Zurich University
of Applied Sciences

School of
Management and Law

Tracking Risk of Exchange Traded Funds Revisited
A Multivariate Regression Approach

SML Working Paper
No. 4

Thomas Merz

SML Working Paper Series ISSN 2296-5025

Januar 2015

ZHAW School of Management and Law, St.-Georgen-Platz 2, 8401 Winterthur, Switzerland

The Tracking Risk of Exchange-Traded Funds Revisited: A Multivariate Regression Approach

Thomas Merz[a,b*]

[a]*UBS Global Asset Management, Stockerstrasse 64, 8098 Zurich, Switzerland*
[b]*Part-time lecturer of core finance indexing at ZHAW Zurich University of Applied Sciences, School of Management and Law, Department of Banking, Finance, Insurance, Technoparkstrasse 2, 8401 Winterthur, Switzerland*

Abstract

This empirical study investigates the ability of exchange-traded funds (ETFs) to replicate the risk-return characteristics of their respective benchmarks accurately. By decomposing ex-post tracking performance, this study finds that the commonly used measure, tracking error, rarely sufficiently explains the deviation from the benchmark and hence has very limited predicting power for assessing the tracking quality of an ETF. The results presented here clearly indicate that in many cases a linear metric is a more reliable predictor for future return deviations and that, therefore, the total cost of administrating an ETF provides a fairly good estimate of its tracking quality.

JEL classification: G11; G23

Keywords: exchange-traded funds; physical ETFs; synthetic ETFs; tracking risk, tracking error; tracking difference

The author is grateful to Hans-Jörg Morath, MSc in Finance student at University of Zurich, for his valuable contributions to this paper. The author also acknowledges the support of Professor Thorsten Hens, Department of Banking & Finance at University of Zurich for his valued input. The views presented in this paper are those of the author and do not necessarily represent those of the afore-mentioned institution.
* Contact details for the author: phone:+41 44 234 30 78, e-mail: thomas.merz@ubs.com

1. Introduction

This study investigates the tracking risk of physical and synthetic European ETFs. The objective of an ETF is to track the returns of its benchmark index as closely as possible. Physical ETFs are designed to replicate the benchmark index by holding the underlying securities, while synthetic ETFs use derivatives, primarily OTC derivatives such as swaps for replication purposes. To overcome some of the challenges of physical replication, the synthetic replication method was introduced in 2001. Synthetic ETFs, using the unfunded swap model, hold an asset basket that usually does not match the index's underlying securities. In order to replicate the index return, the basket return is swapped against the benchmark index return. However, synthetic ETFs (or physical ones, for that matter) cannot guarantee that the performance perfectly matches the returns of the benchmark index. Synthetic ETF providers claim that the synthetic replication method is more efficient and produces a lower tracking risk (i.e., tracking error) when compared with physical ETFs. According to the above-mentioned and other, often contradictory, arguments, there is a fair amount of uncertainty amongst practitioners as to how and on what basis the tracking quality of ETFs should best be evaluated. Also, in academic literature, there is still very little consensus on this particular topic; this is mainly due to different study designs and hence contradicting reported results.

As the main factors that give rise to tracking errors of ETFs, previous empirical studies identified transaction costs, index-composition changes, corporate actions, fund cash flows, index volatility, the reinvestment of dividends, and applied index replication strategies[1]. Furthermore, Shin and Soydemir (2010) argue that the volatility of the exchange rate is also a source of tracking error. This study finds that both physical and synthetic European ETFs are affected by an occasional, albeit sometimes significant, tracking risk. This analysis provides evidence that ETFs that follow a synthetic replication strategy, rather than holding the underlying benchmark securities, are less prone to tracking error; however, in most cases they underperform both with regard to the benchmark return and the physically replicated counterparts.

2. Literature Review

Previous academic research by Elton et al. (2005) documents significant tracking errors generated by US domiciled index funds and ETFs. Frino and Gallagher (2001, 2002) document similar results for Australian ETFs. Roll (1992), Pope and Yadav (1994), and Larsen and Resnick (1998) identify the metrics to measure tracking error as the dispersion of the fund's Net Asset Value (NAV) return relative to the benchmark return. Harper, Madura, and Schnusenberg (2006), in their sample, use market prices (paid prices on exchange) rather than the NAV to evaluate the tracking risk of ETFs. They find that market price tracking error can substantially deviate from NAV tracking error. Elton et al. (2005) and Engle and Sarkar (2006) on the other hand state that market-price deviations from NAVs should theoretically

[1] For more details see Chiang 1998, Elton et al. 2005 and also Frino and Gallagher 2002

disappear quickly due to the arbitrage process[2]. DeFusco, Ivanov, and Karels (2011) show that the pricing deviations of the three most widely known US-domiciled ETFs (tracking the DJIA, NASDAQ, and S&P500) are different from zero. The authors claim that the pricing deviation can be approximated by the cost of administering an ETF. Furthermore, seasonal patterns in tracking errors have been detected by a number of studies such as Frino and Gallagher (2001, 2002), Frino et al. (2004), and also Rompotis (2010a, 2010b). Chu (2011) investigates the tracking errors of ETFs traded in Hong Kong and finds that they are higher compared with those in the US and Australia. This study concludes that one potential explanation is the use of synthetic investment methods over the holding of underlying index constituents.

Other recent research by Blitz, Huij, and Swinkels (2010) examines the tracking error of European index funds and ETFs as measured by their underperformance against the gross total return indices. This study finds that European ETFs underperform with regard to their benchmarks and that withholding taxes on dividends and fund expenses explain most of the observed deviations. They argue that ETF performance could differ from the performance of the underlying indices due to an imperfect replication strategy and market frictions (e.g., transaction costs, management fees, or trading noise). Rompotis (2008) and Milionas and Rompotis (2006) examine the performance of German and Swiss equity ETFs, respectively, using paid-on-exchange prices rather than NAV prices. Both studies report underperformance of ETFs relative to their benchmark indices. They find (not unexpectedly) that tracking error is positively related to the volatility of the respective benchmark indexes. This is, of course, mostly due to the study design, which is based on paid-on-exchange prices rather than NAV prices. Drenovak and Urosevic (2010) also document underperformance of European Bond ETFs between 2007 and 2009 of well above their total expense ratios (TERs). Houweling (2011) analyses US and European domiciled fixed income ETFs and reports that the money market ETFs in his sample, on average, underperform by approximately the average TER. For the bond ETFs in the sample, the study documents an underperformance of well above the average TER. The author argues that higher volatility of corporate bond index constituents increases the transaction costs in the ETF portfolio, resulting in greater underperformance.

Merz (2011) approaches the ETF tracking question from a different angle. The author documents a significant tracking risk for ETFs replicating the MSCI Emerging Markets index and concludes that tracking error has very little forecasting power compared to the observed cumulative tracking difference out of sample. A similar result is reported by Brunner (2011), who applies an in-sample and an out-of-sample test on tracking error. Unfortunately, both studies are based on rather limited data samples of European ETFs[3]. In a similar fashion, the two authors argue that using quadratic TE measures to evaluate ETFs with regard to their tracking quality leads to a substantial selection error owed to a systematic underestimation of the out-of-sample tracking risk of ETFs. According to their results, this systematic error is mainly due to a high serial correlation of the daily return differentials, which may be explained by an accrued cost component for administering the ETF. Further, Blitz, Huij, and

[2] The arbitrage process is based on the optionality of market makers and/or authorized participants to either subscribe or redeem in kind or in cash for ETF units in the case that mispriced ETF quotes are shown in the market

[3] Merz (2011) analyzed only the four biggest ETFs on the MSCI Emerging Market index measured by AuM (VWO, EEM, CSEM, XMEM), while Brunner (2011) also includes only 24 European equity ETFs in his study.

Swinkels (2010) find that the linear cost factors, such as withholding taxes and fund expenses, also result in similar explanations for the observed tracking difference.

To date, when evaluating the tracking risks of passive investment strategies, there has been limited systematic academic research on the appropriateness of quadratic tracking error measures. This study aims to close this gap by analyzing a broad and large sample of European ETFs – in contrast to Merz (2011) and Brunner (2011) – tracking the most important indices for each asset class. It is an attempt to provide some guidance on the most appropriate measure for estimating out-of-sample tracking risks vs. a given benchmark return. This study, therefore, tests the following two hypotheses:

H1: The most popular tracking error (TE) measure (quadratic TE) currently in use has little explanatory power for future ETF return deviation vs. its benchmark return

H2: Cost variables are able to explain the largest portion of out-of-sample ETF return deviation vs. benchmark return

Section 3 describes the ETF sample analyzed and the methodology applied for testing the two hypotheses. Using the quadratic tracking error measure, Section 4 addresses the hypothesis testing and provides evidence that future tracking deviations vs. benchmark are heavily underestimated. For the pooled panel regression, the study reports that cost proxy variables, such as TER and autocorrelation in return differentials, carry much more information about future tracking risks of ETFs than any other analyzed variable. Finally, Section 5 provides a summary and a conclusion.

3. Data and Methodology Used

This section provides descriptive sample statistics and introduces various tracking error measures.

3.1. Data sample Statistics

This study includes a broad sample of 131 European ETFs, covering ETFs replicating a total of 26 indices of the two main asset classes: equity and fixed income (bonds and money market). In order to compare multiple ETFs on the same index, the longest mutual data history has been chosen as the starting date. As documented in Table 1, the longest history of data was observed using a total of 2,468 fixed income ETF return observations (max. of in-sample days plus a max. of out-of-sample days). The smallest ETF data set contains 667 daily returns.

For each ETF, the following information was gathered: product name, Bloomberg ticker, benchmark index, asset class, total expense ratio (TER), dividend policy, dividend payments, dividend payment dates, legal structure, domicile, daily NAV level, daily index level, replications strategy applied, and securities lending status. Based on this information, daily, weekly, and monthly time series of ETF and index returns, TE measures, tracking differentials, autocorrelations, and index volatility figures were calculated. In order to analyze the broadest ETF sample possible, the study design includes all ETFs listed on the SIX Swiss Exchange and on Xetra (Deutsche Börse, ETF segment) on a given index which have more than three years of data history available. To minimize data errors, the majority of data used in this paper were collected from its most original source, i.e. for all product-related data such as NAV levels, dividend's face values and payment dates and other data related to the product structure were requested directly from the fund management company for each of the analyzed ETFs.

Table 1 reports the summary statistics for the European ETF sample as at 31 March 2013. The sample is divided into criteria, including full sample, asset class (index universe), and replication style. There are 131 ETFs replicating in total 26 different benchmarks and representing over EUR 100 billion of invested assets. Physical ETFs account for 50.4% of the total number of ETFs but account for 68.6% of the total AuM, where the average physical ETF has almost double the assets under management compared with the average synthetic ETF. With few exceptions, all types of indices analyzed include both physical and synthetic ETFs. The proportion of physical ETFs is higher only for developed country indices, where, for example, due to the specific nature of the index, the money market ETFs are all synthetically replicated. The average TER (median TER) for equity ETFs accounts for 37 basis points (median at 34bps), where the average TER for fixed income ETFs is only a little more than half, at 20 basis points (20bps). The difference in TER for physical and synthetic ETFs in the European sample at 35 vs. 33 basis points (median at 33 vs. 30 bps) is considerably smaller. For equity ETFs, the lowest TER is 9 basis points and the highest 75 basis points. In contrast, the differential for synthetic ETFs is much higher: 0 for the lowest and over 100 basis points of the highest TER in the sample.

For the observed tracking difference at the end of the out-of-sample period, equity ETFs seem to produce higher return deviations from their benchmark returns than fixed income ETFs on average. Similarly, on average synthetic ETFs appear to be able to keep the cumulative return distance between the ETF and the index portfolio over the out-of-sample period slightly smaller than physical ETFs (25.0 vs. 27.7 bps). At first glance, this result appears to contradict the findings of Shin and Soydemir (2010). However, it is most probably owed to the fact that in Table 1 the European data sample is split between physical and synthetic ETFs regardless of the asset class each ETF is replicating. By dividing the European ETF sample differently, i.e., not only by replication style but also by asset class, the differences between physical and synthetic ETFs are reduced even further. Annually, the average physical fixed income ETF shows a tracking difference of 17.6 basis points, with its synthetic counterpart displaying a tracking difference of 15.2 basis points. Similarly, the difference between annualized physical and synthetic tracking differences for equity ETFs is 30.7 vs. 25.7 basis points. Looking at the minimum as well as the maximum observed tracking differences, one notices that in both cases the inferior tracking difference lies with a synthetic ETF. This clearly indicates a need to be cautious with general statements referring to tracking difference and the applied replication strategy. The results from the European sample statistics illustrate that to appropriately estimate the tracking quality out-of-sample, investors need guidance because in-sample (ex-post) deviations can be quite substantial, even for ETFs.

3.2. Tracking Error Performance Measures

In practice, the majority of professional as well as private investors base their decisions on the widely-used quadratic tracking error ($TE_{RMS(NCE)}$ and $TE_{SD(CE)}$) metric to assess the tracking performance of an ETF (see Goltz 2009).[4]

In this paper, the definitions of tracking error (TE) are limited to the ones most widely accepted by practitioners. Five different tracking error measures are introduced to investigate how well each individual metric is able to capture the tracking quality of a European ETF. A number of different TE definitions can be found in academic literature. Pope and Yadav (1994) or Rudolf, Wolter, and Zimmermann (1999), for example, define TE as the variance of the return differential between portfolio return and benchmark index return. Roll (1992), Clarke, Krase, and Statman (1994) use a different TE metric. In their studies, the TE is defined as simple return differentials of a given interval (e.g. monthly) for portfolio returns vs. benchmark returns. Next to the simple return differential, and the more sophisticated quadratic measures, linear TE measures are also frequently used (see Roll 1992 and Fulmek 2003).

Metrics definition:

The first TE measure introduced for the purpose of this study is the most commonly used non-centric, quadratic measure, the root mean square of the difference between ETF and benchmark index returns. Formally, the non-centric tracking error $TE_{RMS(NCE)}$ is given by

$$\widehat{TE}_{RMS(NCE)_T} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_{p,t} - r_{b,t}\right)^2}\ . \tag{1}$$

Here r_p is the log return of the ETF and r_b is the log return on the benchmark index on day *t*.

A similar quadratic TE measure, and the second TE measure of this study, was initially introduced by Roll (1992) and is calculated as the standard deviation of the return differentials between ETF and benchmark index. Formally, the centric tracking error $TE_{SD(CE)}$ is given by

$$\widehat{TE}_{SD(CE)_T} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_{d,t} - \overline{r_d}\right)^2}\ , \tag{2}$$

with $r_{d,t} = r_{p,t} - r_{b,t}$ and $\overline{r_d} = \frac{1}{T}\sum_{t=1}^{T} r_{d,t}$.

The third quadratic metric of TE introduced in this study is the standard error of the residual of the return regression, where the ETF return r_p is the dependent variable and the benchmark index return r_b the independent variable. Formally, the regression is given by

$$r_{p,t} = \alpha + \beta r_{b,t} + \varepsilon_t\ , \tag{3a}$$

[4] According to the results of the 2009 EDHEC European ETF Survey, 73% of investors use the TE as a measure of tracking performance, while 44% use the correlation-based measure and only 6% use a metric based on co-integration. Other measures include simple comparison of mean return (23%) or asymmetric tracking error (9%).

where r_p is the ETF return, r_b is the return on the benchmark index, and ε_t is the error term of the regression. The TE_{reg} is then the standard error of the above regression (see Pope and Yadav 1994, Cresson et al. 2002 and Rompotis 2010a), where α is not expected to be statistically different from zero, the *i*-th β is not to be expected to be statistically different from unity and the *i*-th R^2 is expected to be close to one. For ETFs which depart from full replication strategies, one expects higher regression standard errors, α-s different from zero, β-s lower than unity and significantly lower R^2s.

The TE_{REG} is therefore formally given by

$$\widehat{TE}_{REG_T} = \sigma_{r_{p,t}}\sqrt{1-\rho_{bp}{}^2}\,, \tag{3b}$$

with $$1-\rho_{bp}{}^2 = \sqrt{\frac{\sum_{t=1}^{T}\left(r_{p,t}-\hat{r}_{p,t}\right)^2}{\sum_{t=1}^{T}\left(r_{p,t}-\bar{r}_{p}\right)^2}}\,.$$

Some literature[5] indicates that quadratic TE measures are unable to detect deviations from the benchmark when *pro rata temporis* differences show a high degree of serial correlation and one therefore expects that the in-sample TE underestimates the out-of-sample deviation. In a passive investment framework, such as with ETFs, one can expect that the daily differences in returns on ETF vs. benchmark are mostly of this nature, depending on the replication method used and the index replicated. More promising in detecting highly autocorrelated return differentials are linear models such as symmetrical or asymmetrical TE measures.

One of the linear models of TE used in this study is calculated as mean absolute deviations (MAD) of the return differential between the ETF and the benchmark index, referred to as TE_{MAD}. Formally, the TE_{MAD} is defined as

$$\widehat{TE}_{MAD_T} = \frac{1}{T}\sum_{t=1}^{T}\left|r_{p,t}-r_{b,t}\right|\,. \tag{4}$$

A second linear measure of TE used in this study which is similar to the one in Equation 4 above is the mean absolute downside deviation (MADD) return differential, referred to in this study to as TE_{MADD}, formally given by

$$\widehat{TE}_{MADD_T} = \frac{1}{T^*}\sum_{t=1}^{T}\left(\left|r_{b,t}-r_{p,t}\right|\cdot 1_{r_{b,t}>r_{p,t}}\right), \tag{5}$$

where $T^* = \sum_{t=1}^{T} 1_{(r_{b,t}>r_{p,t})}$.

Confidence intervals:

To test the different TE measures for their out-of-sample accuracy, the 99% confidence intervals of the expected tracking differences were calculated on a continued basis. To ensure consistent estimation of the confidence intervals, the root mean square TE measure is omitted

[5] See, for example, Roll (1992), Larsen, and Resnick (1998), Baierl and Chen (2000), Frino and Gallagher (2002), Satchell and Hwang (2001), Fulmek (2003) and Harper, and Madura and Schnusenberg (2006).

as it does not correct for the mean tracking difference (cf. Equation 1). Otherwise, the confidence intervals on the non-centric tracking error ($TE_{RMS(NCE)}$) would in our view experience potential bias problems as the distribution properties are not properly known. We therefore calculate the respective confidence intervals only for the centric and regression-based TE measure. The confidence intervals using the centric TE measure are calculated as follows:

$$CI_{99}TE_{SD(CE)_{\mathrm{T+k}}} = \delta_k \pm 2{,}58\sqrt{k} \cdot \widehat{TE}_{SD(CE)_{\mathrm{T}}} \,, \tag{6}$$

and for the quadratic TE measure

$$CI_{99}TE_{REG_{\mathrm{T+k}}} = \delta_k \pm 2{,}58\sqrt{k} \cdot \widehat{TE}_{REG_{\mathrm{T}}} \,, \tag{7}$$

where δ_k is the accumulated daily total expense ratio (TER) around which the confidence interval is centered, with $\delta_k = \sum_{k=1}^{K} dTER_k$ and k as the number of days in the out-of-sample period k equaling 1,…,K

For the linear models, the in-sample TE_{MAD} and TE_{MADD} are multiplied by the number of out-of-sample observations K. For the symmetric measure, the deviation accuracy has been formally tested using the following estimates:

$$TE_{MAD_{\mathrm{T+K}}} = \widehat{TE}_{MAD_{\mathrm{T}}} \cdot K \,, \tag{8}$$

and for the asymmetric model

$$TE_{MADD_{\mathrm{T+K}}} = \widehat{TE}_{MADD_{\mathrm{T}}} \cdot K \,. \tag{9}$$

Multivariate regression:

In a further step, we use the regression formulated below, where the observable tracking difference is attributed to a set of twelve drivers (factors) which are typically discussed by practitioners as having the most influence over the tracking quality of an ETF. This approach aims to address the second hypothesis. We examine all ETF and index return data from the European sample with the following multivariate pooled OLS regression model against the ETF panel data:

$$\begin{aligned} TDIFF_{\mathrm{i},t} = {} & \alpha + \beta_1 \mathrm{TE}_{\mathrm{vol}_{\mathrm{i},t}} + \beta_2 \mathrm{PHY}_{\mathrm{i},t} + \beta_3 \mathrm{OPT}_{\mathrm{i},t} + \beta_4 \mathrm{SEC}_{\mathrm{i},t} + \beta_5 \mathrm{TER}_{\mathrm{i},t} + \beta_6 \mathrm{CH}_{\mathrm{i},t} + \beta_7 \mathrm{DE}_{\mathrm{i},t} + \\ & \beta_8 \mathrm{FR}_{\mathrm{i},t} + \beta_9 \mathrm{IE}_{\mathrm{i},t} + \beta_{10} \mathrm{LU}_{\mathrm{i},t} + \beta_{11} \mathrm{PEA}_{\mathrm{lag1}_{\mathrm{i},t}} + \beta_{12} \mathrm{IND}_{\mathrm{vol}_{\mathrm{i},t}} + \varepsilon_{\mathrm{i},t} \,, \end{aligned} \tag{10}$$

Where *i* is the index of choice[6], $\mathrm{TDIFF_i}$ are the observed (out-of-sample) tracking differences between the ETF and the benchmark index. $\mathrm{TEvol_i}$ is the daily quadratic tracking error measure according to equation (1). This variable represents the tracking quality proxy which is mostly used by practitioners to evaluate the tracking quality of an ETF. $\mathrm{PHY_i}$ is a dummy variable that takes on the value one for ETFs using physical replication and zeros otherwise; $\mathrm{OPT_i}$ is also a dummy variable that takes on the value one for ETFs using an optimized model

[6] All analyzed indices are listed in the captions of Table 3

portfolio to replicate the benchmark index and zeros otherwise; SEC is a dummy variable that takes on the value one for ETFs deploying securities lending activities and zeros otherwise. The variable TER represents the portion of costs accrued against the ETF NAV on a *pro-rata-temporis* basis according to the regulatory guidelines of UCITS for the non-CH domiciled ETFs and, according to FINMA KAG, for the CH domiciled ETFs in the sample[7]. Fund domicile Switzerland (CH_i), fund domicile Germany (DE_i), fund domicile France (FR_i), fund domicile Ireland (IE_i), and fund domicile Luxembourg (LU_i) are all dummy variables that take on the value one for ETFs which are domiciled in these countries, and otherwise take on zero. Autocorrelation of the return differences ETF vs. benchmark index are represented by the variable $PEAlag1_i$. The index volatility is represented by the variable $INDvol_i$ and is calculated as 30 days moving average volatility of the respective benchmark index returns. The last two variables in the regression model represent cost proxy variables which are expected to be more sensible for linear deviations between ETF and benchmark returns. Finally, ε_i represents the unexplained residual term of the regression model.

The panel data have been tested for heteroskedasticity, autocorrelation, and cross-sectional dependencies. To test for heteroskedasticity, the results from the White's test as well as the Breusch-Pagan/Cook-Weisberg test clearly reject H0, hence residuals do not show constant variance. The sample data were also tested for multicollinearity. All calculated variance inflation factors (VIFs) show values clearly below 10. In fact, the VIFs calculated a range from 1.01 to 4.12; hence, multicollinearity does not seem to be severe (values below 5 indicate only weak multicollinearity). Additionally, all variables have been tested for non-linearity whereby H0 was clearly rejected, and thus the linear regression model is applicable. As this study analyzed time series data, autocorrelation was tested applying the Wooldridge test. Based on the results, H0 for no first-order autocorrelation can be rejected. Furthermore, three different testing procedures were executed (Pesaran, Friedman, and Frees) to test for cross-sectional dependencies. Whereas the results from Friedman's and Frees' tests clearly reject H0, the results from Pesaran's test is less conclusive. However, the average absolute value of the off-diagonal elements (correlation) of 0.631 indicates a cross-sectional dependency of the residuals. These test results indicate the presence of heteroskedasticity, autocorrelation, and spatial correlation (cross-sectional dependencies) which all violate the assumptions of standard statistical inference. To ensure consistent estimation of standard errors, we resorted to the method proposed by Driscoll and Kraay (1998), which ensures robustness to heteroskedasticity, autocorrelation, and cross-sectional dependencies for the pooled panel data regression.

4. Empirical Results

The evidence presented in this paper indicates that the widely-used quadratic TE measures are subject to substantial estimation biases towards the out-of-sample tracking difference. The results further demonstrate that the linear models do not properly estimate future tracking

[7] Since the accounting standards for ETFs regulated by UCITS and FINMA KAG are not identical, there are some differences in costs which are accrued and reported as TER for the different fund domiciles. Additionally, and even more important is the fact that all transaction costs incurred by the fund as a result of replicating an index are not reported as costs under TER. This holds true for both regulatory frameworks, UCITS and FINMA KAG. This is of even greater importance in the case of synthetic ETFs as the swap costs do not qualify as costs which have to be reported as TER and are therefore not captured by the TER variable. Nevertheless, they do influence the tracking difference of an ETF vs. its benchmark return.

deviations, and while the linear models overestimate future tracking deviations, the quadratic TE measures, to almost equal parts, both under- as well as overestimate return differences.

The tracking error statistical data in Table 1 provides evidence that there is very little difference between the different quadratic TE measures. In addition, differences between the two linear models (TE_{MAD} and TE_{MADD}) are very small, and thus the decision as to which TE model to apply is expected to have an only insignificant impact on the overall results (if any). Therefore, all TE calculations used for the multivariate regression analysis are based on the non-centric quadratic tracking error formula according to Equation 1.

For the determinants of the observed tracking differences, the study finds that the cost proxy variables, such as TER and autocorrelation of return differences, show the highest explanatory power for the fitted regression line. These results are in line with results from previous studies as reported in Blitz, Huij, and Swinkels (2010), Merz (2011), and Brunner (2011).

4.1. Tracking Error Out-of-Sample Tests

To test whether tracking error is a suitable estimator for future performance deviation against the relevant benchmark return (hypothesis one), the in-sample calculated tracking error is forward-modeled by using Equation 6 and then comparing results to the observed tracking performance difference at the end of the out-of-sample period. To control the results for different calculation variations of the TE, the test was repeated using an additional quadratic TE measure based on Equation 7 as well as the two linear TE measures according to Equations 8 and 9. The results of the additional quadratic TE measure as well as between the two linear TE measures differ insignificantly, as is presented in Table 2.

Upon examination of the estimated tracking difference using the forward-modeled TE, one notices that the quadratic TE measures mislead the estimated tracking difference in more than 45% of all cases to understate the tracking risk. In other words, in more than 45% of cases the modeled TE underestimates the tracking risk of the ETF. As Panel A reports, in over 54% of the cases the in-sample quadratic TE overestimates the level of performance differences; as a result, the ETF shows a return difference vs. the benchmark return which is smaller than investors would have expected using the TE as an estimator for future return deviations over the out-of-sample time period. Only for equity ETFs does this general statement not hold true whereby the ETF sample data suggests that in more than 53% of cases quadratic TE do indeed underestimate the tracking risk using the TE measure.

Panel B of Table 2 examines the severity of the level of misestimating when quadratic TE measures are taken into account. The data reported by Panel B reveals that over 40% of estimating errors lie outside of a 10% threshold; hence, the bias in the estimation is significantly above or below 10% of the total performance difference. The results suggest that using a quadratic TE measure to estimate future tracking difference is not very reliable, especially in the case of equity ETFs. For bond as well as for money market ETFs, all of the estimates lie inside the 10% threshold; for equity ETFs, only 52.7% of all estimations are within this threshold range.

As for the results documented for the quadratic TE, the outcomes were tested for robustness by applying linear TE measures. Both variations show similar results for the over- and underestimation as reported with the quadratic TE measure. Nevertheless, the linear TE models indicate a heavy bias towards overestimation even for equity ETFs. For the 10% threshold analysis, the results point in opposite directions when quadratic and linear models are compared, especially for fixed income ETFs. While the quadratic TE results show 100% within the threshold for bonds and money market ETFs, for the linear measures all estimations were outside the applied threshold.

In line with the formulated hypotheses, tracking error out-of-sample tests of the European ETF sample reported in Table 2 confirm previous results documented in Merz (2011) and Brunner (2011) whereby the TE measures, the quadratic as well as the linear model, do not estimate the out-of-sample tracking difference closely enough.

4.2. Determinants of ETF Tracking Difference

In this step, the entire European ETF sample return differences at the end of the out-of-sample period were analyzed using a pool panel data regression according to Equation 10. By regressing twelve independent variables as described in Subchapter 3.2 against the observed tracking difference, TRACKDIFFACC[8], the respective load of each individual beta factor is calculated and reported in Table 3. A significance test is also deployed using the three most common confidence interval levels 99%, 95%, and 90%. P-values for each beta factor are reported for each individual factor whereby ***, **, and * indicate the respective significance level.

The analyzed data suggests that the regression model from Subchapter 3.2, using the twelve independent variables, can appropriately explain the evident out-of-sample tracking performance at the end of the observation period. This holds true for all the tested benchmark exposures with only five exceptions[9]; in four of these cases, the fitting quality is considerably lower than that of all the other tested exposures, with R^2 showing values of well above 0.8 ranging from 0.83 up to 0.99. In all four cases, the lower fitting quality of the regression model is probably due to the fact that the net return of the benchmark can easily be outperformed by some of the tested ETFs which use an optimal domicile and deploy a securities lending program which then leads to the outperformance vs. the respective benchmark net return. This is typically true for benchmarks for which the total return is calculated using the worst-case assumption with regard to reclaiming withholding taxes on dividends. Such benchmarks typically include either a large portion of US (SPTR500N, NDDUUS) or European equities (MSDEE15N, SX5T), where, for example, on the fund level the taxation on dividends of US ISINs can be reduced by 50% in the case of an IE domicile ETF. In terms of securities lending revenues, European ISINs have provided considerably higher returns than other constituents of standard indices. In this case, some of the variables are responsible for outperformance as opposed to underperformance as calibrated in the model; the fitting quality suffers and hence, R^2 ranges between 0.47 and 0.69 for those four benchmarks. As a consequence, the four benchmarks (SPTR500N, NDDUUS, MSDEE15N

[8] The cumulated tracking difference (TRACKDIFFACC) is calculated using the following formula: $\text{TRACKDIFFACC} = \sum_{t=1}^{K}(r_{p,t} - r_{b,t})$.

[9] SPTR500N, NDDUUS, MSDEE15N, SX5T and LEATRTREU.

and SX5T) are no longer part of the conclusive results. The fifth exception is the benchmark of European government bonds in Panel B (LEATRTREU) of Table 3, where the model shows an R^2 of well below 0.6; this benchmark exposure is excluded from the conclusion, although the reason for the poor quality of the fitting is of a totally different nature than for the equity benchmarks.

For both equity and fixed income ETFs, the results derived from the European ETF sample indicate that only one independent variable, and, to a lesser extent, four, show consistent significance of explanatory power (factor loads) to the observed out-of-sample tracking difference. Panel A for equity ETFs as well as Panel B for fixed income ETFs document significant (99%) p-values for variable TER ($\beta 5TER_i$) for all tested benchmark exposures. To a lesser extent, three other variables for equity ETFs ($\beta 12INDvol_i$, $\beta 2PHY_i$ and $\beta 11PEAlag1_i$) show significant explanatory power for most exposures tested, $\beta 12INDvol_{i;}$ although in two cases on lower confidence levels. All other variables do not demonstrate a consistent level of explanatory power for the European model. In contrast to the equity model, the tracking performance of fixed income ETFs seems to be driven mainly by the TER variable.

To assess the second hypothesis, the reported data of the European ETF sample is now calculated as the portion of the total out-of-sample performance difference explained by each factor (variable) as opposed to absolute values, which are discussed in Table 3. Hence, Table 4 summarizes the relative weights of each variable for explaining the observed tracking difference for all equity and fixed income ETFs in Panels A and B, respectively. Table 4 also offers insights into the difference between physical and synthetic ETFs when it comes to the relative importance of explanatory power of the different variables for both equity and fixed income ETFs in the European ETF sample.

The proportional view documented in Table 4 indicates that the TER is by far the most important variable for explaining the out-of-sample tracking difference. This holds true for both the equity and the fixed income ETFs of the European sample. Even though the fitting quality of the regression model is slightly lower for equity ETFs than for fixed income ETFs, R^2 is at 0.77 for equity and 0.80 for fixed income ETFs. In line with expectations, the relative importance of the TER variable is even higher for synthetic ETFs and accounts for almost 50% (48.77%) of the equity and over 50% (51.88%) of the fixed income ETFs. The second most important factor seems to be the autocorrelation variable. Even though the relative importance is only 11.84% and 13.67% for equity and fixed income ETFs, respectively, the autocorrelation gives some support to the discussion about the first hypothesis, where linear models (captured by TE_{MAD} and TE_{MADD}) seem to be more accurate in capturing the tracking risk of European ETFs[10]. In contrast to the selection criteria applied by most practitioners at fund selection desks, the variable tracking error ranks only at 5.47% for equity and 11.66% for fixed income ETFs.

These results clearly suggest that a fund selection process which aims to select ex-ante the ETF with the highest chance of a minimal tracking difference vs. its benchmark return, needs to focus heavily on the TER variable. This finding is in stark contrast to some selection

[10]See also results documented in Table 2, where quadratic TE measures to almost equal parts (45.8%) underestimate the tracking risk, and linear TE measures mainly overestimate them in the majority of cases.

heuristics advocated by ETF providers and which are documented in a number of different retail publications where ETF selection processes are proposed. In recent years, it appeared that mostly synthetic ETF providers tried to shift investors' eyes towards tracking error (rather than the TER) and towards the tracking difference to assess tracking performance and hence the tracking quality of an ETF. By doing so, investors may easily have overlooked that the chosen ETF with the lowest ex-post tracking error does not necessarily provide the highest tracking quality in terms of minimal tracking difference vs. benchmark return over time.

In order to control model robustness, time series of ETF returns and hence all OLS regression calculations have been repeated using weekly as well as monthly return data based on the same European ETF data sample. The results are also checked for robustness using the two other quadratic TE measures presented in Subchapter 3.1 ($TE_{SD(CE)}$ and TE_{REG}) for calculating the factor variable $\beta 1TEvol_i$. In terms of the significance of each factor load per tested benchmark exposure and the relative explanatory power for each individual variable, the results are similar when using either different return frequency or different quadratic TE measures. Therefore, the results presented in Tables 3 and 4 appear to be robust and thus do not assume a specific tracking error definition, nor do they make assumptions with regard to specific return data calculation frequency.

5. Summary and Concluding Remarks

This study explores the quality of TE measures for estimating out-of-sample tracking differences of European ETFs listed on the SIX Swiss Exchange and Xetra using three quadratic and two linear tracking error models and comparing their estimated tracking deviations with observed tracking differences at the end of the out-of-sample period. Firstly, the findings provide evidence that the most widely-used TE measures are unable to accurately estimate future ETF return deviations vs. the respective benchmark index return. The results reported in this study clearly indicate that TE measures should not be consulted when judging tracking quality for investment decisions among different ETFs. None of the tested TE measures were able to accurately estimate future observed deviations. While the quadratic TE measures equally produced under- as well as overestimations, the linear models in most cases overestimated the out-of-sample tracking difference. This applies to different asset classes such as equity, bonds and money market ETFs, irrespective of the replication strategies used.

Secondly, the study tests the validity of the widespread opinion amongst practitioners that the observed tracking error best explains the tracking difference. While tracking error for all equity ETFs in the sample accounts for only 5.47% of the total explanatory power, the explanatory power of the variable tracking error is significantly different for physical and synthetic equity ETFs (8.57% and 3.02%, respectively). Slightly higher values are documented for fixed income ETFs in the sample. Nevertheless, by far the biggest influence on the observed tracking difference between ETF and its benchmark index is the TER variable, with 44.28% for all equity ETFs, 39.09% for physical, and 48.77% for synthetic equity ETFs. Inversely, slightly lower values occur in the case of fixed income ETFs (36.84%, 32.82%, and 51.88%, respectively). Lower explanatory power, albeit on average still higher than the ones for TE variables, is documented for the autocorrelation variable, which represents an additional proxy variable for cost-driven return differences between ETF and benchmark index returns. This provides evidence that linear TE models are expected to produce slightly better out-of-sample tracking risk estimates when comparing synthetic ETFs from different ETF managers.

This study aimed to close a gap in academic literature and to provide guidance to investment practitioners on how to select ETFs when the tracking risk of competing ETFs is to be evaluated, and how specific ETFs should be selected based on future prediction of the lowest tracking risk. By providing an overview of individual forecasting qualities of the different tracking error measures and by quantifying the explanatory power of various variables, which all have an impact on the tracking risk of an ETF, this study concludes that the TER variable of a specific ETF has the biggest forecasting power with regard to the tracking risk, regardless of the replication method applied by individual ETFs. This analysis of a large and broad European ETF sample is able to clearly accept therefore both hypotheses stated in Chapter 2.

References

Baierl, G.T., Chen, P., 2000. Choosing Managers and Funds. Journal of Portfolio Management 26(2), pp. 47-53.

Blitz, D., Huij, J., Swinkels, L., 2012. The Performance of European Index Funds and Exchange Traded Funds. European Financial Management 18(4), 649-662.

Brunner, C., 2011. Tracking Error – Wie aussagekräftig ist diese Risikokennzahl in Bezug auf die Abbildungsgenauigkeit verschiedener ETFs? Unpublished Master's thesis, FHS St. Gallen.

Chiang, W.C., 1998. Optimizing Performance, in: Neubert, A.S. (Ed.), Indexing for maximum investment results. Global Professional Publishers, Chicago, Illinois, USA, pp.307-335.

Chu, P.K., 2011. Study on the tracking errors and their determinants: Evidence from Hong Kong exchange traded funds. Applied Financial Economics 21(5), 309-315.

Clarke, R.G, Krase, S. Statman, M., 1994. Tracking errors, regret, and tactical asset allocation. The Journal of Portfolio Management 20(3), 16-24.

DeFusco, R.A., Ivanov, S.I., Karels, G.V., 2011. The exchange traded funds' pricing deviation: analysis and forecasts. Journal of Economics and Finance 35(2), 181-197.

Driscoll, J.C., Kraay, A.C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. Review of economics and statistics, 80(4), 549-560.

Drenovak, M., Urošević, B., 2010. Exchange-traded funds of the Eurozone sovereign debt. Ekonomski anali 55(187), 31-60.

Goltz, F., 2009. The EDHEC-Risk European ETF Survey 2009.

Elton, E.J., et al., 2005. Spiders: where are the bugs? Springer Berlin, Heidelberg.

Engle, R.F., Sarkar, D., 2006. Premiums-discounts and exchange-traded funds. ETF and Indexing 2006(1), 35-53.

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

Frino, A., Gallagher, D.R., 2002. Is Index Performance Achievable? An Analysis of Australian Equity Index Funds. Abacus 38(2), 200-214.

Fulmek, J., 2003. Untersuchungen zum Thema Tracking Error. Working paper. Available from: www.ziel.at/content/_common/attachments/TrackingError.pdf

Harper, J.T., Madura, J., Schnusenberg, O., 2006. Performance comparison between exchange-traded funds and closed-end country funds. Journal of International Financial Markets, Institutions and Money 16(2), 104-122.

Houweling, P., 2012. On the Performance of Fixed-Income Exchange-Traded Funds. The Journal of Index Investing 3(1), 39-44.

Larsen Jr, G.A., Resnick, B.G., 1998. Empirical insights on indexing. The Journal of Portfolio Management 25(1), 51-60.

Merz, T., 2011. Kriteriologie des ETF-Ratings am Beispiel Emerging Markets, in: Everling, O., Kirchhoff, G.J. (Eds.), Exchange Traded Fund-Rating: Marktüberblick, Einsatzkriterien und Praxiseinsatz. Bank-Verlag Medien GmbH, Cologne, pp. 263-280.

Milonas, N.T., Rompotis, G.G., 2006. Investigating European ETFs: The case of the Swiss exchange traded funds, in: Paper for the Annual HFAA Conference in Thessaloniki, Greece.

Rudolf, M., Wolter, H.J., Zimmermann, H., 1999. A linear model for tracking error minimization. Journal of Banking & Finance 23(1), 85-103.

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

Rompotis, G.G., 2008. Performance and trading characteristics of German passively managed ETFs. International Research Journal of Finance and Economics 15, 210-223.

Rompotis, G.G., 2010a. Searching for seasonal patterns in exchange traded funds' trading characteristics. American Journal of Finance and Accounting 2(2), 155-180.

Rompotis, G.G. 2010b. Investigating the Tracking Error of Exchange Traded Funds, International Journal of Management Research 1, 23-42.

Satchell, S.E., Hwang, S., 2001. Tracking error: Ex ante versus ex post measures. Journal of Asset Management 2(3), 241-246.

Shin, S., Soydemir, G., 2010. Exchange-traded funds, persistence in tracking errors and information dissemination. Journal of Multinational Financial Management 20(4), 214-234.

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

Table 1

Sample Summary Statistics

This table reports the sample statistics of all European ETFs included in this study. The table includes the summarized characteristics of the European ETF sample in terms of number of indices replicated, number of ETFs analyzed, AuM (in Euro), total expense ratios (TER), the observed tracking difference at the end of the period on an annualized basis (Tracking Diff. p.a.), in-sample calculated quadratic tracking error measures according to equation (1), (2), and (3b), ($TE_{RMS(NCE)}$, $TE_{SD(CE)}$, TE_{REG},), in sample linear tracking error measures according to equation (4) and (5), respectively (TE_{MAD} and TE_{MADD}), and the number of observations included in the calculations (In-Sample Days and Out-of-Sample Days). Statistics are computed by full sample (All ETFs), by type of asset class exposure (equity ETFs and Fixed Inc. ETFs), and by replication strategy (Physical ETFs and Synthetic ETFs).

European ETF Sample	Full Sample	Asset Class		Replication Method	
	All ETFs	Equity ETFs	Fixed Inc. ETFs	Physical ETFs	Synthetic ETFs
No. of Indices repl.	26	16	10	20	18
No. of ETFs	131	112	19	66	65
AuM (EUR)	100'971'679'016	90'235'032'700	10'736'646'316	69'245'006'571	31'726'672'445
Avg. AuM (EUR)	782'726'194	820'318'479	565'086'648	1'049'166'766	503'597'975
Median AuM (EUR)	192'200'372	193'989'821	164'060'218	201'071'104	181'287'852
Avg. TER	0.34%	0.37%	0.20%	0.35%	0.33%
Median TER	0.30%	0.34%	0.20%	0.33%	0.30%
Min. TER	0.00%	0.00%	0.10%	0.09%	0.00%
Max. TER	1.05%	1.05%	0.45%	0.75%	1.05%
Avg. Tacking Diff. p.a.	-0.2638%	-0.2799%	-0.1706%	-0.2770%	-0.2500%
Median Tracking Diff. p.a.	-0.2432%	-0.2678%	-0.1546%	-0.2518%	-0.1959%
Min. Tracking Diff. p.a.	-1.3047%	-1.3047%	-0.6736%	-1.1850%	-1.3047%
Max. Tracking Diff. p.a.	0.8254%	0.8254%	0.0973%	0.8254%	0.6537%
Avg. $TE_{RMS(NCE)}$	0.2359%	0.2292%	0.2743%	0.3649%	0.1006%
Median $TE_{RMS(NCE)}$	0.1067%	0.1033%	0.2368%	0.1679%	0.0596%
Min. $TE_{RMS(NCE)}$	0.0111%	0.0170%	0.0111%	0.0253%	0.0111%
Max. $TE_{RMS(NCE)}$	2.1965%	2.1965%	1.7876%	2.1965%	0.6447%
Avg. $TE_{SD(CE)}$	0.2310%	0.2237%	0.2735%	0.3628%	0.0930%
Median $TE_{SD(CE)}$	0.1048%	0.0991%	0.2373%	0.1661%	0.0454%
Min. $TE_{SD(CE)}$	0.0086%	0.0089%	0.0086%	0.0152%	0.0086%
Max. $TE_{SD(CE)}$	2.1983%	2.1983%	1.7887%	2.1983%	0.6422%
Avg. TE_{REG}	0.2208%	0.2133%	0.2639%	0.3498%	0.0856%
Median TE_{REG}	0.0969%	0.0891%	0.2364%	0.1592%	0.0384%
Min. TE_{REG}	0.0082%	0.0088%	0.0082%	0.0152%	0.0082%
Max. TE_{REG}	2.1887%	2.1887%	1.6877%	2.1887%	0.5154%
Avg. TE_{MAD}	0.0094%	0.0092%	0.0108%	0.0145%	0.0041%
Median TE_{MAD}	0.0042%	0.0039%	0.0091%	0.0064%	0.0025%
Min. TE_{MAD}	0.0005%	0.0006%	0.0005%	0.0010%	0.0005%
Max. TE_{MAD}	0.1009%	0.1009%	0.0609%	0.1009%	0.0204%
Avg. TE_{MADD}	0.0092%	0.0089%	0.0107%	0.0143%	0.0038%
Median TE_{MADD}	0.0041%	0.0036%	0.0087%	0.0065%	0.0024%
Min. TE_{MADD}	0.0006%	0.0006%	0.0006%	0.0008%	0.0006%
Max. TE_{MADD}	0.1040%	0.1040%	0.0606%	0.1040%	0.0174%
Avg. In- sample Days	666	661	695	655	677
Median In-sample Days	604	604	581	598	630
Min. In-sample Days	445	563	445	445	563
Max. In-sample Days	1234	1140	1234	1234	1140
Avg. Out-of-sample Days	672	668	695	328	345
Median Out-of-sample Days	604	604	581	296	315
Min. Out-of-sample Days	445	563	445	222	282
Max. Out-of-sample Days	1234	1140	1234	617	570

Table 2

Tracking Error Estimation Quality from Out-of-Sample Tests

This table reports the results of the tracking error out-of-sample tests conducted. Panel A and B report the results for all ETFs in the sample comparing the out-of-sample quadratic tracking error ($TE_{SD(CE)}$ and TE_{REG}) as well as the two linear tracking error measures TE_{MAD} and TE_{MADD} with the observed tracking difference at the end of the out-of-sample period. In order to evaluate the quality of the quadratic tracking error estimates (out-of-sample), the in-sample TE has been forward-modeled with a 99% confidence interval by using equation (6) and (7) as well as (8) and (9), respectively. Panel A reports the percentage of ETFs in the sample where the observed tracking difference at the end of the out-of-sample period is smaller compared to the estimated tracking difference using the in-sample TE. It summarizes the results by categorizing the estimated tracking difference into two categories: overestimation of the tracking difference and underestimation of the tracking difference. Panel B reports the results applying a threshold of 10% in order to evaluate whether or not the estimated tracking difference is accurate enough.

Panel A: Over & Underestimation

	No.	in %	No.	in %	No.	in %	No.	in %
All ETFs	TESD		TEReg		TERMAD		TEMADD	
Overestimated	71	54.2%	66	50.4%	118	91.5%	116	89.9%
Underestimated	60	45.8%	65	49.6%	11	8.5%	13	10.1%
Equity ETFs								
Overestimated	52	46.4%	47	42.0%	99	90.0%	97	88.2%
Underestimated	60	53.6%	65	58.0%	11	10.0%	13	11.8%
Bond ETFs								
Overestimated	16	100.0%	16	100.0%	16	100.0%	16	100.0%
Underestimated	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Money Market ETFs								
Overestimated	3	100.0%	3	100.0%	3	100.0%	3	100.0%
Underestimated	0	0.0%	0	0.0%	0	0.0%	0	0.0%

Panel B: Within 10% Threshold

	No.	in %	No.	in %	No.	in %	No.	in %
All ETFs	TESD		TEReg		TEMAD		TEMADD	
Within 10%	78	59.5%	78	59.5%	13	10.1%	12	9.3%
Outside 10%	53	40.5%	53	40.5%	116	89.9%	117	90.7%
Equity ETFs								
Within 10%	59	52.7%	59	52.7%	13	11.8%	12	10.9%
Outside 10%	53	47.3%	53	47.3%	97	88.2%	98	89.1%
Bond ETFs								
Within 10%	16	100.0%	16	100.0%	0	0.0%	0	0.0%
Outside 10%	0	0.0%	0	0.0%	16	100.0%	16	100.0%
Money Market ETFs								
Within 10%	3	100.0%	3	100.0%	0	0.0%	0	0.0%
Outside 10%	0	0.0%	0	0.0%	3	100.0%	3	100.0%

Table 3

Determinants of Tracking Performance

This table reports the results from the OLS pooled regression on determinants of the tracking performance for sampled ETFs domiciled in Europe. Panel A summarizes the results for all equity ETFs in the European sample. The regression model includes the following independent variables: the average quadratic, non-centric tracking error measure (TEvol), physical replication (PHY), which is a dummy variable that takes on the value one for ETFs using physical replication, optimized replication (OPT), which is a dummy variable that takes on the value one for ETFs using an optimized model to replicate the benchmark index, securities lending (SEC), which is a dummy variable that takes on the value one for ETFs deploying securities lending activities, total expense ratio (TER), fund domicile CH (CH), fund domicile Germany (DE), fund domicile France (FR), fund domicile Ireland (IE), fund domicile Luxembourg (LU), autocorrelation of the return differences ETF vs. benchmark index (PEAlag1), and the 30-days average index volatility (INDvol) of the respective benchmark index. These variables are regressed (OLS regression) against the observed cumulative tracking difference (TRACKDIFFACC) using the following regression equation:

$$TDIFF_{i,t} = \alpha + \beta_1 TE_{vol_{i,t}} + \beta_2 PHY_{i,t} + \beta_3 OPT_{i,t} + \beta_4 SEC_{i,t} + \beta_5 TER_{i,t} + \beta_6 CH_{i,t} + \beta_7 DE_{i,t} + \beta_8 FR_{i,t} + \beta_9 IE_{i,t} + \beta_{10} LU_{i,t} + \beta_{11} PEA_{lag1_{i,t}} + \beta_{12} IND_{vol_{i,t}} + \varepsilon_{i,t}$$

The reported factor loads β1,…,β12 from the regression model are calculated as average numbers for each of the sixteen equity and the nine fixed income exposures. Equity exposures include SMI (SMIC), Dax (DAX), S&P500 (SPTR500N), MSCI USA (NDDUUS), Dow Jones Industrial Average (DJUS), FTSE100 (TUKXG), MSCI Brazil (NDUEBRAF), MSCI Japan (NDDUJN), MSCI Europe (MSDEE15N), STOXX Europe 600 (SXXR), Euro Stoxx 50 (SX5T), MSCI World (NDDUWI), MSCI Emerging Markets (NDUEEGF), MSCI Pacific ex Japan (NDDUPXJ), MSCI EM Latin America (NDUEEGFL), and NASDAQ 100 (NDX). Fixed income exposures include SBI Domestic Government 1-3 (SBGM1T), SBI Domestic Government 7-15 (SBGM7T), iBoox Euro Corporate Bonds (IB8A), Barclays Euro Corporate Bonds (LECPTREU), Euro Government Bonds (LEATRTREU), Euro Aggr. Bonds (LBEATREU), Global Govt. Bonds (SBG7U), Emerging Markets Bonds (JPEICORE), and Eonia (EONIA). Values denoted with ***, **, * indicate a significance at the 1%, 5%, and 10% level, respectively. The respective p-values are documented for each beta factor from the regression model. The tracking performance difference estimated by the regression model ($\widehat{TDIFF_i}$) using all beta factor is documented as Regression at the end of the table. The observed out-of-sample tracking difference between ETF and the benchmark index is reported as TRACKDIFFACC.

Panel A **Equity ETFs**	**SMIC**	**DAX**	**SPTR500N**	**NDDUUS**	**DJUS**	**TUKXG**	**NDUEBRAF**	**NDDUJN**	**MSDEE15N**	**SXXR**	**SX5T**	**NDDUWI**	**NDUEEGF**	**NDDUPXJ**	**NDUEEGFL**	**NDX**
No. of ETFs	5	4	9	9	3	10	6	5	6	6	16	8	9	7	4	4
R^2	0.83	0.93	0.62	0.69	0.94	0.91	0.91	0.99	0.64	0.83	0.47	0.95	0.86	0.96	0.99	0.90
$R^2_{adj.}$	0.68	0.86	0.38	0.47	0.88	0.82	0.83	0.97	0.40	0.68	0.20	0.90	0.74	0.93	0.97	0.81
T	1061	1140	586	641	761	604	572	647	662	849	591	692	580	563	636	630
Constant	-0.0058***	0.0006***	-0.0014***	0.0059***	0.0026***	-0.0006***	-0.0006***	-0.0094***	0.0006***	-0.3421***	-0.0011*	-0.0019***	-0.0033***	-0.0019***	-0.009***	-0.0031***
p-value	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*0.0700*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*
β1TEvoli	0	0	5.5485***	-8.9474***	0	-4.1144***	-3.5206***	0	4.9358***	0	2.4239***	9.1332***	-32.7734***	21.3201***	0	0
p-value			*<0.001*	*<0.001*		*<0.001*	*<0.001*		*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*		
β2PHYi	0.0209***	0.0719***	0.0034***	0.002***	-0.0045***	0	-0.0015***	-0.0709***	0.0023***	0	0.001***	0.0013***	0.0527***	-0.0024***	0.0008***	0.0017***
p-value	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*
β3OPTi	0	0	-0.0024***	-0.0002***	0	0	0	0	0.0014***	0	0	0	0	0	0	0
p-value			*<0.001*	*<0.001*					*<0.001*							
β4SECi	0.0061***	-0.0695***	0	-0.0074***	-0.009***	0.0005***	0	-0.0093***	0	0.3538***	0.0009***	-0.0071***	0	0.0012***	0	0.0003***
p-value	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*		*<0.001*		*<0.001*	*<0.001*	*<0.001*		*<0.001*		*<0.001*
β5TERi	-2.9496***	-1.4827***	0.0631***	-1.1222***	-3.1126***	-1.6146***	-1.7954***	-1.7112***	-0.4924***	-0.3068***	2.0905***	-1.1423***	-2.095***	-1.4257***	-2.1412***	2.7007***
p-value	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*
β6CHi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p-value																
β7DEi	-0.0186***	0	0	0	0	0	0	0	0	0	0.0065***	0	0	0	0	0
p-value	*<0.001*										*<0.001*					
β8FRi	0	0	0.002***	0	0	-0.0013***	-0.0032***	0	0	0.5883***	0	-0.002***	0.0044***	0.0009***	-0.0035***	0.0049***
p-value			*<0.001*			*<0.001*	*<0.001*			*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*
β9IEi	0	0	0	-0.0053***	0	-0.0006***	0	0.0172***	-0.0046***	0.2895***	-0.0021***	0	0	0	0	0
p-value				*<0.001*		*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*					
β10LUi	0	0	0.0011***	0.0004***	0	0	0	0	-0.0031***	0.2474***	0	0.0069***	0.0019***	0.0008***	0	0
p-value			*<0.001*	*<0.001*					*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*		
β11PEAlag1i	0.0033***	-0.1894***	0.0062***	0.0089***	0	-0.0055***	-0.0132***	-0.1655***	0.0048***	0.5239***	0.0083***	0.0145***	0.0265***	0.0021***	-0.0238***	0
p-value	*<0.001*	*<0.001*	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	*<0.001*	
β12INDvoli	0.0386**	-0.0249***	-0.0264***	0	0.0217***	-0.0085**	0.019***	-0.0111***	0	0.0249***	0.0677***	-0.0182*	-0.0636**	0	-0.3271***	-0.0505***
p-value	*0.0440*	*<0.001*	*<0.001*		*<0.001*	*0.0260*	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*0.0690*	*0.0150*		*<0.001*	*<0.001*
$\widehat{TDIFF}_i$	-0.0246	-0.0055	-0.0001	-0.0066	-0.0296	-0.0094	-0.0175	-0.0164	-0.0036	-0.0024	0.0095	-0.0114	-0.0215	-0.0105	-0.0283	0.0115
TRACKDIFFACC	-0.0338	-0.0057	0.0009	-0.0062	-0.0295	-0.0095	-0.0182	-0.0161	-0.0035	-0.0023	0.0102	-0.0108	-0.0234	-0.0113	-0.0236	0.0137

Table 3 (continued)

Determinants of Tracking Performance

Panel B **Fixed Income ETFs**	**SBGM1T**	**SBGM7T**	**IB8A**	**LECPTREU**	**LEATTREU**	**LBEATREU**	**SBG7U**	**JPEICORE**	**EONIA**
No. of ETFs	3	3	2	2	2	2	1	1	3
R^2	0.79	0.94	0.71	0.74	0.54	0.70	0.85	0.97	0.96
$R^2_{adj.}$	0.61	0.88	0.50	0.53	0.27	0.47	0.72	0.93	0.92
T	530	581	906	445	459	449	1002	1234	1039
Constant	-0.0408***	-0.0317***	-0.0019***	0.0003*	0.0021***	0	0.0017***	00	-0.0035***
p-value	*<0.001*	*<0.001*	*<0.001*	*0.0890*	*<0.001*		*<0.001*		*<0.001*
β1TEvoli	199.2275***	213.7063***	0	0	0	0	0	0	239.2476***
p-value	*<0.001*	*<0.001*							*<0.001*
β2PHYi	0	0	0.0013***	0	0	0	0	0	0
p-value			*<0.001*						
β3OPTi	0	0	0	0	0	0	0	0	0
p-value									
β4SECi	0	0	0	0.001***	-0.003***	0	0	0	0
p-value				*<0.001*	*<0.001*				
β5TERi	-0.7156***	-2.306***	-1.0311***	-1.3936***	0	-0.3481***	-1.1319***	-2.0709***	-1.3329***
p-value	*<0.001*	*<0.001*	*<0.001*	*<0.001*		*<0.001*	*<0.001*	*<0.001*	*<0.001*
β6CHi	0	0	0	0	0	0	0	0	0
p-value									
β7DEi	0	0	0	0	0	0	0	0	0
p-value									
β8FRi	0	0	0	0	0	0	0	0	0
p-value									
β9IEi	0	0	0	0	0	0	0	0	0
p-value									
β10LUi	0	0	0	0	0	0	0	0	0
p-value									
β11PEAlag1i	-0.1156***	-0.1377***	0	0	0	0.026***	0	0	-0.0035***
p-value	*<0.001*	*<0.001*				*<0.001*			*<0.001*
β12INDvoli	0	0	0.1759***	-0.3471***	-0.1513**	0	0	0.176***	2.3858***
p-value			*<0.001*	*<0.001*	*0.0270*			*<0.001*	*<0.001*
$\widehat{TDIFF}_i$	-0.0018	-0.0069	-0.0055	-0.0035	0.0002	-0.0008	-0.0030	-0.0309	-0.0059
TRACKDIFFACC	-0.0017	-0.0068	-0.0058	-0.0033	-0.0004	-0.0019	-0.0033	-0.0330	-0.0057

Table 4

Explanatory Power of Determinants of the Tracking Performance

This table reports the results from the OLS pooled regression on determinants of the tracking performance for sampled ETF domiciled in Europe. Panel A summarizes the results for all equity ETFs and Panel B for all fixed income ETFs in the European sample. The regression model includes the following independent variables: the average quadratic, non-centric tracking error measure (TEvol), physical replication (PHY), which is a dummy variable that takes on the value one for ETFs using physical replication, optimized replication (OPT), which is a dummy variable that takes on the value one for ETFs using optimized model to replicate the benchmark index, securities lending (SEC), which is a dummy variable that takes on the value one for ETFs deploying securities lending activities, the total expense ratio (TER), fund domicile CH (CH), fund domicile Germany (DE), fund domicile France (FR), fund domicile Ireland (IE), fund domicile Luxembourg (LU), autocorrelation of the return differences ETF vs. benchmark index (PEAlag1), and the 30-days average index volatility (INDvol) of the respective benchmark index. These variables are regressed (OLS regression) against the observed cumulative tracking difference using the following regression equation:

$$TDIFF_{i,t} = \alpha + \beta_1 TE_{vol_{i,t}} + \beta_2 PHY_{i,t} + \beta_3 OPT_{i,t} + \beta_4 SEC_{i,t} + \beta_5 TER_{i,t} + \beta_6 CH_{i,t} + \beta_7 DE_{i,t} + \beta_8 FR_{i,t} + \beta_9 IE_{i,t} + \beta_{10} LU_{i,t} + \beta_{11} PEA_{lag1_{i,t}} + \beta_{12} IND_{vol_{i,t}} + \varepsilon_{i,t}$$

The values for measuring the explanatory power are derived from the respective regression factor and are calculated as the percentage value of the total explained by the regression model. This table also reports the number of ETFs included, the average R^2 of the regression model as well as the total number of observations (T) included. The tracking performance difference estimated by the regression model ($\widehat{TDIFF}_i$) using all beta factor values is documented as Regression at the end of the table. The observed out-of-sample tracking difference between ETF and the benchmark index is reported as TRACKDIFFACC.

Panel A Equity ETFs	All ETFs	Physical ETFs	Synthetic ETFs	**Panel B** Fixed Income ETFs	All ETFs	Physical ETFs	Synthetic ETFs
No. of ETFs	112	51	61	No. of ETFs	19	15	4
R^2	0.77	0.78	0.77	R^2	0.80	0.78	0.96
$R^2_{adj.}$	0.59	0.60	0.58	$R^2_{adj.}$	0.63	0.60	0.92
T	668	666	668	T	695	612	1006
Influence $\beta_1 TEvol_i$	5.47%	8.57%	3.02%	Influence $\beta_1 TEvol_i$	11.66%	11.26%	13.18%
Influence $\beta_2 PHY_i$	8.06%	17.39%	0.18%	Influence $\beta_2 PHY_i$	0.90%	1.14%	0.00%
Influence $\beta_3 OPT_i$	0.58%	1.26%	0.00%	Influence $\beta_3 OPT_i$	0.00%	0.00%	0.00%
Influence $\beta_4 SEC_i$	5.13%	6.63%	3.86%	Influence $\beta_4 SEC_i$	3.91%	4.95%	0.00%
Influence $\beta_5 TER_i$	44.28%	39.09%	48.77%	Influence $\beta_5 TER_i$	36.84%	32.82%	51.88%
Influence $\beta_6 CH_i$	0.00%	0.00%	0.00%	Influence $\beta_6 CH_i$	0.00%	0.00%	0.00%
Influence $\beta_7 DE_i$	0.57%	1.22%	0.00%	Influence $\beta_7 DE_i$	0.00%	0.00%	0.00%
Influence $\beta_8 FR_i$	2.48%	0.00%	4.55%	Influence $\beta_8 FR_i$	0.00%	0.00%	0.00%
Influence $\beta_9 IE_i$	3.60%	4.66%	2.64%	Influence $\beta_9 IE_i$	0.00%	0.00%	0.00%
Influence $\beta_{10} LU_i$	3.03%	2.06%	3.87%	Influence $\beta_{10} LU_i$	0.00%	0.00%	0.00%
Influence $\beta_{11} PEAlag1_i$	11.84%	8.85%	14.26%	Influence $\beta_{11} PEAlag1_i$	13.67%	15.67%	6.17%
Influence $\beta_{12} INDvol_i$	2.77%	1.52%	3.79%	Influence $\beta_{12} INDvol_i$	3.28%	3.81%	1.28%
$\widehat{TDIFF}_i$	-0.00792	-0.00808	-0.00781	$\widehat{TDIFF}_i$	-0.00510	-0.00484	-0.00610
TRACKDIFFACC	-0.00806	-0.00830	-0.00789	TRACKDIFFACC	-0.00534	-0.00516	-0.00601