EKONOMIA I PRAWO. ECONOMICS AND LAW
Volume 24, Issue 2, June 2025
p-ISSN 1898-2255, e-ISSN 2392-1625
www.apcz.umk.pl/EiP.pl

Original article
received 20.02.2025; accepted 05.06.2025; published 30.06.2025
doi: 10.12775/EiP.2025.10

Patterns in ETF Tracking Errors: ESG vs. Non-ESG Passive Equity Exchange-traded funds

KATARZYNA DANILUK
corresponding author
University of Lodz, Faculty of Economics and Sociology, Department of International Finance and Investment, ul. POW 3/5, 90-255 Łódź, Poland
katarzyna.daniluk@uni.lodz.pl
orcid.org/0000-0003-3346-8788

Abstract

Motivation: Passive exchange-traded funds (ETFs) are designed to replicate index performance but face tracking errors due to costs and market inefficiencies. ESG ETFs, which incorporate environmental, social, and governance factors, have become increasingly popular in Europe. However, limited research has examined patterns in their tracking errors in comparison to non-ESG ETFs.
Aim: This study is among the first to systematically assess the determinants of tracking error in passive ESG equity ETFs listed on European exchanges and compare them to non-ESG counterparts. Employing dynamic panel GMM models, we investigate key factors influencing tracking error, including past tracking error, total expense ratio (TER), assets under management (AUM), benchmark volatility, and fund age.
Results: Analysing 48 ESG and 86 non-ESG equity ETFs from January 2021 to June 2024, we find that past tracking error, total expense ratio, and benchmark volatility significantly impact tracking error for both fund categories. Assets under management significantly affect tracking errors only in ESG ETFs. Fund age has no significant impact on either group. Although the Chow test indicated structural differences in how fund size and costs affect tracking errors in ESG versus non-ESG ETFs, these effects were not confirmed in the interaction

model. These findings indicate that the tracking error determinants are largely similar for ESG and non-ESG ETFs.

Keywords: Exchange-Traded Funds, ESG Investing, Tracking Error, Passive Investment
JEL: G14, G15, G23

1. Introduction

Passive exchange-traded funds (ETFs) play a key role in today's investment market, enabling investors to closely replicate the performance of specific market indices. Simultaneously, over the past few years, there has been growing interest in ESG investing in Europe, which, along with traditional financial measures, considers non-financial factors including environmental, social, and corporate governance (ESG) factors. According to Morningstar (2025), at the end of 2024, ESG-focused equity ETFs in Europe accounted for USD 325.3 billion in assets, and their share of the industry's total assets under management represented 11%. The passive ESG strategy entails the integration of ESG factors into the process of index creation. These ESG indices serve as the foundation for passive ESG exchange-traded funds. ETFGI (2024) reports that as of February 2023, there were 1,458 ESG ETFs listed globally with assets of USD 530.64 billion.

Even though the primary objective of both passive ESG ETFs and their non-ESG counterparts is to replicate the benchmark index, in practice, perfect replication remains unreachable due to transaction costs, market imperfections, and legal restrictions (Frino & Gallagher, 2001). The tracking ability of an ETF indicates the extent to which a fund follows a passive investment strategy (Roll, 1992). One key measure of a passively managed ETF's performance is the tracking error (TE), which measures the extent to which a fund's returns deviate from its underlying index.

Understanding the specific determinants of tracking error in passive ESG ETFs is crucial for ensuring accurate index replication, which is fundamental to building investor trust and fostering the growth of this market. Although several studies have explored tracking error determinants (Frino & Gallagher, 2001; Romp Otis, 2009; Chu, 2011; Qadan & Yagil, 2012; Blitz et al., 2012; Osterhoff & Kassirer, 2016), they primarily focus on non-ESG ETFs in the US and other developed markets, with limited research on ESG ETFs in Europe. Given the growing significance of ESG investing, understanding tracking error patterns in this segment is crucial for investors and fund managers.

This study addresses the gap in the literature on the determinants of tracking error of passive ESG ETFs listed in European exchanges by examining them in comparison to their non-ESG counterparts. Using dynamic GMM panel models, the study reveals key differences between these groups of funds.

The main objectives of the study are:

- To identify the determinants of tracking error (TE) in passive equity ETFs listed in European exchanges, accounting for the distinction between ESG and non-ESG funds.
- To compare the determinants of tracking error in passive ESG equity ETFs listed in European exchanges and their non-ESG counterparts.

The study answers the following research questions:

- What are the key determinants of tracking error of passive equity ESG ETFs listed in European exchanges and their non-ESG counterparts?
- Do passive ESG equity ETFs listed in European exchanges differ from their non-ESG counterparts in terms of key tracking error determinants?

The article is the first that identify and compare the determinants of tracking error for passive ESG and non-ESG equity ETFs listed in European exchanges. The results provide useful insights for investors and fund managers, indicating the specific tracking error patterns of passive ESG equity ETFs. The paper is based on the empirical examination of the tracking errors and their determinants of 48 passive broad market ESG equity ETFs listed on European exchanges compared to their 86 non-ESG counterparts over the period from January 1, 2021, to June 30, 2024. The results of the study indicate that lagged tracking error, total expense ratio, and benchmark risk impact the tracking error of both ESG and non-ESG ETFs, while the size of assets under management proved to have a significant impact only for the tracking error of ESG ETFs. Then, we demonstrated that determinants of tracking error are similar across ESG and non-ESG ETFs. The introduction of interaction terms in a pooled model revealed that none of the interaction effects were statistically significant, supporting the conclusion that tracking error determinants operate comparably in both fund types.

The article is structured as follows. The next section reviews the key factors influencing tracking error. This is followed by a presentation of the research methods. The next sections provide the results of the empirical analysis and discuss the results in the context of previous research. The last section provides conclusions, the limitations of the study, and the possible directions for future research.

2. Literature review

ETFs are designed to replicate the performance of an underlying index, but tracking errors persist due to various factors. The previous research on the tracking error has primarily focused on assessing the tracking error of ETFs listed on the US market (Elton et al., 2002; Poterba & Shoven, 2002), other developed markets (Gallagher & Segarra, 2006; Romposis, 2008; Johnson et al., 2013), as well as in emerging markets (Lin & Chou, 2006; Kuok-Kun Chu, 2011; Khan et al., 2015; Miziołek & Feder-Sempach, 2019). Feder-Sempach and Miziołek (2022) examined the tracking efficiency of 14 European ETFs replicating the Euro Stoxx 50 Index over the 2012–2021 period, covering both economic growth and the COVID-19 crisis. They found that ETFs, especially those with accumulating share classes, demonstrate high tracking precision and highlight that both the tracking error calculation method and return interval frequency significantly influence the results.

The ETF total expense ratio (Frino & Gallagher, 2001; Rompotis, 2009; Blitz et al., 2012; Osterhoff & Kaserer, 2016), ETF size (Chu, 2011), and market volatility (Qadan & Yagil, 2012; Drenovak et al., 2014) were indicated as the main factors determining the tracking error of ETFs. However, research on ESG ETFs remains scarce. Lee (2020) investigated the performance of ESG ETFs in Korea by price disparity ratio and tracking error. Then, Nguyen (2023) showed that ESG ETFs display low tracking errors during the COVID-19 market stress. Despite these contributions, little is known about the determinants of ESG ETF tracking errors in European markets.

The Sustainable Finance Disclosure Regulation (European Commission, 2021) shapes how ESG indices are constructed by requiring funds to disclose how they integrate sustainability and by encouraging the use of climate-focused benchmarks like the EU Climate Transition Benchmark (CTB) and the Paris-Aligned Benchmark (PAB), which often exclude high-emission sectors or companies that do not meet environmental criteria. These exclusions can limit diversification and change the index's risk structure (Ascioglu & Saatcioglu, 2024). Still, all passive ETFs—whether ESG or not—aim to closely follow their benchmark. They are exposed to the same market frictions, such as transaction costs, bid-ask spreads, and rebalancing needs, which affect tracking error regardless of the index type (Charupat & Miu, 2013). Therefore, even if ESG and non-ESG indices differ in composition, the main drivers of tracking error are expected to operate similarly in both types of ETFs.

Most ETFs are passive investment products whose ultimate goal is to track the performance of a selected index. Tracking error refers to the volatility of the differences in the returns of an ETF and its underlying benchmark (Elton and Gruber, 2013), which assesses the stability of these deviations. Previous research often indicates that the tracking errors of ETFs display an autoregressive nature. DeFusco et al. (2011) used the ARCH (Autoregressive Conditional Heteroskedasticity) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) models to describe the tracking errors and price deviations in equity ETFs. Similarly, Ivanov (2015a) demonstrated the usefulness of autoregressive models for modeling the tracking errors of currency ETFs. Based on the mentioned review, the first research hypothesis states that:

H1. The tracking errors of both ESG and non-ESG passive equity ETFs listed in European exchanges exhibit an autoregressive pattern.

One of the key factors affecting an ETF's ability to accurately track an index is its total expense ratio (TER). Previous research consistently shows a negative relationship between TER and ETF tracking accuracy. Numerous authors have evidenced that with an increase in the fund total expense ratio, its tracking error rises (Frino and Gallagher, 2001; Rompotis, 2009; Agapova, 2011; Blitz et al., 2012; Elia, 2012; Osterhoff & Kaserer, 2016). Charupat & Miu (2013) further confirm that a higher TER weakens ETF performance compared to the benchmark. Based on the review, we formulated the second research hypothesis:

H2. There is a positive relationship between the total expense ratio (TER) and the tracking error of both ESG and non-ESG passive equity ETFs listed on European exchanges.

Another study suggests that the ETF tracking error depends on the fund size. Chu (2011) found that larger funds tend to have lower tracking errors because they benefit from economies of scale. In contrast, smaller funds may struggle to maintain a portfolio structure that closely follows the index due to limited resources (Chu, 2011). Based on these findings, the third research hypothesis states that:

H3. There is a negative relationship between assets under management (AUM) and the tracking error of both ESG and non-ESG passive equity ETFs listed in European exchanges.

Then, Rompotis (2011) found that an ETF's age affects its tracking error. As newer funds often have lower liquidity, smaller asset bases, and higher transaction costs, they often display higher tracking errors. In contrast, older ETFs typically have larger assets and better cost optimization, which can help reduce tracking errors. Based on these findings, the fourth research hypothesis states that:

H4. There is a negative relationship between the fund age and the tracking error of both ESG and non-ESG passive equity ETFs listed in European exchanges.

Finally, Qadan and Yagil (2012) and Drenovak et al. (2014) argued that an increased volatility of the underlying benchmark leads to a higher tracking error of an ETF. Increased volatility raises transaction costs and requires more frequent portfolio rebalancing. Based on these findings, we formulated the fifth research hypothesis:

H5. There is a positive relationship between the underlying benchmark volatility and the tracking error of both ESG and non-ESG passive equity ETFs listed in European exchanges.

3. Methods

This study employs dynamic panel models using the Arellano-Bond GMM estimator to examine tracking error determinants in 48 ESG and 86 non-ESG ETFs from January 2021 to June 2024. This methodology accounts for potential endogeneity and autocorrelation (Arellano and Bond, 1991). Table 1 presents the inclusion and exclusion criteria, while Table 2 details the research sample.

The study examines key factors influencing tracking errors, including lagged tracking errors, total expense ratio (TER), assets under management (AUM), fund age, and benchmark volatility. The selection of variables is based on previous studies confirming their impact on the tracking error of non-ESG ETFs (Blitz et al., 2012; Poterba & Shoven, 2002; Elton et al., 2004; Johnson, 2009; Dyer & Guest, 2022). Table 3 provides detailed information on the variables used. The data were sourced from the Bloomberg database, which classifies ETFs into ESG and non-ESG categories.

We obtained the NAVs (Net Asset Values) of the funds from the database on a weekly basis to calculate logarithmic returns according to the formula:

$$R_{i,t} = \ln\left(P_{i,t} + D\right) - \ln\left(P_{i,t-1}\right), \tag{1}$$

where:

$R_{i,t}$ – logarithmic return of an ETF in t period

$P_{i,t}$ – NAV price of an ETF in t period

D – dividend paid during the t period (if applicable)

$P_{i,t-1}$ – NAV price of an ETF in $t-1$ period.

Analogously, we calculated logarithmic returns of ETF's underlying benchmarks as follows:

$$R_{b,t} = \ln\left(P_{b,t}\right) - \ln\left(P_{b,t-1}\right), \tag{2}$$

where:

$R_{b,t}$ – logarithmic return of the benchmark index in period t;

$P_{b,t}$ – closing price of the benchmark index in period t;

$P_{b,t}$ – closing price of the benchmark index in period $t-1$.

To calculate the tracking error, the standard deviation of the differences in fund and index returns was used (ESMA, 2012: 43). This method measures volatility relative to the benchmark and assumes no autocorrelation of differences. Using weekly data helps capture important trends while reducing the impact of short-term volatility (Frino and Gallagher, 2001; Rompotis, 2009). The monthly tracking error is calculated as the standard deviation of the weekly return differences between the ETF and its benchmark:

$$TE_{i,t} = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}\left(\left(R_{i,t} - R_{b,t}\right) - \frac{1}{n}\sum_{t=1}^{n}\left(R_{i,t} - R_{b,t}\right)\right)^2}, \tag{3}$$

where:

$TE_{i,t}$ – tracking error of the ETF i in the period t;

$R_{i,t}$ – the return of the ETF i in the period t;

$R_{b,t}$ – the return of the benchmark index b in the period t;

n – the number of periods.

Data used for the regression analysis were normalized using the min-max scaling method from Han et al. (2011: 51). The panel data regression was utilized to assess the impact of selected variables on the tracking error of ETFs.

First, two models were developed to explain the tracking error—one for ESG ETFs and another for non-ESG ETFs. The models are described by the following equation:

$$\begin{aligned}\mathrm{TE}_{i,t} = \alpha + \gamma \mathrm{TE}_{i,t-1} + \beta_1 \mathrm{TER}_{i,t} + \beta_2 \mathrm{AUM}_{i,t} + \beta_3 \mathrm{AGE}_{i,t} + \beta_4 \mathrm{RISK}_{i,t} \\ + \beta_5 \mathrm{ESG}_{i,t} + \beta_6 \mathrm{REP}_{i,t} + u_{i,t};\end{aligned} \tag{4}$$

where:

$TE_{i,t}$ – Tracking Error for ETF i at time t,

$TE_{i,t}$ – Tracking Error for ETF i at time t-1

α – the intercept representing the baseline level of TE when all other variables are zero,

γ, β_1, β_2, β_3, β_4, β_5, β_6 – estimated coefficients for the respective variables,

TER – total expense ratio of an ETF i at time t,

AUM – assets under management of an ETF i at time t,

AGE – age of an ETF i at time ***t*,**

RISK – the underlying benchmark i risk at time t,

REP – dummy variable equal to 1 if an ETF uses a synthetic replication method,

ESG – dummy variable equal to 1 if an ETF tracks an ESG index,

u_{it} – the error term representing the residuals of the model.

To assess whether the impact of key determinants differs between ESG and non-ESG ETFs, interaction terms between the ESG dummy variable and selected independent variables were introduced. In the regression model including interactions with the ESG variable, all continuous variables were first scaled to a 0–1 range using the min-max method. Subsequently, the scaled variables were centred on their mean values. Centring the variables improves the interpretation of the interaction coefficients and reduces potential collinearity problems (Aiken and West, 1991).

The following equation extends the baseline model by incorporating the interaction effects:

$$\begin{aligned} \mathrm{TE}_{i,t} = \alpha + \gamma \mathrm{TE}_{i,t-1} + \beta_1 \mathrm{TER}_{i,t} + \beta_2 \mathrm{AUM}_{i,t} + \beta_3 \mathrm{AGE}_{i,t} + \beta_4 \mathrm{RISK}_{i,t} \\ + \beta_5 \mathrm{ESG}_{i,t} + \beta_6 (ESG \times \mathrm{TER}_{i,t}) + \beta_7 (ESG \times \mathrm{AUM}_{i,t}) \\ + \beta_8 ((ESG \times \mathrm{AGE}_{i,t)} + \beta_9 (ESG \times \mathrm{RISK}_{i,t}) + u_{i,t}\ ; \end{aligned} \quad (5)$$

where:

$ESG \times TER_{i,t}$ – interaction of the ESG variable with the respective variable.

This study initially considered static panel data models, including fixed effects (FE) and random effects (RE). However, diagnostic tests revealed several statistical issues: heteroskedasticity (Breusch-Pagan test, BP = 67.23, p = 0.00), serial correlation (Wooldridge test, χ^2 = 845.56, p = 0.00; Breusch-Godfrey test, LM = 792.38, p = 0.00). These findings indicate that FE and RE models produce biased estimates and are unsuitable for analyzing tracking error determinants.

To address these issues, we applied the Arellano-Bond GMM estimator, which corrects for endogeneity and accounts for dynamic relationships in tracking error behaviour (Arellano and Bond, 1991; Blundell and Bond, 1998). Lagged values of tracking error are used as instruments, and their validity is confirmed through the Hansen test. To address the problem of heteroskedasticity, we used the robust standard errors. To verify the validity of the instrument set, we conduct the Arellano-Bond tests (AR (1) and AR (2)). Details of the linear dynamic panel model specifications are provided in Table 4. This approach allows for unbiased and efficient estimation of tracking error determinants.

Given the sample size (N = 134, T = 42), this study uses a significance level of 0.1 to capture potential relationships that might be missed with a stricter threshold. This is particularly useful in analyses with fewer observations or when identifying subtle effects (Greene, 2008: 498-506).

4. Results

Based on the estimations of the model depicting the tracking error (TE) in ESG ETFs (see Table 5), it can be seen that the key determinants of TE are past tracking error (lag(TE,1)), total expense ratio (TER), asset size (AUM) and benchmark risk (RISK). Notably, the lagged dependent variable has the strongest effect (coefficient of 0.16), indicating that tracking errors tend to persist over time. Additionally, the model indicates that a one-unit increase in TER raises TE by 0.02 units. Then, a one-unit increase in AUM reduces TE by 0.03 units. Finally, a one-unit increase in RISK increases TE by 0.03 units. The AGE variable is not statistically significant, which implies that the fund's age does not affect its tracking error.

In line with the model assumptions, the results of the diagnostic tests shown in Table 6 demonstrate first-order autocorrelation in the residuals, as indicated by the AR test (1). Next, the AR test (2) does not reveal significant second-order autocorrelation. The Hansen test confirms that the selected instruments are uncorrelated with the model errors and have been appropriately selected. Furthermore, the high significance level of the Wald test implies that the explanatory variables significantly predict the dependent variable. Overall, these results confirm the good model fit, as well as the stability and reliability of the estimates.

Based on the estimations of the model depicting the tracking error (TE) in non-ESG ETFs (see Table 7), we can see that the key determinants of TE are past tracking error (lag (TE,1)), total expense ratio (TER) and benchmark risk (RISK). Similarly to ESG ETFs, the lagged dependent variable has the strongest effect, with a coefficient of 0.10. Additionally, the model indicates that a one-unit increase in TER raises TE by 0.06 units. Next, a one-unit increase in RISK increases TE by 0.03 units. The AGE variable is not statistically significant. Notably, in opposition to ESG ETFs, assets under management (AUM) do not significantly affect the tracking error of non-ESG ETFs. The diagnostic tests shown in Table 8 validate the model. As intended, AR (1) detects first-order autocorrelation and AR (2) shows no second-order autocorrelation. The results of the Hansen test confirm the proper selection of instruments, and the Wald test confirms the joint significance of the parameters.

The Chow test results (Table 9) indicate that AUM has a significantly different impact on tracking errors between ESG and non-ESG ETFs. Then, we observe that TER and AGE exhibit significance at a 0.1 level, suggesting potential differences in their effects. In contrast, the impact of lagged tracking error and benchmark risk remains statistically similar across both fund types.

The estimations of the GMM model for the full sample (see Table 10) account for the interactions between all variables and the binary ESG variable. The model introduces the interaction terms with the ESG factor for only those variables that showed significant differences in the Chow test. The results of the diagnostic tests for this model are presented in Table 11. They confirm the model's validity and provide a foundation for reliable conclusions. The estimation results indicate that the ESG variable has a negative coefficient (-0.02), suggesting that ESG ETFs tend to exhibit lower tracking errors than non-ESG ETFs, after accounting for other factors. The positive and significant coefficient on the lagged tracking error means that tracking error is persistent over time. Finally, the negative and significant effect of the synthetic replication variable (REP) suggests that this replication method improves tracking accuracy.

However, after introducing interactions with the ESG variable, we observed relevant differences from previous models. Specifically, in contrast to individual models, we observed the lack of a significant impact of TER, AUM, and RISK on the tracking error of ETFs. This can be explained by increased variance when both fund types are combined, reducing the precision of coefficient estimates (Greene, 2008: 500–502). The inclusion of broader explanatory factors, such as the ESG dummy and replication method (REP), may also absorb some of the explanatory power of TER, AUM, and RISK (Aiken & West, 1991: 36–42). Finally, none of the interaction terms between ESG and the key explanatory variables is statistically significant. This means that the effects of TER and AUM on tracking error do not differ meaningfully between ESG and non-ESG ETFs.

In conclusion, our empirical analysis demonstrates that passive ESG and non-ESG equity ETFs share broadly similar tracking error determinants. Even though we confirmed that assets under management significantly reduce tracking errors in ESG ETFs, but do not affect the TE of non-ESG ETFs, these differences turned out to be insignificant in the extended model that included interaction terms with the ESG variable. Therefore, we showed that after controlling for the ESG classification and the replication method, differences in the tracking error determinants lose significance.

5. Discussion

Based on the study using the Arellano-Bond GMM models, we identified key tracking error (TE) determinants of ESG and non-ESG passive equity ETFs listed in European exchanges. The lagged tracking error proved to be the most significant factor influencing TE, with the highest estimates in all the composed models. Therefore, fund managers and investors should take into account the autoregressive pattern of tracking errors and mind that reducing them requires long-term corrective actions. These results are in line with Ivanov (2015b), who observed a strong link between current and past tracking errors in funds tracking currency indices. In that way, the study confirmed H1, stating that the tracking errors of both ESG and non-ESG passive equity ETFs listed in European exchanges exhibit an autoregressive pattern.

Next, the study demonstrated that the total expense ratio (TER) is a key factor influencing tracking error. In line with the previous research conducted by Frino and Gallagher (2001), Rompotis (2009), Agapova (2011), Blitz et al. (2012), Elia (2012), and Osterhoff and Kaserer (2016), we showed that higher costs lead to higher tracking errors. This relationship holds for both

ESG and non-ESG funds. In that way, we confirmed H2, which states that there is a positive relationship between the total expense ratio (TER) and the tracking error of both ESG and non-ESG passive equity ETFs listed on European exchanges.

On the other hand, we found that the impact of assets under management (AUM) on the tracking errors of ETFs varies depending on the type of fund. Specifically, in ESG ETFs, a larger AUM led to a reduction in tracking errors, while in non-ESG funds, the relationship was not statistically significant. The results contradict Chu (2011), who found the magnitude of tracking errors of ETFs traded in Hong Kong to be negatively related to size. The contradictory findings could be explained by a certain size threshold reached in many non-ESG funds, beyond which a further increase in AUM no longer results in a noticeable improvement in replication quality. Simultaneously, ESG ETFs still benefit from economies of scale. Therefore, based on the findings, we rejected hypothesis H3, which states a negative relationship between assets under management (AUM) and the tracking error of both ESG and non-ESG passive equity ETFs listed In European exchanges.

The lack of a significant effect of fund age (AGE) on the tracking error in any of the examined models contradicts some previous studies (Rompotis, 2011; Chu, 2011). However, this may be due to the increasing professionalization and standardization of operational processes in the European ETF market (ESRB, 2019). High regulatory standards and greater market transparency may have reduced the impact of a fund's maturity on its ability to track the index. Therefore, based on the results, we rejected hypothesis H4, stating that there is a negative relationship between the fund age and the tracking error of both ESG and non-ESG passive equity ETFs listed in European exchanges.

The analysis revealed that ESG and non-ESG funds experience substantial tracking error growth when their underlying index demonstrates higher risk levels. This is in line with the research of Vardharaj et al. (2004), who showed that market volatility increases trading costs, which makes index replication challenging for ETF managers. Our research validated H5, which states that there is a positive relationship between the underlying benchmark volatility and the tracking error of both ESG and non-ESG passive equity ETFs listed in European exchanges.

The research established that the method used for replication directly affects the tracking performance of ETFs. The negative REP variable confirms previous research, which indicated that ETFs that adopt synthetic replication methods have lower tracking errors than ETFs using physical replication (Elia, 2012; Osterhoff and Kaserer, 2016).

The results indicate that the determinants of tracking error do not differ significantly between passive ESG equity ETFs listed on European exchanges and their non-ESG counterparts. The Chow test showed some possible differences, especially in the role of fund size and costs. However, these differences were not confirmed in the interaction model, which gives a more direct and reliable test. The impact of TER and AUM on tracking error turned out to be context-dependent, appearing significant in separate models but losing explanatory power when structural differences between ESG and non-ESG ETFs, and the replication methods, are controlled for. The interaction terms were not significant, meaning that the effects of key factors like TER and AUM are similar for both ESG and non-ESG ETFs. Therefore, any differences found should be seen as weak and not statistically reliable.

6. Conclusion

This study contributes to a better understanding of factors affecting tracking errors of passive ESG and non-ESG equity ETFs listed in European exchanges. Using dynamic panel models, we verified key tracking error determinants of 48 ESG and 86 non-ESG ETFs. The research covered the period January 2021 to June 2024. We showed that the factors that determine replication quality consist of historical tracking performance, costs, together with benchmark volatility, replication method, and the ESG status of the underlying benchmark. Then, we proved that assets under management impact the tracking error only in ESG ETFs, suggesting that larger fund size may enhance their ability to track the index. Finally, we indicated that the fund age did not significantly affect the tracking error of ETFs, likely due to the growing professionalization of ETF management.

The research provides useful information to both investors and ETF providers. The findings show that ESG ETFs replicate their benchmark indices with a high level of accuracy and even tend to exhibit lower tracking errors than their non-ESG counterparts. As a result, investors can confidently include ESG ETFs in their portfolios without compromising on tracking quality. Investors should apply the same evaluation process to ESG ETFs that they would use for traditional non-ESG passive funds. The assessment of both ESG and non-ESG ETFs requires evaluation of their historical tracking error performance, together with their total expense ratio, benchmark index risk level, and the replication method. It is also important to consider the size of the fund when selecting an ESG ETF. ESG ETFs with larger assets tend to have lower tracking errors, which means that they follow their index more accurately. The study also challenges the results of some studies (Rompotis, 2011; Chu, 2011) about fund age importance while confirming the importance of universal tracking error determinants.

The study suggests that ESG ETF managers need to expand their fund size to achieve improved replication accuracy. The strategic growth of ESG ETFs requires specific promotional efforts to reach the target audience. Promoting passive ESG investing means clearly explaining its main strengths, such as low fees, alignment with sustainable goals, and the ability to follow index performance with accuracy. The ESG ETF market can grow faster if these funds are easier to access, for example, by listing them on more stock exchanges or offering them on global investment platforms.

This study has some limitations that might be addressed in future research. Specifically, we focused on one unified market of passive broad-market equity ETFs listed in European exchanges to provide the baseline for comparisons. This limits the generalizability of the findings to other markets. Additionally, due to the novelty of ESG ETFs, the study period (2021–2024) is relatively short. This prevents the identification of long-term patterns influencing ETF tracking errors. Therefore, future research should consider other geographical regions and extend the time horizon. Furthermore, including additional factors, such as market liquidity and index composition, might provide a deeper understanding of the mechanisms influencing the tracking error of ESG and non-ESG ETFs.

References

Agapova, A. (2011). Conventional mutual index funds versus exchange-traded funds. *Journal of Financial Markets*, 14(2), 323–343. https://doi.org/10.1016/j.finmar.2010.10.005.

Aiken, L.S. & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Arellano, M., and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies.* 58(2), p. 277. https://doi.org/10.2307/2297968.

Ascioglu, A., Saatcioglu, K., 2024. Factor exposures of environmental, social, and governance (ESG) indexes. Review of Financial Economics. https://doi.org/10.1002/rfe.1200.

Blitz, D., Huij, J. & Swinkels, L. (2012). The performance of European index funds and exchange-traded funds: European index funds and exchange-traded funds. *European Financial Management*, 18(4), 649–662. https://doi.org/10.1111/j.1468-036x.2010.00550.x.

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), pp. 115–143. https://doi.org/10.1016/S0304-4076(98)00009-8.

Charupat, N. & Miu, P. (2013). The pricing efficiency of leveraged exchange-traded funds: Evidence from the u.S. Markets: Pricing efficiency of leveraged exchange-traded funds. *Journal of Financial Research*, 36(2), 253–278. https://doi.org/10.1111/j.1475-6803.2013.12010.x.

Chu, P.K.-K. (2011). Study on the tracking errors and their determinants: evidence from Hong Kong exchange traded funds. *Applied Financial Economics*, *21*(5), 309–315. https://doi.org/10.1080/09603107.2010.530215.
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. https://doi.org/10.1007/s12197-009-9090-6.
Drenovak, M., Urošević, B. & Jelic, R. (2014). European bond ETFs: Tracking errors and the sovereign debt crisis. *European Financial Management*, *20*(5), 958–994. https://doi.org/10.1111/j.1468-036X.2012.00649.x.
Dyer, T., Guest, N. (2022). A tale of two index funds: Full replication vs. representative sampling. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4057537
Elia, M. (2012). Tracking error of traditional and synthetic European exchange-traded funds. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2084929.
Elton, E.J., Gruber, M.J., Comer, G. & Li, K. (2002). Spiders: Where are the bugs? *The Journal of Business*, 75(3), 453–472. https://doi.org/10.1086/339891.
ETFGI. (2024). ETFGI Reports Assets of 530.64 Billion US Dollars Invested in ESG ETFs Listed Globally. Retrieved February 3, 2025, from https://etfgi.com/news/press-releases/2024/03/etfgi-reports-assets-53064-billion-us-dollars-invested-esg-etfs-listed#:~:text=At%20the%20end%20of%20February%202024%2C%20ESG%20ETFs,providers%20listed%20on%2046%20exchanges%20in%2036%20countries.
European Commission. (2021). Regulation (EU) 2019/2088 on sustainability-related disclosures in the financial services sector (Sustainable Finance Disclosure Regulation – SFDR). *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R2088
European Securities and Markets Authority. (2012). Consultation paper: Guidelines on ETFs and other UCITS issues (ESMA/2012/44). https://www.esma.europa.eu/sites/default/files/library/2015/11/2012-44_0.pdf.
European Systemic Risk Board (ESRB). (2019). *Can ETFs contribute to systemic risk?* European Systemic Risk Board Report, Frankfurt: ESRB. Available at: https://www.esrb.europa.eu/pub/pdf/asc/esrb.asc190617_9_canetfscontributesystemicrisk~983ea11870.sv.pdf.
Feder-Sempach, E. & Miziołek, T. (2022). How Precisely European Equity ETFs Mirror Their Flagship Benchmarks? Evidence from Funds Replicating Performance of Euro Stoxx 50 Index. Journal of Asset Management. 24(2), 121-135. https://doi.org/10.1057/s41260-022-00287-9
Frino, A. & Gallagher, D.R. (2001). Tracking S&P 500 index funds. *The Journal of Portfolio Management*, *28*(1), 44–55. https://doi.org/10.3905/jpm.2001.319822.
Gallagher, D.R., & Segara, R. (2006). The Performance and Trading Characteristics of Exchange-Traded Funds. *Journal of Investment Strategy*, 1(1), 47–58.

Greene, W.H. (2008). Econometric analysis. 6th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
Han, J., Kamber, M. & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.).* Morgan Kaufmann.
Ivanov, S.I. (2015a). An empirical examination of currency ETFs tracking error. *International Journal of Bonds and Derivatives*, 1(4), 333. https://doi.org/10.1504/ijbd.2015.073794.
Ivanov, S.I. (2015b). Intraday analysis of currency ETFs. *International Journal of Managerial Finance*, 11(4), 438–450. https://doi.org/10.1108/ijmf-10-2014-0161.
Johnson, B., Bioy, H., Kellett, A. & Davidson, L. (2013). On the right track: *measuring tracking efficiency in ETFs. The Journal of Index Investing*, 4(3), 35–41. https://doi.org/10.3905/jii.2013.4.3.035.
Johnson, W.F. (2009). Tracking errors of exchange traded funds. *Journal of Asset Management*, 10(4), 253–262. https://doi.org/10.1057/jam.2009.10
Khan, A., Bacha, O. & Masih, A.M. (2015). Performance and trading characteristics of exchange Traded Funds: Developed vs emerging markets. *Capital Markets Review*, 23(1 & 2), 40–64. https://econpapers.repec.org/article/mfajournl/v_3a23_3ay_3a2015_3ai_3a1&2_3ap_3a40-64.htm.
Lee, S.W. (2020). Performance of ESG ETF by price disparate ratio and tracking error. *Korea Association of Business Education*, *35*(6), 309–329. https://doi.org/10.23839/kabe.2020.35.6.309.
Lin, A. & Chou, A.-T. (2006). *The tracking error and premium/ discount of Taiwan's first exchange traded fund.* Retrieved February 3, 2025 from https://www.semanticscholar.org/paper/ac5bd87cb7d4f2ac6740373998dcb1efe78f6a5b.
Miziołek, T. & Feder-Sempach, E. (2019). Tracking ability of exchange-traded funds. Evidence from Emerging Markets Equity ETFs. *Bank i Kredyt*, 50(3), 221–248.
Morningstar. (2025). European ETFs Attract Record Inflows in 2024. Retrieved February 3, 2025 from https://www.morningstar.co.uk/uk/news/259560/european-etfs-attract-record-inflows-in-2024-.aspx.
Nguyen, H. (2023). COVID-19: PERFORMANCE OF ESG ETFS AND, ESG ETFS VS. THEIR DECLARED INDEXES. *Applied Finance Letters*, 12(1), 33–43. https://doi.org/10.24135/afl.v12i1.647.
Osterhoff, F. & Kaserer, C. (2016). Determinants of tracking error in German ETFs – the role of market liquidity. *Managerial Finance*, 42(5), 417–437. https://doi.org/10.1108/mf-04-2015-0105.
Poterba, J.M. & Shoven, J. B. (2002). Exchange-traded funds: A new investment option for taxable investors. *American Economic Review*, 92(2), 422–427. https://doi.org/10.1257/000282802320191732.
Qadan, M. & Yagil, J. (2012). On the dynamics of tracking indices by exchange traded funds in the presence of high volatility. *Managerial Finance*, 38(9), 804–832. https://doi.org/10.1108/03074351211248162.

Roll, R. (1992). A mean/variance analysis of tracking error. *The Journal of Portfolio Management*, 18(4), 13–22. https://doi.org/10.3905/jpm.1992.701922.
Rompotis, G.G. (2009). Interfamily competition on index tracking: The case of the vanguard ETFs and index funds. *Journal of Asset Management*, 10(4), 263–278. https://doi.org/10.1057/jam.2009.11.
Rompotis, G.G. (2011). Predictable Patterns in ETF's Return and Tracking Error. *Studies in Economics and Finance*, *28*(1), 14–35.
Vardharaj, R., Fabozzi, F. & Jones, F. n.d. *Determinants of Tracking Error for Equity Portfolios | Semantic Scholar.* https://www.semanticscholar.org/paper/Determinants-of-Tracking-Error-for-Equity-Vardharaj Fabozzi/a03131ac5c238c586e1bbc34b5e2fa77b7de4f8a.

Appendix

Table 1. Summary of inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
fund listed in the European market	fund established after January 1, 2021
passively managed fund	leveraged fund
broad-market equity fund	inverse fund
fund currency consistent with the currency of the benchmark index	fund adopting currency hedging
existence of an ESG/non-ESG counterpart	

Source: Own preparation.

Table 2. Number of ETFs under study in division to geographical exposure

Category	Index Short Name	ESG ETFs	Non-ESG ETFs	Total
Global Indices	FTSE Developed Index	1	2	33
	MSCI World Index	5	7	
	MSCI World Minimum Volatility Index	2	2	
	FTSE Emerging Index	1	2	
	MSCI Emerging Markets Index	3	8	
Regional Indices	MSCI Europe Index	8	9	50
	MSCI Europe Minimum Volatility Index	1	3	
	STOXX Europe 600 Index	3	6	
	MSCI EMU Index	5	9	
	MSCI Pacific ex Japan	1	5	

Category	Index Short Name	ESG ETFs	Non-ESG ETFs	Total
Country-specific Indices	MSCI Japan Index	4	6	51
	FTSE Japan Index	1	2	
	S&P 500 Index	4	12	
	MSCI USA Index	6	11	
	MSCI USA Small Cap Index	2	1	
	MSCI USA Minimum Volatility Index	1	1	
Total		48	86	134

Source: Own preparation based on Bloomberg database.

Table 3. Explanation of variables used in the study.

Variable symbol	Description
TE	Dependent variable, tracking error of an ETF calculated as the standard deviation of the return differences between an ETF and its benchmark)
TER	Independent variable, total expense ratio of an ETF: the annual cost of holding an ETF, expressed as a percentage of the fund's average net assets. This is a measure of the total costs associated with managing and operating an investment fund, TER includes management fees, administrative fees, operating costs, custodian fees, and registration fees.
AUM	Assets under management of an ETF. This is a key metric to assess the size of the ETF and refers to the total market value of the assets held within the ETF's portfolio (securities, cash holdings, derivatives, commodities, and any other financial instruments the ETF might hold).
AGE	Natural logarithm of an ETF age, the difference between the current date and the inception date
RISK	Benchmark risk calculated as the standard deviation of an index monthly logarithmic returns
REP	Independent dummy variable equal to 1 if an ETF uses a synthetic replication method
ESG	Independent dummy variable equal to 1 if an ETF tracks an ESG index

Source: Own preparation.

Table 4. Specifications of the linear dynamic panel models

Specification	Description
Lagged Dependent Variable	$TE_{i,t-1}$ (First-order lag of tracking error)
Instrumental Variables	lag (TE_2)
Unit Effects	Individual fixed effects
Estimation Approach	One-step Generalized Method of Moments (GMM), Arellano-Bond
Data Transformation	First-difference transformation (Δ)
Robust Standard Errors	Heteroskedasticity and autocorrelation-consistent (HAC) robust standard errors

Source: Own preparation.

Table 5. Estimation results of the GMM model for ESG ETFs

Model Formula	**TE_1 ~ lag (TE, 1) + TER + AUM + RISK + AGE | lag(TE_1, 3)**			
Variable	**Estimate**	**Std error**	**Statistic**	**p-value**
lag(TE_1, 1)	0.16	0.05	3.03	0.00
TER	0.02	0.01	1.81	0.07
AUM	-0.03	0.01	-3.48	0.00
RISK	0.05	0.01	3.47	0.00
AGE	0.03	0.02	1.51	0.13

Source: Own preparation.

Table 6. Diagnostic tests of the GMM model for ESG ETFs

Test	**Statistic**	**p-value**
Autocorrelation Test AR(1)	-3.67	0.00
Autocorrelation Test AR(2)	-0.32	0.75
Hansen Test	χ^2 = 47.95	0.99
Wald Test	χ^2 = 126.74	0.00

Source: Own preparation.

Table 7. Estimation results of the GMM model for non-ESG ETFs

Model Formula	**TE_1 ~ lag (TE, 1) + TER + AUM + RISK + AGE | lag (TE_1, 24)**			
Variable	**Estimate**	**Std error**	**Statistic**	**p-value**
lag (TE_1, 1)	0.10	0.05	1.93	0.05
TER	0.06	0.02	3.10	0.00
AUM	0.08	0.05	1.60	0.11
RISK	0.03	0.01	2.74	0.01
AGE	-0.02	0.02	-1.19	0.24

Source: Own preparation.

Table 8. Diagnostic tests of the GMM model for non-ESG ETFs

Test	**Statistic**	**p-value**
Autocorrelation Test AR(1)	-4.67	0.00
Autocorrelation Test AR(2)	-0.64	0.52
Hansen Test	χ^2 = 116.97	0.22
Wald Test	χ^2 = 86.45	0.00

(Source: Own preparation)

Table 9. Chow test results for GMM models of ESG and non-ESG ETFs

Variable	ESG Estimate	Non-ESG Estimate	Z Statistic	p-value
lag (TE_1, 1)	0.16	0.10	0.85	0.40
TER	0.02	0.06	-1.79	0.07
AUM	-0.03	0.08	-2.16	0.03
RISK	0.05	0.03	1.41	0.16
AGE	0.03	-0.02	1.77	0.08

Source: Own preparation.

Table 10. Estimation results of the GMM model including interactions with the ESG variable for the full sample of ETFs

Model Formula	TE_1 ~ lag (TE, 1) + TER + AUM + RISK + AGE + ESG + REP + TER: ESG + AUM: ESG | lag (TE_1, 20)			
Variable	**Estimate**	**Std error**	**Statistic**	**p-value**
ESG	-0.05	0.01	-4.94	0.00
lag (TE_1, 1)	0.16	0.10	1.55	0.00
TER	0.01	0.01	0.98	0.61
AUM	0.01	0.04	0.27	0.84
RISK	-0.01	0.01	-0.76	0.43
AGE	0.02	0.02	1.01	0.37
REP	-0.02	0.01	-3.39	0.00
ESG: lag (TE, 1)	0.46	0.32	1.44	0.15
ESG: TER	-0.02	0.04	0.71	0.21
ESG: AUM	-0.08	0.06	-1.30	0.19

Source: Own preparation)

Table 11. Diagnostic tests of the GMM model including interactions with the ESG variable for the full sample of ETFs

Test	Statistic	p-value
Autocorrelation Test AR(1)	-4.22	0.00
Autocorrelation Test AR(2)	0.51	0.60
Hansen Test	χ^2 = 110.40	0.22
Wald Test	χ^2 = 243.72	0.00

Source: Own preparation.