as *Tracking Difference* (TD) and *Tracking Error* (TE), have been introduced both in industry and academia in order to assess ETF tracking quality. TD and TE are defined respectively as: the difference between the total return of the fund and the total return of the benchmark index, and the volatility of the differences in returns between a fund and its benchmark[2]. In order to avoid confusion, we point out that several other research investigations denote returns differences between the tracker and its benchmark as *"tracking error"* and thus its volatility as *"tracking error volatility"* (e.g., Basak et al., 2006; Hassine and Roncalli, 2013; Roll, 1992). In the rest of this paper, we will use the first nomenclature *Tracking Difference* (TD) and *Tracking Error* (TE) defined above. The main reason is that this notation agrees with the IOSCO CR05/12 consulting report and the ESMA (European Securities and Markets Authority) report *"Guidelines on ETFs and other UCITS issues."*

2.1 Data

In order to illustrate our analysis with concrete examples and to enhance the necessity of using different indicators in the ETF selection procedure depending on the investor profile, we will consider three equity indices and five European ETFs (i.e., funds domiciled in Europe) tracking each of them. For each index, we consider ETFs with a creation date earlier or equal to January 2011, and we select five funds among the ten with the highest market capitalization. In the interests of fairness and in order to not promote any particular provider, ETFs considered are anonymized[3] and simply denoted by a letter from A to E. The three benchmark indices that will be used are as follows: 1) MSCI World Daily Total Return Net Index - USD, 2) MSCI Emerging Markets Net Total Return Index - USD, and 3) MSCI Daily Net Total Return Europe Euro Index - EUR. Data are obtained from Bloomberg and correspond to daily NAV total returns of ETFs (data type: DAY_TO_DAY_TOT_RETURN_GROSS_DVDS, corresponding to Net Asset Value with dividends reinvested) and daily last prices of indices (data type:

[2] Roll (1992), Pope and Yadav (1994), Larsen and Resnick (1998), and Gallagher and Segara (2005) use different methods to evaluate tracking error.

[3] Data are available upon request to the author.

PX_LAST) from 1st January 2011 to 31th July 2014[4]. Characteristics, such as expense ratio and replication model (physical or synthetic) of ETFs considered, are presented in Table 1. We remind readers that expense ratio includes administration and management fees but excludes transaction fees. Regarding the replication model, physical (or full) replication consists of investing in all underlying assets of the benchmark index, whereas synthetic replication implies enter into a swap contract with an investment bank paying the index return against fees (not included in expense ratio) and returns of collateral assets. Looking through Table 1, we can observe that we cannot establish a general rule about replication model, administration and management fees: there is not a single systematically cheaper methodology.

Panel 1: ***MSCI World Daily Total Return Net Index - USD***		
ETF	**Expense Ratio**	**Replication Model**
ETF A	0.50%	Physical
ETF B	0.45%	Synthetic
ETF C	0.20%	Physical
ETF D	0.30%	Physical
ETF E	0.45%	Synthetic
Panel 2: ***MSCI Emerging Markets Net Total Return Index - USD***		
ETF	**Expense Ratio**	**Replication Model**
ETF A	0.68%	Physical
ETF B	0.20%	Synthetic
ETF C	0.68%	Physical
ETF D	0.65%	Synthetic
ETF E	0.45%	Physical
Panel 3: ***MSCI Daily Net Total Return Europe Euro Index - EUR***		
ETF	**Expense Ratio**	**Replication Model**
ETF A	0.35%	Physical
ETF B	0.30%	Synthetic
ETF C	0.30%	Physical
ETF D	0.28%	Synthetic
ETF E	0.33%	Physical

Table 1: Expense ratio and replication model of ETFs tracking the three indices considered.

[4]We download indices prices and not returns in order to calculate ourselves the indices returns the trading days of the considered ETF.

2.2 Relative risk

In this section, we focus on the analysis of the tracking error (TE). This indicator is defined as the volatility of return difference between the tracker and its benchmark index and is a well-known statistic as most investors rely on tracking error as the main measure of replication quality. By definition, it gives a measure of the dispersion of the return difference observations around the average return difference. Intuitively, the higher the tracking error, the higher the risk of incurring a return gap with respect to the index different from the previous one. In practice, computing tracking error can be useful to monitor if it stays below a target level and thus if it corresponds to a degree of relative risk suitable to the investor. Computed daily, it offers a reliable measure of the risks faced by investors to suffer from a gap with the index when trading at Net Asset Value (NAV). The calculation can be easily done using historical data (ex-post estimate). Moreover, a tracking error confidence interval can be derived assuming a particular distribution for return difference. If Gaussian, the tracking error confidence interval is a function of the tracking error sample estimate, the sample size, and chi-squared distribution quantiles[5]. The width of the interval tells how the TE estimate is likely to be the exact volatility of the distribution generating returns observations. Nevertheless, tracking error is not representative of the under or outperformance of the ETF with respect to its index. In particular, an ETF can have a strictly zero tracking error and underperform its benchmark every day by a constant non-zero percentage[6]. Hence tracking error is not intended to measure relative performance of the tracker but instead the relative risk of variation of daily deviations. This indicator is thus of interest and appropriate for short-term and quite dynamic investors selection. However, TE should not be perceived as an issue for mid and long-term investors, which remains the norm in the institutional space, although too high tracking error levels might induce unwanted risk at the time of trading the funds.

We compute daily tracking error for each ETF relative to its benchmark index mentioned in section 2.1. The results are annualized over a 252-day basis and presented in Table 2. Moreover, we provide the ranking of each ETF among others tracking the same index, according to this indicator. Two key findings can be derived. First, the levels of daily tracking error vary significantly across the different ETFs tracking the same index. Indeed, in Panel 1 of Table 2, we can see that the tracking error varies from 0.02% to 0.22% for ETFs tracking the MSCI World Index. The results are even more striking in Panel 2 regarding the MSCI Emerging Markets Index, where tracking errors range from 0.08% to 1.56%[7]. Finally, the MSCI Europe (Panel 3) is the index for which ETF tracking errors are the closest related to each other: 0.04% to 0.14%. The second finding concerns the link between expense ratio (reported in Table 1) and tracking error: there is no clear and direct relation between the levels of expense ratio and ETFs tracking errors. In other words, low (resp. high) expense ratio is not necessarily synonym of low (resp. high) tracking error. Indeed, for the three equity indices considered, the ETF with the lowest expense ratio is never the one with the lowest TE (except ETF D for the MSCI Europe Index). Taking the MSCI World Index for example, we can see that the fund with the lowest fees (0.20%) has the fourth highest tracking error. However, we note that for both the MSCI World and the MSCI Emerging Markets, ETFs with the highest expense ratio, respectively, ETF A with 0.50% and ETF C with 0.68%, have the highest tracking error: 0.22% and 1.56%. Hence, our conclusion is that low management and administrative costs are not sufficient for ensuring

[5]See Cochran (1934) for the derivation of a confidence interval for the statistics of samples from Gaussian distributions.

[6]Similarly in absolute terms, an asset with a constant daily return of $-x\%$ (with $x > 0$) will have a zero volatility estimate.

[7]Given that MSCI Emerging Markets Index relates to fragmented exchanges with multiple currencies, incompatible clearing and settlement channels, it is a priori the most difficult index to replicate in our sample.

low tracking error, but it seems that it helps to maintain a correct TE level with respect to other ETFs charging significantly higher fees. We also assert that for all indices considered, ETFs using a synthetic replication model (i.e., indirect exposure) systematically present lower tracking error than ETFs using a physical replication. Even thought our example concerns only three equity benchmarks and does not allow us to generalize this last conclusion, our explanation is that daily deviations are smoothed and tracking error risk is transferred to the investment bank through the swap contract concluded by providers of synthetic ETFs. Thus, investors looking for ETFs presenting low tracking error should definitely monitor closely the replication method of their ETF candidates. We add that in order to further monitor the risk of daily deviations, looking carefully to extreme quantiles of daily return difference of an ETF with respect to its benchmark can also be of great interest. Indeed, TE is by definition a measure of variation, while a short-term investor may also be interested in the worst (daily) deviations he or she can undergo.

Panel 1: ***MSCI World Daily Total Return Net Index - USD***		
ETF	**Tracking Error**	**ETF Ranking**
ETF A	0.22%	5
ETF B	0.03%	2
ETF C	0.21%	4
ETF D	0.17%	3
ETF E	0.02%	1
Panel 2: ***MSCI Emerging Markets Net Total Return Index - USD***		
ETF	**Tracking Error**	**ETF Ranking**
ETF A	1.37%	4
ETF B	0.42%	2
ETF C	1.56%	5
ETF D	0.08%	1
ETF E	1.33%	3
Panel 3: ***MSCI Daily Net Total Return Europe Euro Index - EUR***		
ETF	**Tracking Error**	**ETF Ranking**
ETF A	0.08%	2
ETF B	0.04%	1
ETF C	0.14%	3
ETF D	0.04%	1
ETF E	0.08%	2

Table 2: Tracking errors and ranking of ETFs considered. Tracking errors are computed using daily data from 2011-01-01 to 2014-07-31 and annualized over 252 days.

2.3 Relative performance

Unlike tracking error, the tracking difference (TD) is set to assess the potential under-performance, or under certain circumstances, the potential out-performance of the tracker with respect to its benchmark. This indicator is defined as the cumulative total return difference between an ETF and its index. This statistic is generally computed on the whole available sample and then annualized for standards issues. Therefore, for a long-term investor, looking at tracking difference of ETF candidates with a long track record is particularly adapted for its selection. Its level provides a very solid indication of the total costs of holding the instrument. Indeed, if replication was perfect, tracking difference and expense ratio (administration and management fees set ex-ante) would be exactly equal in magnitude; thus, low expense ratio would imply low (absolute) tracking difference. Figure 1 shows a scatter plot of ETF expense ratio versus tracking error and the straight line of equation $y = -x$. The latter would correspond to the "perfect replication line". ETFs that are above (resp. below) deliver a greater (resp. lower) performance than what an investor would expect given the expense ratio. Readers should not be surprised that some ETFs manage to lie above this line. Indeed, we only consider *Net Total Return* indices meaning that ETF benchmarks do not include 100% of dividends. Therefore, we observe that on the MSCI World Index, ETFs considered are quite close to the continuous line. The only ETF below (D) presents a tracking difference of -0.43% for an expense ratio of 0.30%. The situation is quite different for the MSCI Emerging Markets Index. Indeed, no ETFs are above the line $y = -x$. Furthermore, ETFs exhibit tracking difference far away from their expense ratio in magnitude, especially ETF B with TD equals -0.65% and fees to 0.20%. Finally, looking at Panel 3 of Figure 1 corresponding to ETFs tracking the MSCI Europe Index, we can observe much better results. All ETFs manage to lay above the continuous line, and some even achieve a positive tracking difference (ETFs A, C, and D), thereby generating a positive excess return with respect to the benchmark. Overall, ETFs with the lowest expense ratio (on the left of different Panels of Figure 1) are not always above others (i.e., with "better" tracking difference). The first conclusion of this analysis is thus similar to section 2.2: even if low expense ratio clearly helps, the management quality seems to make the difference between two funds charging the same expense ratio[8].

Similarly to the previous section, we present in Table 3 the tracking differences of ETFs considered as well as their ranking according to this indicator. Our second conclusion is that the ranking according to the tracking difference is not the same as that obtained using the tracking error in Table 2. Meaning that an investor caring about TD will not select the same ETFs as one interested in low tracking error. In particular, we note that for all indices, the ETF with the lowest tracking error (i.e. ranked first in Table 2) has one of the worst tracking differences and is thus, in the last three positions of Table 3. We also add that the replication model (physical or synthetic) does not have an influence on the ranking of ETFs according to the tracking difference, meaning they both face comparable friction and trading costs.

[8]Elton et al. (2004) addresses this issue considering S&P 500 index funds.

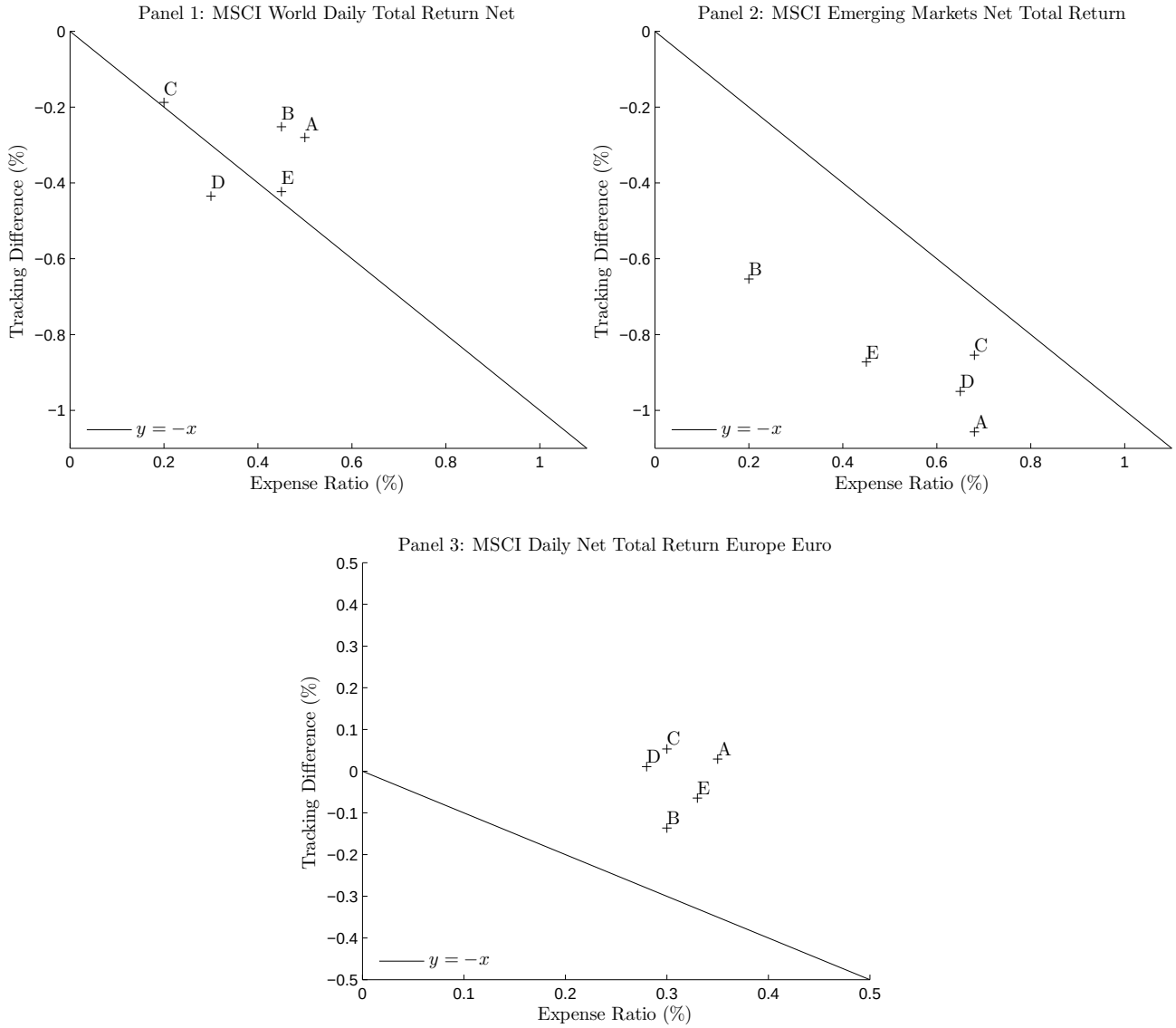

Figure 1: Scatter plot representing expense ratio against tracking difference of ETFs considered. Tracking differences are computed using daily data from 2011-01-01 to 2014-07-31 and annualized over 252 days.

Panel 1: ***MSCI World Daily Total Return Net Index - USD***		
ETF	**Tracking Difference**	**ETF Ranking**
ETF A	-0.28%	3
ETF B	-0.25%	2
ETF C	-0.19%	1
ETF D	-0.43%	5
ETF E	-0.42%	4
Panel 2: ***MSCI Emerging Markets Net Total Return Index - USD***		
ETF	**Tracking Difference**	**ETF Ranking**
ETF A	-1.06%	5
ETF B	-0.65%	1
ETF C	-0.85%	2
ETF D	-0.95%	4
ETF E	-0.87%	3
Panel 3: ***MSCI Daily Net Total Return Europe Euro Index - EUR***		
ETF	**Tracking Difference**	**ETF Ranking**
ETF A	0.03%	2
ETF B	-0.14%	5
ETF C	0.05%	1
ETF D	0.01%	3
ETF E	-0.06%	4

Table 3: Tracking differences and ranking of ETFs considered. Tracking differences are computed using daily data from 2011-01-01 to 2014-07-31 and annualized over 252 days.

As any investment vehicle, ETF performance varies over time and may not be consistent; thus, even if tracking difference is indicative of the global relative performance of an ETF on the whole period considered, it neither reflect its stability over time nor the robustness of the investment process. Indeed, sub-periods of under-performance can be hidden by outperformance sub-periods. For instance, just consider an ETF with a track record of two years whose ordinary return is +6.01% for year 1 and +4.00% for year 2. Suppose its corresponding index benchmark has performed +5.00% for both years 1 and 2. Then, the tracking difference estimate using the whole sample will be close to nil (as the cumulative return of the two assets are 10.25%); however, an investor buying the tracker at the beginning of year 2 will suffer an excess return of -1.00% with respect to the benchmark. Therefore a TD equal to zero does not mean that the tracker has not significantly under-performed its benchmark during some sub-periods of the sample. Of course, if the corrections occur over a short period of time, this drag is less problematic but as illustrated in the simple example above, investors can be aggrieved according to their time of purchase.

One way to avoid missing out on these sub-periods is to look, at each point in time, to the trailing tracking difference with a look-back period of interest (e.g., one year). In other words, compute on each date of the available sample the TD using a moving window on the last H days (where H is the number of days of the investment horizon). Note that if this calculation is done using a growing window instead of a rolling one, it would not be specific to the investment horizon and thus dilute the potential under-performance periods. Moreover, this method delivers a more accurate view on the risk of divergence in returns on the horizon of interest (specific to the investor) rather than computing tracking difference only every H days to match the relevant period. In particular, a dynamic investor likely to frequently buy and sell ETF shares will prioritize funds with stable rolling tracking difference even if globally, the TD is lower on the period. This choice would avoid a buy high/sell low situation in relative terms. Figure 2 presents the one-year trailing tracking difference of ETFs previously considered (plotted period corresponds to 1st January 2012 to 31th July 2013, i.e., one year after the first date of the returns sample). We can observe that curve aspects clearly differ from a fund to another and even more from a benchmark to another. Again, the MSCI Emerging Markets Index (Panel 2) seems to be the more difficult to replicate with huge positive and negative spikes especially on ETFs A and C. This is not the case for ETFs B and D, which manage to present trailing difference quite stable over time despite their levels always being significantly negative. Panels 1 and 2 also show off very clear disparities among trackers of the MSCI World Index and the MSCI Europe Index.

Statistics on the trailing tracking difference sample can also be derived for further analysis, typically mean, standard deviation, and extreme quantile. In Table 4, we present the lowest (i.e., worst) one-year tracking difference of funds considered. In other words, we report the minimum value reached by each curve of Figure 2. The overall ranking of ETFs according to this value is again different from that obtained using the tracking difference computed on the whole period. The interpretation is that, on several ETFs, sub-periods of under-performance are compensated by outperformance sub-periods and vice-versa. The example of ETF C tracking the MSCI World Index is striking. It is at the top of its ranking according to the tracking difference (cf. Panel 1 of Table 3) with a TD equal to -0.19%, but as we can observe on Panel 1 of Figure 2, it alternates one-year periods of outperformance and under-performance. This leads to the worst one-year cumulative return difference of -0.51%: more than 30 basis points below the TD computed on the whole period. Such oscillations can also be considered as a drawback even for investors seeking to keep ETF shares in their portfolio over a sufficiently long period: inevitably, they also face the risk of selling their shares at some point of particularly large under-performance. Again, how much importance to give to such statistic should

be assessed by investors based on their specific needs. Nonetheless, we remark that for each index, the ETF exhibiting the worst tracking difference is also the one with the worst one-year cumulative return difference, showing the link between these two indicators.

Panel 1: ***MSCI World Daily Total Return Net Index - USD***		
ETF	**Worst 1-Y Return Difference**	**ETF Ranking**
ETF A	-0.61%	4
ETF B	-0.41%	1
ETF C	-0.51%	2
ETF D	-0.69%	5
ETF E	-0.56%	3
Panel 2: ***MSCI Emerging Markets Net Total Return Index - USD***		
ETF	**Worst 1-Y Return Difference**	**ETF Ranking**
ETF A	-2.35%	5
ETF B	-1.25%	2
ETF C	-2.30%	4
ETF D	-1.17%	1
ETF E	-1.56%	3
Panel 3: ***MSCI Daily Net Total Return Europe Euro Index - EUR***		
ETF	**Worst 1-Y Return Difference**	**ETF Ranking**
ETF A	-0.20%	2
ETF B	-0.35%	4
ETF C	-0.20%	2
ETF D	-0.07%	1
ETF E	-0.22%	3

Table 4: Lowest (i.e., worst) one-year rolling return difference and ranking of ETFs considered. For each ETF, Worst 1-Y Return Difference corresponds to the minimum value of the rolling tracking differences computed daily from 2012-01-01 to 2014-07-31 using the last year of returns.

Figure 2: One-year rolling tracking difference of ETFs considered. Rolling tracking differences are computed daily from 2012-01-01 to 2014-07-31 using the last year of returns.

2.4 Trading and liquidity issues

If the replication quality of ETFs is a complex question that has major impact on the overall performance, trading the instrument is itself a serious matter. Like index funds, some investors prefer trading at Net Asset Value (requiring to source shares of ETFs outside of the regulated market) and thus avoid the potential bid-ask spread unavoidable on the market place. However, this is not always feasible for all investors, mainly for constraints on minimum volume. Hence, evaluating the liquidity via bid-ask spread risk and execution depth is often necessary.

Deville et al. (2013) find that ETF liquidity depends on its underlying basket's liquidity and volatility but also on ETF trading volume. Moreover, Hassine and Roncalli (2013) propose a spread measure for an investor wishing to trade a given notional, potentially non-executable at the best first limit order price. This measure is a first step toward the assessment of the ETF market depth and liquidity. However the calculation require intraday data. The latter allow greater accuracy but are sometimes not available with a very long track record and difficult to collect on certain specific markets. Roncalli and Zheng (2014) investigate European ETFs liquidity, both on a daily and intraday basis, and find that liquidity varies significantly between providers. Hence, the problem of assessing ETF liquidity and execution depth is of capital interest and still remains. This is also due to the off-exchange execution of large orders, especially in Europe where additionally, the market structure is fragmented between numerous exchanges. Thus, institutional investors should keep a close look to the market capitalization and average daily volume of their ETF candidates in order to evaluate if they are fitted for their typical size-orders and if the latter can be absorbed by the market without resulting in adverse market impact. The European market is in this sense very different from its US counterpart, where secondary trading has shown to be a relevant and robust source of liquidity in the shares of ETFs.

Conclusion

Choosing an ETF over another is complex and challenging. Once a benchmark is selected, assessing the tracking quality of ETF candidates is a specific task for each investor given that each fund has strengths and weaknesses. Indeed, we considered three indices and ranked five ETFs tracking each of them according to different indicators: none of them gave the same overall ranking. Furthermore, we find that expense ratio could not be used as a fair indicator of the quality of the investment vehicles. We did not manage to find any clear correlation between the level of fees put forward by the different ETFs and the overall quality of the products. Thus, investors would be wise not to over-focus on that element, as the tracking difference is a better representation of the total cost of ownership. We notice widespread differences among ETF, and significant variation of the yearly tracking difference over the period considered. Besides, the replication methodology, i.e., physical or synthetic, should not be considered as a crucial selection criterion: the reality is that any replication model has intrinsic risks that result in tracking error, tracking difference, or counterparty risk. It is the role of the ETF manager to design optimal implementation schemes and organize the elements of precaution and protection to mitigate those risks. Investors should thus only concentrate on statistics relevant for their investment purpose based on the available data.

References

Amenc, N., F. Goltz, and A. Lodh (2012). Choose your betas: Benchmarking alternative equity index strategies. *The Journal of Portfolio Management 39*(1), 88–111.

Amenc, N., F. Goltz, A. Lodh, and L. Martellini (2012). Diversifying the diversifiers and tracking the tracking error: Outperforming cap-weighted indices with limited risk of underperformance. *The Journal of Portfolio Management 38*(3), 72–88.

Basak, S., A. Shapiro, and L. Tepla (2006). Risk management with benchmarking. *Management Science 52*(4), 542–557.

Cochran, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 30, pp. 178–191. Cambridge Univ Press.

Deville, L., A. Calamia, F. Riva, et al. (2013). Liquidity in European Equity ETFs: What Really Matters? *Bankers Markets & Investors* (124), 60–73.

Elton, E. J., M. J. Gruber, and J. A. Busse (2004). Are investors rational? Choices among index funds. *The Journal of Finance 59*(1), 261–288.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business*, 34–105.

Frino, A. and D. R. Gallagher (2001). Tracking S&P 500 index funds. *The Journal of Portfolio Management 28*(1), 44–55.

Fuhr, D. (2001). Exchange-traded funds: A primer. *Journal of Asset Management 2*(3), 260–273.

Gallagher, D. R. and R. Segara (2005). The performance and trading characteristics of exchange-traded funds. *Journal of Investment Strategy 1*(1), 47–58.

Gastineau, G. L. (2001). Exchange-traded funds: An introduction. *The Journal of Portfolio Management 27*(3), 88–96.

Gastineau, G. L. (2002). Equity index funds have lost their way. *The Journal of Portfolio Management 28*(2), 55–64.

Gastineau, G. L. (2004). The benchmark index ETF performance problem. *The Journal of Portfolio Management 30*(2), 96–103.

Hassine, M. and T. Roncalli (2013). Measuring performance of exchange-traded funds. *The Journal of Index Investing 4*(3), 57–85.

Hsu, J. C., T.-m. Chow, V. Kalesnik, and B. Little (2011). A survey of alternative equity index strategies. *Financial Analysts Journal 67*(5).

Kostovetsky, L. (2003). Index mutual funds and exchange-traded funds. *The Journal of Portfolio Management 29*(4), 80–92.

Larsen, G. A. and B. G. Resnick (1998). Empirical insights on indexing: How capitalization, stratification and weighting can affect tracking error. *Journal of Portfolio Management 25*(1), 51.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance 7*(1), 77–91.

Pope, P. F. and P. K. Yadav (1994). Discovering errors in tracking error. *The Journal of Portfolio Management 20*(2), 27–32.

Roll, R. (1992). A mean/variance analysis of tracking error. *The Journal of Portfolio Management 18*(4), 13–22.

Roncalli, T. and B. Zheng (2014). Measuring the Liquidity of ETFs: An Application to the European Market. *Available at SSRN 2404313*.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance 19*(3), 425–442.

Svetina, M. and S. Wahal (2010). Exchange traded funds: Performance and competition. *Journal of Applied Finance 20*(2), 130.