Bank i Kredyt 57(5), 2026, 569-588
DOI: 10.5604/01.3001.0055.9228

BANK I KREDYT
BANK & CREDIT

Analiza porównawcza błędu odwzorowania funduszy ETF w podziale na filary E, S oraz G

Michał Stańczyk, Monika Czerwonka*

Nadesłany: 15 stycznia 2026 r. Zaakceptowany: 13 kwietnia 2026 r.

Jak cytować: Stańczyk M., Czerwonka M. (2026), Analiza porównawcza błędu odwzorowania funduszy ETF w podziale na filary E, S oraz G, *Bank i Kredyt*, 57(5), 569–588, DOI: 10.5604/01.3001.0055.9228.

Streszczenie

Celem pracy jest analiza wpływu kategorii czynników zrównoważonego rozwoju (Environmental, Social, Governance) na dokładność odwzorowania indeksu w latach 2020–2024.

Do oceny dokładności śledzenia wykorzystano zarówno błąd odwzorowania (*tracking error*, TE), jak i różnicę odwzorowania (*tracking difference*, TD). Ponadto zastosowano miary alternatywne, w tym błąd odwzorowania oparty na średnim odchyleniu bezwzględnym (TE-MAD), błąd odwzorowania kroczącego (okna 60-dniowe) oraz błąd odwzorowania spadkowego. Do oceny statystycznej zastosowano metryki regresyjne (alfa, beta, R^2 i RMSE). Właściwości dystrybucyjne stóp zwrotu zbadano z wykorzystaniem skośności, kurtozy i testu Jarquego-Bery.

Fundusze ETF o charakterze środowiskowym charakteryzują się najmniejszym błędem odwzorowania i najwyższą spójnością replikacji benchmarku, pomimo relatywnie wyższej zmienności stóp zwrotu. Z kolei fundusze ETF reprezentujące kategorię Governance wykazują systematycznie wyższe błędy odwzorowania i słabsze dopasowanie regresji, co wskazuje na mniej precyzyjną replikację.

Według wiedzy autorów jest to jedno z pierwszych badań wpływu kategorii czynników zrównoważonego rozwoju na dokładność odwzorowania indeksu przez fundusz ETF.

Słowa kluczowe: błąd odwzorowania, ETF, ESG, rynek kapitałowy

JEL: G11, G12, G23, Q56

* Szkoła Główna Handlowa w Warszawie; e-mail: monika.czerwonka@sgh.waw.pl.

1. Wprowadzenie

Fundusze typu *exchange-traded funds* (ETF) stanowią obecnie jeden z najdynamiczniej rozwijających się segmentów instrumentów finansowych notowanych na giełdach, szczególnie w sferze inwestycji pasywnych i długoterminowych. Wzrost ich popularności bezpośrednio wiąże się z tym, że inwestorzy coraz większą wagę przywiązują do kosztów. Przejawia się to preferowaniem instrumentów o niskim wskaźniku kosztów całkowitych (*total expense ratio* – TER), który w przypadku funduszy ETF jest zazwyczaj znacznie niższy niż w tradycyjnych funduszach aktywnie zarządzanych. Inwestorzy coraz częściej dostrzegają, że wysokie opłaty za zarządzanie w dłuższej perspektywie istotnie obniżają stopę zwrotu z portfela, co skłania ich do wybierania rozwiązań indeksowych.

Zarówno w skali globalnej, jak i na polskim rynku kapitałowym obserwuje się systematyczny wzrost zainteresowania sektorem funduszy indeksowych. Potwierdzają to dane rynkowe, wskazujące na dynamiczny przyrost wartości aktywów pod zarządzaniem (*assets under management* – AuM) ulokowanych w tych instrumentach, oraz dodatnie saldo napływów kapitału netto (Miziołek 2023). Wiąże się to z upowszechnieniem edukacji finansowej, która dostarcza inwestorom wiedzy o mechanizmach działania replikacji fizycznej i syntetycznej oraz o korzyściach płynących z dywersyfikacji przy niskim progu wejścia, co czyni ETF-y atrakcyjnym narzędziem dla szerokiego grona odbiorców.

Równolegle kluczową rolę na rynkach finansowych zaczęła odgrywać koncepcja zrównoważonego rozwoju. Wynika to przede wszystkim z intensyfikacji wymogów regulacyjnych, zainicjowanej m.in. przez Unię Europejską w ramach planu działania na rzecz finansowania zrównoważonego wzrostu (np. rozporządzenie SFDR czy taksonomia UE). Regulacje te nakładają na instytucje finansowe obowiązek implementacji procedur dotyczących raportowania niefinansowego oraz uwzględniania ryzyka klimatycznego w procesach inwestycyjnych (por. Pyka, Nocoń 2026; Czerwonka 2025). Wzrost zainteresowania inwestycjami społecznie odpowiedzialnymi (*socially responsible investing* – SRI), realizowanymi również za pośrednictwem przeznaczonych do tego funduszy ETF, staje się widoczny także w segmencie inwestorów indywidualnych (Rompotis 2022). Osiągnięcie spójności w tym obszarze wymaga symetrycznego zaangażowania: ze strony emitentów za pośrednictwem strategii społecznej odpowiedzialności biznesu (CSR) oraz ze strony inwestorów przez skierowanie kapitału do podmiotów spełniających wysokie standardy etyczne i środowiskowe (por. Dziawgo 2009).

Najczęściej stosowanym podejściem do analizy wpływu czynników niefinansowych na instrumenty giełdowe jest kompleksowa ocena oparta na zagregowanym ratingu ESG (ESG score). W niniejszym badaniu skoncentrowaliśmy się jednak na dekompozycji tego wpływu, analizując oddziaływanie poszczególnych komponentów (E, S oraz G) na błąd odwzorowania (*tracking error* – TE) oraz wyniki inwestycyjne (*performance*) funduszy ETF. Zróżnicowana natura danych niefinansowych oraz wielość metod ich interpretacji determinują profil ryzyka inwestycyjnego i wymuszają na zarządzających specyficzne podejście do konstrukcji portfela. Wspomniane wcześniej uwarunkowania regulacyjne wpływają na zmienność oraz sposób replikacji indeksów, co może mieć odmienne skutki w perspektywie krótko- i długoterminowej. Istotnym czynnikiem różnicującym jest także dynamika zmian składu portfeli indeksowych, dobór agencji ratingowych oraz zróżnicowana metodyka stosowania wag i filtrów wykluczających. Wszystkie te elementy bezpośrednio rzutują na efektywność naśladowania benchmarku przez fundusz.

Pomimo dynamicznego rozwoju literatury poświęconej zrównoważonemu inwestowaniu wiele obszarów wciąż wymaga pogłębionej analizy empirycznej. Dotychczasowa aktywność badawcza koncentrowała się na ocenie wyników inwestycyjnych strategii ESG oraz na wpływie czynników środo-

wiskowych na wycenę aktywów (np. Pastor, Stambaugh, Taylor 2023; Bolton, Kacperczyk 2023). Wiele uwagi poświęcono również mechanizmom funkcjonowania funduszy ETF oraz ich oddziaływaniu na płynność i efektywność rynków (np. Ben-David i in. 2023). Relatywnie niewiele miejsca zajmuje jednak bezpośrednia analiza jakości odwzorowania benchmarków przez fundusze oparte na indeksach zrównoważonego rozwoju. Co więcej, rzadko podejmuje się próbę dekompozycji tego wpływu na poszczególne filary koncepcji ESG: środowiskowy (Environmental), społeczny (Social) oraz ład korporacyjny (Governance). W praktyce instrumenty te istotnie różnią się konstrukcją, poziomem koncentracji sektorowej oraz kryteriami selekcji spółek, co bezpośrednio wpływa na zdolność funduszu do precyzyjnej replikacji indeksu. Niniejsze badanie stanowi odpowiedź na wskazane ograniczenia i oferuje kompleksową ocenę jakości odwzorowania benchmarków ESG w podziale na filary E, S oraz G.

Po przeprowadzeniu przeglądu literatury oraz zdefiniowaniu celu badania sformułowano następujące hipotezy badawcze:

H1. Istnieją statystycznie istotne różnice w poziomie błędu odwzorowania pomiędzy funduszami ETF koncentrującymi się na poszczególnych filarach zrównoważonego rozwoju: E, S i G.

H2. Wybór bazy danych (wartość aktywów netto – NAV vs. ceny rynkowe) istotnie wpływa na uzyskiwane wartości błędu odwzorowania funduszy ETF.

H3. Efektywność replikacji benchmarków, mierzona za pomocą współczynnika determinacji (R^2) oraz błędu średniokwadratowego (RMSE), istotnie zmienia się zależnie od przynależności funduszu do określonej grupy czynników ESG.

H4. Rozkłady stóp zwrotu funduszy ETF z segmentu ESG odbiegają od rozkładu normalnego oraz różnią się pod względem skośności i kurtozy pomiędzy grupami E, S i G.

Struktura artykułu została podporządkowana realizacji celu badawczego. W pierwszej części przybliżono teoretyczne aspekty zrównoważonych indeksów giełdowych, ze szczególnym uwzględnieniem metodyki pomiaru błędu odwzorowania. Część druga stanowi syntetyczny przegląd literatury przedmiotu w zakresie wpływu ratingów ESG na wyniki inwestycyjne oraz efektywność replikacji funduszy ETF. W końcowej części opracowania przedstawiono wyniki badań własnych oraz sformułowano wnioski końcowe.

2. Zrównoważone indeksy giełdowe

2.1. Specyfika i klasyfikacja indeksów ESG

Indeksy ESG (Environmental, Social and Governance) są to instrumenty bazowe (benchmarki), przy których konstrukcji wykroczono poza tradycyjne kryteria kapitalizacji rynkowej i w procesie selekcji komponentów uwzględniono czynniki pozafinansowe. Według literatury przedmiotu (zob. Schoenmaker, Schramade 2018) czynniki te zalicza się do trzech filarów. Filar E – środowiskowy (Environmental) – obejmuje aspekty związane z oddziaływaniem na ekosystem, w tym intensywność emisji gazów cieplarnianych, efektywność energetyczną oraz gospodarkę obiegu zamkniętego. Filar S – społeczny (Social) – odnosi się do kapitału ludzkiego, standardów pracy, różnorodności (*gender balance*) oraz relacji z lokalnymi społecznościami. Filar G – ład korporacyjny (Governance) – dotyczy struktury organów nadzorczych, etyki biznesu, przejrzystości podatkowej oraz ochrony praw akcjonariuszy mniejszościowych.

W praktyce indeksy te odgrywają rolę filtrów inwestycyjnych. Przy ich konstrukcji stosuje się podejście *positive screening* (premiowanie liderów zrównoważonego rozwoju) lub *negative screening* (wykluczanie branż kontrowersyjnych, takich jak sektor tytoniowy, zbrojeniowy czy wydobycie paliw kopalnych). Zgodnie z typologią rynkową, np. metodyką MSCI, oraz badaniami Amel-Zadeha i Serafeima (2018) indeksy zrównoważone można podzielić na:

1) *ESG leaders*: obejmujące spółki o najwyższym scoringu w danej klasie aktywów;

2) *single pillar* (*E/S/G-only*): skoncentrowane wyłącznie na wybranym aspekcie, np. niskoemisyjności;

3) *best-in-class*: selekcjonujące liderów w ramach poszczególnych sektorów przy zachowaniu neutralności branżowej;

4) *exclusionary*: oparte na szerokich indeksach rynkowych, ale pomijające podmioty, które naruszają normy (np. UN Global Compact);

5) *thematic*: ukierunkowane na konkretne trendy, np. technologie czystej energii.

Współczesny rynek ESG ewoluuje w stronę standaryzacji. O ile dotychczas ujawnienia opierały się na zasadzie dobrowolności (*soft law*), o tyle wprowadzenie takich regulacji jak SFDR (Sustainable Finance Disclosure Regulation) czy taksonomia UE (rozporządzenie w sprawie ustanowienia ram ułatwiających zrównoważone inwestycje) wymusza na emitentach i funduszach ścisłą spójność raportowania, co ma na celu ograniczenie zjawiska *greenwashing* (por. Dempere, Alamash, Mattos 2024).

2.2. Błąd odwzorowania

Efektywność funduszu ETF rozpatrywana jest w dwóch wymiarach: cenowym oraz replikacyjnym. Pierwszy dotyczy relacji między ceną giełdową a wartością aktywów netto. Występowanie premii lub dyskonta stwarza okazje do arbitrażu, przy czym – jak wskazują Gaba i Kumar (2021) – mechanizm kreacji i umarzania jednostek niweluje tę nieefektywność znacznie szybciej na rynkach rozwiniętych niż na wschodzących. Kluczowym miernikiem jakości zarządzania pasywnego jest jednak błąd odwzorowania (*tracking error* – TE). Określa on, czy kurs rynkowy ETF-ów jest wyższy czy niższy od wartości benchmarku (odpowiedniego indeksu, a w przypadku ETN i ETC – indeksu lub aktywów bazowych, np. surowca).

Błąd odwzorowania można zapisać następującym równaniem:

$$TE_{i,t} = \frac{\sum_{t=1}^{n} \left| R_{ETF_t} - R_{BM_t} \right|}{n}$$

gdzie:

TE – błąd odwzorowania,

R_{ETF} – stopa zwrotu funduszu ETF,

R_{BM} – stopa zwrotu indeksu (benchmarku).

Badania empiryczne dowodzą, że większość funduszy ETF generuje wyniki niższe od benchmarku, co wynika z kosztów operacyjnych i barier transakcyjnych (Gaba, Kumar 2021). Należy zauważyć, że inwestor nie ma możliwości bezpośredniego nabycia indeksu, może jedynie kupić replikujące go instrumenty. Wykorzystanie rynku terminowego (kontrakty *futures*) wiąże się natomiast z ryzykiem bazy, kosztami rolowania pozycji oraz wymogami depozytowymi, które często są nieakceptowalne dla inwestorów indywidualnych (por. Miziołek, Roman 2026).

Zgodnie z analizą Gaby i Kumara (2021) na poziom błędu odwzorowania oddziałuje wiele determinant, wśród których kluczowe znaczenie mają: struktura kosztów, wolumen obrotu, zmienność funduszu oraz wartość zgromadzonych aktywów, która wpływa wprost proporcjonalnie na efektywność replikacji. Istotną rolę odgrywa również ryzyko, kwantyfikowane za pośrednictwem odchylenia standardowego stóp zwrotu i wykazujące dodatnią korelację z TE, a także wiek funduszu, przy czym dłuższy staż rynkowy sprzyja ograniczaniu błędu odwzorowania. Katalog ten uzupełniają czynniki operacyjne i rynkowe, takie jak wahania kursów walutowych, występowanie premii lub dyskonta względem NAV, poziom rezerw gotówkowych, wykorzystanie optymalizacji podatkowej, szerokość spreadu w arkuszu zleceń oraz koszty związane z obsługą kontraktów swap.

3. ETF-y ESG – przegląd literatury

3.1. Mechanizmy i determinanty błędu odwzorowania w funduszach ETF

W literaturze przedmiotu fundusze ETF (*exchange-traded funds*) są definiowane jako instrumenty zapewniające płynną i efektywną kosztowo ekspozycję na indeksy rynkowe (Ben-David i in. 2023). Mimo wysokiej efektywności mechanizmów arbitrażowych fundusze te generują odchylenia od benchmarków, określane jako błąd odwzorowania (TE). Ben-David i in. (2023) upatrują przyczyn tego zjawiska w kosztach transakcyjnych, przyjętych strategiach replikacji oraz ograniczeniach płynności aktywów bazowych. Omawiając specyfikę funduszy zrównoważonych, Climent i Soriano (2011) oraz Plagge i Grim (2020) wskazują na niejednoznaczny związek między scoringiem ESG a wynikami finansowymi i sugerują, że integracja kryteriów niefinansowych może modyfikować profil ryzyka portfela. Kluczową determinantą TE okazuje się również skala działalności. Daniluk (2025) dowodzi, że wartość aktywów pod zarządzaniem jest istotnie skorelowana z dokładnością replikacji funduszy ETF ESG, podczas gdy wiek funduszu pozostaje czynnikiem statystycznie nieistotnym. W świetle tych ustaleń nadrzędne znaczenie dla minimalizacji błędu odwzorowania przypisuje się rygorystycznej kontroli kosztów operacyjnych oraz aktywnemu zarządzaniu ryzykiem (Daniluk 2025).

3.2. Efektywność i błąd odwzorowania funduszy ETF ESG

Fundusze ETF ESG podlegają podwójnej wycenie: rynkowej (giełdowej) oraz niefinansowej, opartej na ratingach zrównoważonego rozwoju. Jak zauważają Janicka i Miziołek (2022), dominujący nurt literatury koncentruje się na środowiskowym aspekcie ESG, co uzasadnia konieczność pogłębionej analizy pozostałych komponentów. Okres 2020–2024, charakteryzujący się destabilizacją gospodarczą wywołaną pandemią COVID-19 oraz konfliktem zbrojnym w Ukrainie, przyniósł nowe wyzwania w zakresie raportowania czynników niefinansowych (Adams, Abhayawansa 2022). Wstrząsy te wymusiły redefinicję paradygmatów funkcjonowania rynków oraz modyfikację modeli finansowych (Altig i in. 2020).

Empiryczne analizy Rompotisa (2022) oraz Nguyena (2023) wskazują, że w czasie kryzysu pandemicznego efektywność replikacji funduszy ESG ulegała istotnym wahaniom, co wiązało się ze specyficzną zmiennością tych instrumentów (Kaminskyi, Baiura, Nehrey 2022). Whelan, Atz i Clark

(2021) podkreślają, że wyniki ETF w odniesieniu do zrównoważonego rozwoju są silnie determinowane przez warunki makroekonomiczne. Gebauer, Kreuzer i Schmidhammer (2026) wykazali natomiast, że w okresach podwyższonej niepewności geopolitycznej fundusze o profilu ESG wykazują tendencję do osiągania niższych stóp zwrotu względem tradycyjnych benchmarków, co wynika ze zwiększonej ekspozycji na ryzyko systematyczne. Kontrastuje to z wynikami badań Supatgiata i in. (2025), którzy sugerują, że w okresie 2018–2023 wysokie ratingi ESG pełniły funkcję stabilizującą, ograniczając wrażliwość na ekstremalne szoki rynkowe. Z kolei analiza Rompotisa (2024) dotycząca rynku w Hongkongu w latach 2022–2024 wskazuje, że fundusze zrównoważone generowały wyniki niższe od indeksu Hang Seng, a jednocześnie cechowały się wyższymi kosztami zarządzania i podwyższonym profilem ryzyka.

3.3. Nowoczesna teoria portfelowa i typologie finansów zrównoważonych

Wykorzystanie czynników ESG w procesach inwestycyjnych prowadzi do powstania strukturalnych różnic między indeksami zrównoważonymi a tradycyjnymi, głównie w zakresie selekcji i wag komponentów. Pastor, Stambaugh i Taylor (2023) argumentują, że rosnący popyt na „zielone" aktywa generuje premię cenową, co wpływa na wyceny rynkowe funduszy ETF ESG. Równolegle Bolton i Kacperczyk (2023) wskazują na istotny wpływ ryzyka klimatycznego na wycenę instrumentów finansowych – różnice w emisji CO_2 przekładają się na dyskonto lub premię z tytułu transformacji energetycznej.

W warunkach rynkowej efektywności Lindsey (2011) postuluje zachowanie koncepcyjnej równowagi między składowymi ESG, przestrzegając przed redukcjonistycznym podejściem koncentrującym się wyłącznie na ekologii. Aktywność inwestorów, mierzona m.in. zainteresowaniem w sieci (Google Trends), stanowi istotny predyktor stóp zwrotu funduszy ESG (Zong, Zhu, Huang 2024). Olofsson i Carlsson (2023) wskazują jednak na brak spójnych korelacji między scoringiem ESG a wynikami inwestycyjnymi funduszy zaliczanych do różnych klas aktywów, co sugeruje, że efektywność inwestycyjna pozostaje pochodną specyfiki indeksu bazowego.

Współczesna dynamika rynków, napędzana postępem technologicznym, wymusza stosowanie zaawansowanych metod ilościowych. Badania Cho i in. (2023) wykazały, że algorytmy uczenia maszynowego pozwalają na dokładniejsze przewidywanie błędów odwzorowania funduszy ETF niż tradycyjne modele liniowe oraz lepszą identyfikację kluczowych czynników, takich jak udział aktywów amerykańskich i wskaźnik kosztów. Ajay i in. (2026) dowodzą z kolei, że modele uczenia maszynowego mają przewagę nad klasycznymi modelami statystycznymi w prognozowaniu wyników funduszy ETF, szczególnie w sektorach o wysokiej zmienności, takich jak infrastruktura i bankowość. W sektorach bardziej stabilnych, m.in. farmaceutycznym czy w opiece zdrowotnej, tradycyjne modele nadal uzyskują wysoką trafność prognoz. Wyniki te wskazują na znaczny potencjał wykorzystania technik uczenia maszynowego do prognozowania cen ETF oraz kształtowania strategii inwestycyjnych.

Uwzględnienie czynników zrównoważonego rozwoju znajduje odzwierciedlenie w ewolucji modeli wyceny. Zerbib (2020) zaproponował model S-CAPM (Sustainable CAPM), który modyfikuje klasyczne podejście Sharpe'a (1964) przez uwzględnienie podziału inwestorów na grupy o różnych preferencjach etycznych (grupy „obojętne" i „zrównoważone") oraz podział aktywów na dostępne i wykluczone. Zgodnie z tą teorią aktywa wykluczane przez inwestorów zrównoważonych muszą oferować wyższą oczekiwaną stopę zwrotu jako rekompensatę za ryzyko wykluczenia, co bezpośrednio wpływa na strukturę portfeli ETF.

Procesy te systematyzuje klasyfikacja Schoenmakera (2017), która ilustruje ewolucję od prymatu wartości dla udziałowców (Finanse 1.0) do paradygmatu wspólnego dobra (Finanse 3.0). Dalszy rozwój technologii prowadzi do koncepcji Finansów 4.0 (cyfryzacja i Big Data) oraz Finansów 5.0. Te ostatnie definiuje się jako ekosystem skoncentrowany na człowieku i etyce, wykorzystujący sztuczną inteligencję i technologie kwantowe w służbie odpowiedzialnego rozwoju (Gulati, Singla, Saini 2025).

4. Metodyka badania własnego

Materiał empiryczny wykorzystany w niniejszym opracowaniu został pozyskany z bazy danych Reuters Workspace. Podstawę obliczeń stanowią dzienne logarytmiczne stopy zwrotu, wyznaczone na podstawie giełdowych kursów zamknięcia. W celu zachowania rzetelności analizy zastosowano podejście *total return*, uwzględniające wypłacane dywidendy. Do próby badawczej wybrano dziewięć funduszy ETF, przy zachowaniu parytetu trzech instrumentów dla każdej subkategorii ESG (Environmental, Social, Governance). Fundusze te reprezentują poszczególne filary odpowiedzialnego inwestowania. Filar E obejmuje ETF-y skupiające się na redukcji śladu węglowego, zgodnie z Climate Transition Benchmark (CTB). Filar S to ETF-y eliminujące spółki naruszające zasady UN Global Compact; koncentruje się na odpowiedzialności społecznej. Filar G to ETF-y wybierające spółki z wysokimi wynikami ESG, kładące nacisk na ład korporacyjny. Szczegółowa charakterystyka techniczna wybranych funduszy została przedstawiona w tabeli 1.

Należy podkreślić, że na dobór jednostek do próby badawczej wpłynęły istotne ograniczenia rynkowe. Mimo dynamicznego wzrostu sektora inwestycji zrównoważonych podaż instrumentów o precyzyjnie zdefiniowanym i jednorodnym profilu tematycznym dla każdego z trzech filarów pozostaje relatywnie niska. Znaczna część dostępnych produktów finansowych charakteryzuje się podejściem zintegrowanym (mieszanym), co utrudnia wyodrębnienie wpływu pojedynczych komponentów ESG na wynik inwestycyjny oraz błąd odwzorowania.

W związku z powyższym dla każdego segmentu przeanalizowano po trzy reprezentatywne fundusze. Przyjęte podejście metodologiczne znajduje bezpośrednie uzasadnienie w literaturze przedmiotu – analogiczną strukturę badania (opartą na analizie trzech funduszy typu *core* oraz trzech odpowiadających im kategorii ESG) zastosował Noman (2023), badając premie oraz błędy odwzorowania w grupie funduszy iShares. Taka selekcja pozwala na zachowanie wysokiej precyzji analitycznej przy jednoczesnym uwzględnieniu faktycznych barier w dostępie do wyspecjalizowanych instrumentów finansowych.

W celu eliminacji ryzyka walutowego i zapewnienia pełnej porównywalności z benchmarkami szeregi czasowe funduszy notowanych we Frankfurcie przeliczono z euro na dolary amerykańskie, co umożliwiło ujednolicone wyznaczenie błędu odwzorowania na podstawie dziennych logarytmicznych stóp zwrotu w ujęciu *total return*.

Wskaźnik TE obliczono na podstawie cen rynkowych oraz porównano z alternatywnym podejściem – TE obliczanym na podstawie wycen NAV. Wartość TE obliczono na podstawie dziennych stóp zwrotu ETF oraz odpowiadających im benchmarków.

W celu oceny dokładności replikacji zastosowano dwie komplementarne miary: (1) błąd odwzorowania (TE), który wyznacza się jako odchylenie standardowe różnic stóp zwrotu funduszu i benchmarku, oraz (2) różnicę odwzorowania (TD) – wprowadzoną jako miara skali niedopasowania,

rozumianego jako prosta różnica między stopą zwrotu funduszu a stopą zwrotu benchmarku w danym podokresie. TD mierzy średnią różnicę pomiędzy stopą zwrotu funduszu ETF a stopą zwrotu benchmarku. W przypadku idealnego odwzorowania indeksu wartość tej miary powinna być bliska zera.

$$TD = \frac{1}{n}\sum_{t=1}^{n}\left(R_{ETF_t} - R_{BM_t}\right)$$

gdzie:

TD – różnica odwzorowania,
R_{ETF_t} – stopa zwrotu ETF w okresie t,
R_{BM_t} – stopa zwrotu benchmarku w okresie t,
n – liczba obserwacji.

Dodatkowo obliczono kroczący błąd odwzorowania (*rolling tracking error*) w oknach kroczących, co pozwala ocenić stabilność jakości odwzorowania benchmarków w czasie.

$$TE_t^{(k)} = \sqrt{\frac{1}{k-1}\sum_{i=t-k+1}^{t}\left(R_{ETF_t} - R_{BM_t}\right)^2}$$

gdzie $TE_t^{(k)}$ to błąd odwzorowania w czasie t, liczony dla okna o długości k.

W celu analizy zachowania funduszy w okresach spadków rynku zastosowano *downside tracking error* (DTE). Uwzględnia on jedynie obserwacje, dla których stopa zwrotu benchmarku była ujemna.

$$DTE \quad \sqrt{\frac{1}{n}\sum_{t=1}^{n}\min\left(0, R_{ETF_t} \quad R_{BM_t}\right)^2}$$

Do oceny stopnia dopasowania modelu użyto współczynnika determinacji R^2 oraz błędu dopasowania RMSE.

Aby zwiększyć porównywalność wyników, wszystkie miary TE przedstawiono również w ujęciu rocznym, przez zastosowanie procedury annualizacji opartej na liczbie dni sesyjnych w roku.

W sekcji dotyczącej metodyki rozszerzono aparat statystyczny o test analizy wariancji (ANOVA) w celu zbadania istotności różnic między grupami oraz test Jarquego-Bery służący do weryfikacji normalności rozkładów stóp zwrotu.

Dane dzienne zostały przetworzone statystycznie i zwizualizowane z wykorzystaniem RStudio. Badanie obejmuje okres od 1 stycznia 2020 r. do 31 grudnia 2024 r.

5. Błąd odwzorowania indeksów ETF ESG – wyniki analizy empirycznej

Na pierwszym etapie oceniono jakość odwzorowania indeksów referencyjnych przez fundusze ETF ESG przy użyciu standardowych miar: błędu odwzorowania (TE), różnicy odwzorowania (TD), alternatywnej miary opartej na średnim bezwzględnym odchyleniu (TE_MAD), kroczącego błędu odwzorowania (*rolling TE*), a także *downside tracking error*, uwzględniającego wyłącznie okresy spadków benchmarku (*downside TE*). Dodatkowo oszacowano model regresyjny opisujący zależność między stopami zwrotu

ETF oraz benchmarków i uzyskano wartości parametrów alfa i beta, współczynnik determinacji (R^2) oraz błąd dopasowania RMSE. Zbiorcze wyniki prezentuje tabela 2.

Analiza miar opartych na cenach rynkowych (TE_PRICE, R^2, RMSE) wskazuje, że najwyższą precyzją odwzorowania cechuje się kategoria Environmental, natomiast najniższą – Social. Grupa Governance zajmuje pozycję środkową.

Porównanie błędu odwzorowania w ujęciu cen rynkowych i NAV

Istotnym wynikiem analizy jest wyższy poziom TE opartego na wartości aktywów netto w stosunku do TE liczonego według cen rynkowych. Relacja ta jest nietypowa, gdyż w literaturze często wskazuje się NAV jako bardziej bezpośrednie odzwierciedlenie wartości portfela (por. Frino, Gallagher 2001; Ackert, Tian 2008; Rompotis 2009; Blitz, Huij 2012). Większe odchylenie NAV od benchmarku, szczególnie widoczne w grupie Governance, może wynikać z opóźnień w wycenie składników indeksu, ich ograniczonej płynności lub specyficznej konstrukcji benchmarków.

Stabilność odwzorowania w czasie zbadano za pomocą kroczącego błędu odwzorowania z oknem 60-dniowym (wykres 1).

Najwyższą stabilnością cechuje się grupa E (niski poziom i brak skoków RTE). Fundusze z grupy G wykazują umiarkowaną zmienność, natomiast te z kategorii S odznaczają się najwyższym poziomem i epizodyczną zmiennością RTE. Przyczyną tych fluktuacji mogą być problemy z płynnością składników, trudności w rebalancingu lub wysoka koncentracja indeksów.

Różnica odwzorowania

Wyniki wskazują, że TD dla cen rynkowych oscyluje wokół zera, natomiast dla NAV przyjmuje wartości nieznacznie ujemne, szczególnie w kategorii Governance. Potwierdza to, że średnie stopy zwrotu ETF są niższe od indeksów referencyjnych, co jest spójne z badaniami wskazującymi na koszty zarządzania, rebalancingu i transakcji jako źródła systematycznego odchylenia (Frino, Gallagher 2001; Elton i in. 2002; Blitz, Huij 2012; Johnson i in. 2013).

Analiza regresji

Oszacowane parametry alfa bliskie zera oraz współczynniki beta zbliżone do jedności potwierdzają pasywny charakter funduszy. Fundusze środowiskowe (E) wykazują silniejszą reakcję na zmiany benchmarku (beta > 1), podczas gdy grupy S i G reagują słabiej. Najwyższy współczynnik determinacji (R^2) odnotowano dla kategorii E (0,8143), co w powiązaniu z najniższym błędem RMSE (0,0063) oraz najwyższą korelacją (0,8929) potwierdza, że ta grupa najwierniej odwzorowuje benchmarki. W grupach S i G poziom niewyjaśnionej zmienności jest istotnie wyższy (R^2 odpowiednio: 0,5308 i 0,4939).

Analiza ryzyka w okresach spadkowych

Dodatkowym elementem badania była ocena jakości odwzorowania benchmarków w okresach spadków rynku. Analiza *downside tracking error* wykazała, że w czasie spadków rynkowych odchylenia funduszy od indeksów rosną. Najniższe wartości DTE uzyskała kategoria E, a najwyższe kategoria S, co sugeruje pogorszenie jakości replikacji przy podwyższonej zmienności.

Alternatywne miary błędu odwzorowania

W celu zwiększenia odporności analizy na obserwacje skrajne obliczono również alternatywną miarę TE opartą na średnim bezwzględnym odchyleniu stóp zwrotu. Uzyskane wyniki wskazują, że ETF-y z kategorii G charakteryzują się najwyższym przeciętnym błędem odwzorowania.

Zróżnicowanie profilu odwzorowania według filarów ESG – ujęcie syntetyczne

Agregacja wyników w podziale na kategorie E, S i G pozwala na sformułowanie wniosków dotyczących specyfiki replikacji każdej z grup. Z przeprowadzonej analizy (tabela 2, wykres 2) wynika, że kategoria E charakteryzuje się najwyższą precyzją i stabilnością odwzorowania. Potwierdzają to nie tylko najniższe wartości TE i RMSE, ale także najwyższa korelacja (0,89) oraz najwyższy stopień wyjaśnienia zmienności przez benchmark (R^2 = 0,8143).

W przeciwieństwie do nich kategoria S odznacza się największym rozproszeniem jakości replikacji oraz najsłabszymi parametrami w warunkach dekoniunktury (najwyższy *downside TE*). Z kolei w kategorii G stwierdzono systematycznie najwyższe błędy odwzorowania w ujęciu wartości aktywów netto oraz najwyższy błąd bezwzględny (TE_MAD). Fundusze te są również najbardziej podatne na negatywne szoki rynkowe, co potwierdzają parametry rozkładu: najniższa (najbardziej ujemna) skośność oraz najwyższa kurtoza. Zbiorcze wyniki prezentuje tabela 3.

Interesującym dopełnieniem obrazu efektywności jest analiza wskaźnika *mean premium* (por. tabela 2). Ceny rynkowe funduszy E utrzymują się przeciętnie powyżej wartości NAV (premia 0,3899), co może dowodzić premii za „zielony" charakter aktywów, tymczasem fundusze G wykazują tendencję do handlu z dyskontem (-0,2341).

Test Jarquego-Bery (p-value = 0) jednoznacznie wskazuje na brak normalności rozkładów stóp zwrotu we wszystkich grupach, przy czym to kategoria G charakteryzuje się najbardziej „grubymi ogonami". W połączeniu z niskim dopasowaniem regresji (R^2 = 0,49) sugeruje to, że na wyniki tej grupy istotnie wpływają czynniki specyficzne, niezwiązane z samym benchmarkiem. Brak normalności oraz występowanie obserwacji skrajnych zobrazowano histogramami rozkładów stóp zwrotu wybranych funduszy (wykres 4).

W celu statystycznego potwierdzenia zaobserwowanych różnic między grupami E, S i G przeprowadzono analizę wariancji (ANOVA). Wyniki zestawiono w tabeli 4.

Przeprowadzony test nie wykazał statystycznie istotnych różnic między grupami funduszy ESG w zakresie analizowanych miar ($p > 0,05$). Oznacza to brak podstaw do odrzucenia hipotezy zerowej o równości średnich. Należy jednak zaznaczyć, że mimo braku istotności statystycznej (co może wynikać z relatywnie małej liczebności prób oraz wysokiej zmienności wewnątrzgrupowej) odnotowano wyraźne różnice ekonomiczne między poziomami TD i profilami ryzyka. Wyniki sugerują, że choć klasyfikacja ESG nie jest jedyną determinantą precyzji replikacji, jednak wpływa na strukturę odchyleń i charakterystykę statystyczną stóp zwrotu funduszy.

6. Omówienie i interpretacja wyników

Analiza dowodzi, że fundusze z kategorii Environmental wykazują najwyższą precyzję i spójność replikacji, co może wynikać z większej przejrzystości indeksów oraz płynności ich składników. Najniższe wartości błędu odwzorowania oraz najwyższy współczynnik determinacji w tej grupie sugerują wysoką standaryzację indeksów środowiskowych. Można to wiązać z obserwacjami Supatgiata i in. (2025), którzy sugerują, że wysokie ratingi ESG (szczególnie w kategorii E) pełnią funkcję stabilizującą, ograniczając wrażliwość na szoki rynkowe. Ich wysoka precyzja może również wynikać z opisywanej przez Pastora, Stambaugha i Taylora (2023) premii cenowej za aktywa „zielone", która stymuluje płynność i zainteresowanie inwestorów, ułatwiając zarządzającym dokładne naśladowanie benchmarku. Fundusze z kategorii Governance cechują się systematycznie wyższym błędem odwzorowania i słabszym dopasowaniem modelu regresji, natomiast kategoria Social cechuje się największym zróżnicowaniem wyników wewnątrz grupy. Istotnym i nietypowym rezultatem badania jest stwierdzenie wyższego poziomu błędu odwzorowania opartego na NAV w stosunku do cen rynkowych, co jest szczególnie widoczne w grupie Governance. Wynik ten jest odmienny niż w klasycznych pracach, takich jak Frina i Gallaghera (2001) czy Rompotisa (2009), które zazwyczaj wskazują na NAV jako bardziej bezpośrednie i precyzyjne odzwierciedlenie wartości portfela. Przyczyną tej rozbieżności, zgodnie z sugestiami Daniluk (2025), może być kluczowa rola skali działalności (AuM) dla precyzji replikacji funduszy ESG. Niższa płynność niektórych składników portfeli może prowadzić do opóźnień w wycenie NAV, co w sytuacji dynamicznych zmian rynkowych zwiększa błąd odwzorowania względem benchmarku. Jednocześnie brak statystycznie istotnych różnic w teście ANOVA między grupami E, S i G (mimo wyraźnych różnic ekonomicznych) sugeruje, że przynależność do danej grupy ESG nie stanowi jedynej determinanty precyzji, co jest zgodne z wnioskami Olofssona i Carlssona (2023). Uzyskane ujemne wartości TD są spójne z literaturą przedmiotu oraz potwierdzają, że koszty operacyjne i transakcyjne uniemożliwiają pełną replikację benchmarku (Elton i in. 2002; Gastineau 2010; Gaba, Kumar 2021).

Badania empiryczne pozwoliły na odniesienie się do postawionych hipotez badawczych, łącząc surowe dane statystyczne z ich interpretacją ekonomiczną:

H1. Pomimo numerycznych różnic (najniższy TE dla kategorii E) analiza wariancji (ANOVA) wykazała $p > 0{,}05$. Brak podstaw do odrzucenia hipotezy zerowej o równości średnich oznacza, że różnice między grupami nie są istotne statystycznie.

H2 (potwierdzona). Analiza wykazała, że błąd odwzorowania liczony na podstawie NAV jest systematycznie wyższy niż dla cen rynkowych, co dowodzi, że przyjęta baza danych istotnie wpływa na ocenę jakości replikacji.

H3 (potwierdzona). Wyniki regresji wskazują na zróżnicowane dopasowanie: fundusze E osiągają najwyższe R^2 oraz najniższe RMSE (najlepsza jakość replikacji), podczas gdy grupy S i G cechują się słabszymi parametrami modelu.

H4 (potwierdzona). Test Jarquego-Bery ($p = 0$) jednoznacznie odrzucił normalność rozkładów we wszystkich grupach. Uzyskane wartości skośności i kurtozy potwierdzają asymetrię oraz występowanie zjawiska „grubych ogonów", co oznacza, że w analizowanych danych występuje znacznie wyższe prawdopodobieństwo pojawienia się zdarzeń ekstremalnych (dużych spadków lub wzrostów), niż wynikałoby z klasycznego rozkładu Gaussa.

7. Podsumowanie i wnioski

Niniejsze badanie dowodzi, że fundusze ETF oparte na indeksach ESG stanowią efektywne narzędzie inwestowania pasywnego. Zastosowanie szerokiego zestawu miar klasycznych i alternatywnych pozwoliło wykazać, że analizowane podmioty skutecznie odwzorowują zachowanie benchmarków, a błędy odwzorowania kształtują się na poziomach typowych dla tradycyjnych funduszy indeksowych. Odnotowane niewielkie, systematycznie ujemne wartości TD są spójne z wynikami opisanymi w literaturze przedmiotu i wynikają z kosztów operacyjnych oraz transakcyjnych. Analiza regresyjna potwierdziła wysoki stopień wyjaśniania zmienności stóp zwrotu ETF przez indeksy referencyjne, choć precyzja ta różni się w zależności od kategorii ESG, co może być pochodną zróżnicowanej płynności składników portfela lub stosowania strategii replikacji niepełnej (*sampling*).

Wkład artykułu do literatury obejmuje kompleksową ocenę trzech filarów ESG przy użyciu rozszerzonego instrumentarium badawczego (m.in. *downside tracking error* oraz miary odporne na obserwacje skrajne). Wyniki sugerują, że mimo dodatkowych ograniczeń inwestycyjnych fundusze te zachowują wysoką jakość replikacji, co w kontekście rosnącej świadomości inwestorów i postępu technologicznego (AI) pozwala prognozować dalszą marginalizację błędów odwzorowania w przyszłości.

Należy jednak wskazać na istotne ograniczenia badania, takie jak pięcioletni horyzont czasowy obejmujący okresy podwyższonej zmienności rynkowej, a także trudności z precyzyjnym przypisaniem niektórych indeksów do konkretnych filarów ESG. Potrzebne są zatem dalsze analizy, które mogłyby objąć szerszą próbę funduszy akumulujących dywidendy lub badać wpływ częstotliwości zmian składu indeksów (wynikających z rewizji scoringów ESG) na ich wyniki. Ponadto, biorąc pod uwagę niską wiarygodność obecnych oznaczeń „zielonych" funduszy w ramach rozporządzenia SFDR (Trotta i in. 2026), w przyszłych badaniach należy uwzględniać ewolucję standardów raportowania, która może istotnie wpłynąć na stabilność i jakość replikacji strategii zrównoważonego rozwoju.

Bibliografia

Ackert L.F., Tian Y.S. (2008), Arbitrage, liquidity, and the valuation of exchange traded funds, *Financial Markets, Institutions and Instruments*, 17(5), 331–362.

Adams C.A., Abhayawansa S. (2022), Connecting the COVID-19 pandemic, environmental, social and governance (ESG) investing and calls for 'harmonisation' of sustainability reporting, *Critical Perspectives on Accounting*, 82, 102309.

Ajay J., Kavitha R., Ashwin V.S., Dhanya M. (2026), A comparative analysis of ETF performance using machine learning algorithms and traditional models, w: Fong S., Dey N., Joshi A. (red.), *ICT Analysis and Applications. Proceedings of ICT4SD 2025*, Springer.

Altig D. i in. (2020), Economic uncertainty before and during the COVID-19 pandemic, *Journal of Public Economics*, 191(4), 104274.

Amel-Zadeh A., Serafeim G. (2018), Why and how investors use ESG information: evidence from a global survey, *Financial Analysts Journal*, 74(3), 87–103.

Ben-David I., Franzoni F., Kim B., Moussawi R. (2023), Competition for attention in the ETF space, *The Review of Financial Studies,* 36(3), 987–1042.

Blitz D., Huij J. (2012), Evaluating the performance of global emerging markets equity exchange-traded funds, *Emerging Markets Review,* 13, 149–158.

Bolton P., Kacperczyk M. (2023), Global pricing of carbon transition risk, *The Journal of Finance*, 78(6), 3677–3754.

Cho J.E., Lee G.H., Lee W., Kim B. (2023), Machine learning approach for predicting U.S. ETFs' tracking errors – implications on U.S. invested fund, *SSRN Electronic Journal*, 1–26.

Climent F., Soriano P. (2011), Green and good? The investment performance of US environmental mutual funds, *Journal of Business Ethics*, 103(2), 275–287.

Czerwonka M. (2025), *Zrównoważone i społecznie odpowiedzialne inwestowanie*, Oficyna Wydawnicza SGH.

Daniluk K. (2025), Patterns in ETF tracking errors: ESG vs. non-ESG passive equity exchange-traded funds, *Ekonomia i Prawo*, 24(2), 181–199.

Dempere J., Alamash E., Mattos P. (2024), Unveiling the truth: greenwashing in sustainable finance, *Frontiers in Sustainability*, 5, 1362051.

Dziawgo L. (2009), Społeczna odpowiedzialność biznesu jako czynnik konkurencyjny na współczesnym rynku kapitałowym, *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, Rynki Finansowe, Ubezpieczenia*, 16, 269–278.

Elton E., Gruber M., Comer G., Li K. (2002), Spiders: Where are the bugs?, *Journal of Business*, 75(3), 37–59.

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

Gaba A., Kumar R. (2021), Tracking error and pricing efficiency of exchange traded funds: a systematic literature review, *Orissa Journal of Commerce*, 42(3), 42–58.

Gastineau G. (2010), *The Exchange-Traded Funds Manual*, John Wiley & Sons.

Gebauer L., Kreuzer C., Schmidhammer C. (2026), Sustainability in calm and rough waters: an empirical investigation of European ESG ETFs, *Journal of Asset Management*, 27(4).

Gulati T., Singla A., Saini P. (2025), Sustainable digital finance and Finance 5.0: a systematic review and research agenda, *South Asian Journal of Business Studies*, 15(3), 283–307, DOI: 10.1108/SAJBS-03-2025-0131.

Janicka M., Miziołek T. (2022), *Finanse zrównoważone. ESG – przedsiębiorstwa – sektor finansowy*, Polskie Wydawnictwo Ekonomiczne.

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.

Kaminskyi A., Baiura D., Nehrey M. (2022), ESG investing strategy through COVID-19 turmoil: ETF-based comparative analysis of risk-return correspondence, *Intellectual Economics*, 16(2), 95–120.

Lindsey T.C. (2011), Sustainable principles: common values for achieving sustainability, *Journal of Cleaner Production*, 19(5), 561–565.

Miziołek T. (2023), Active management in Polish domestic Treasury bond funds, *Annales Universitatis Mariae Curie-Skłodowska. Sectio H – Oeconomia*, 57(1), 137–153·

Miziołek T., Roman A. (2026), Pricing efficiency of exchange-traded funds listed on the Warsaw Stock Exchange, *Central European Economic Journal*, 13(60), 37–55.

Nguyen H. (2023), COVID-19: performance of ESG ETFS and ESG ETFS vs their declared indexes, *Applied Finance Letters*, 12, 33–43.

Noman A. (2023), Premium, tracking errors and performance of iShares Core versus ESG S&P ETFs, *Banking and Finance Review*, 14(1), 25–37.

Olofsson E., Carlsson J. (2023), *ESG and Fund Performance. Comparing Funds with Different Strategic Benchmarks*, University of Gothenburg.

Pastor L., Stambaugh R.F., Taylor L.A. (2023), ESG returns face slowdown. Climate shocks drove the outperformance of green stocks but that is changing, *The Financial Times*, https://www.ft.com/content/f3d9f74e-df3d-4ec5-b3ae-04746c4bdde7.

Plagge J.C., Grim D.M. (2020), Have investors paid a performance price? Examining the behavior of ESG equity funds, *The Journal of Portfolio Management*, 46(3), 123–140.

Pyka I., Nocoń A. (2026), Adaptacja strategii ESG w bankach komercyjnych w Polsce, *Bank i Kredyt*, 57(3), 263–286, DOI: 10.5604/01.3001.0055.7777.

Rompotis G.G. (2009), Active vs. passive management: new evidence from exchange traded funds, *SSRN Electronic Journal*, 1–18.

Rompotis G. (2022), Environmental, Social and Governance Responsibility, financial performance and assets: a study of exchange traded fund, *Research Papers in Economics and Finance*, 6(2), 23–49.

Rompotis G.G. (2024), The performance of ESG ETFs vis-à-vis non-ETFs in Hong Kong, *Climate Economics and Finance*, 2(1), 12–23.

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

Schoenmaker D. (2017), *Investing for the Common Good. A Sustainable Finance Framework,* Bruegel.

Schoenmaker D., Schramade W. (2018), *Principles of Sustainable Finance*, Oxford University Press.

Supatgiat C., Phiromswad P., Sahin O., Sarajoti P. (2025), Investigating the impact of ESG ratings on ETF performance during market disruptions: evidence from the COVID-19 pandemic and Russian (full-scale) invasion of Ukraine, *Research in International Business and Finance*, 77, Part A.

Trotta A., Piluso F., Strano E., Ceraso D. (2026), Exploring the climate resilience of ESG and conventional ETFs: evidence from the European region, *Research in International Business and Finance*, 84, 103313.

Whelan T., Atz U., Clark C. (2021), *ESG and Financial Performance. Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020*, NYU Stern.

Zerbib O.D. (2020), A Sustainable Capital Asset Pricing Model (S-CAPM): evidence from green investing and sin stock exclusion, *SSRN Electronic Journal*, 1–101.

Zong S., Zhu D., Huang T. (2024), Does investor attention impact ESG ETF performance?, *The Journal of Investing*, 33(5), 87–97.

Aneks

Tabela 1
ETF-y wykorzystane do badania wpływu czynnika ESG na wynik funduszu ETF oraz jego błąd odwzorowania indeksu

Filar ESG	Nazwa ETF	ISIN
(E) Environmental	iShares MSCI EMU ESG Enhanced CTB UCITS ETF EUR (Dist)	IE00BHZPHZ28
	iShares Global Clean Energy	US4642882249
	Invesco Water Resources	US46137V1420
(S) Social	UBS MSCI World Socially Responsible UCITS ETF USD dis	LU0629459743
	Impact Shares NAACP Minority Empwrmt ETF	US45259A2096
	SPDR MSCI USA Gender Diversity	US78468R7474
(G) Governance	Amundi MSCI USA ESG Leaders Extra UCITS ETF DR – USD (D)	IE0008TKP6O7
	WisdomTree EM ex-State-Owned Enterprises	US97717X5784
	Freedom 100 Emerging Markets	US02072L6074

Źródło: opracowanie własne.

Tabela 2
Wyniki zbiorcze

Miary	E	S	G
TE_PRICE_ANNUAL	0,102	0,1839	0,1713
TE_NAV_ANNUAL	0,2993	0,3343	0,5965
TD_PRICE	-0,0001	-0,0001	0
TD_NAV	-0,0004	-0,0001	-0,0007
TE_MAD_PRICE_ANNUAL	0,0726	0,0747	0,1024
DOWNSIDE_TE_PRICE_ANNUAL	0,1029	0,195	0,1772
ALPHA_ANNUAL	-0,0416	0,0228	0,0059
BETA	1,1081	0,6323	0,6977
R^2_PRICE	0,8143	0,5308	0,4939
RMSE_PRICE	0,0063	0,0076	0,0092
CORRELATION_PRICE	0,8929	0,6561	0,6833
MEAN_PREMIUM	0,3899	0,0048	-0,2341

Źródło: opracowanie własne.

Tabela 3
Statystyki opisowe

Miary	E	S	G
Średnia (MEAN)	0,0002	0,0004	0,0002
Zmienność (SD)	0,0167	0,0127	0,0136
Skośność	-0,4788	-0,6306	-0,9314
Kurtoza	8,3602	9,1699	10,3054
Test Jarquego-Bery (JB_STAT)	11 391,4633	13 783,4671	17 643,4296
JB_PVALUE	0	0	0

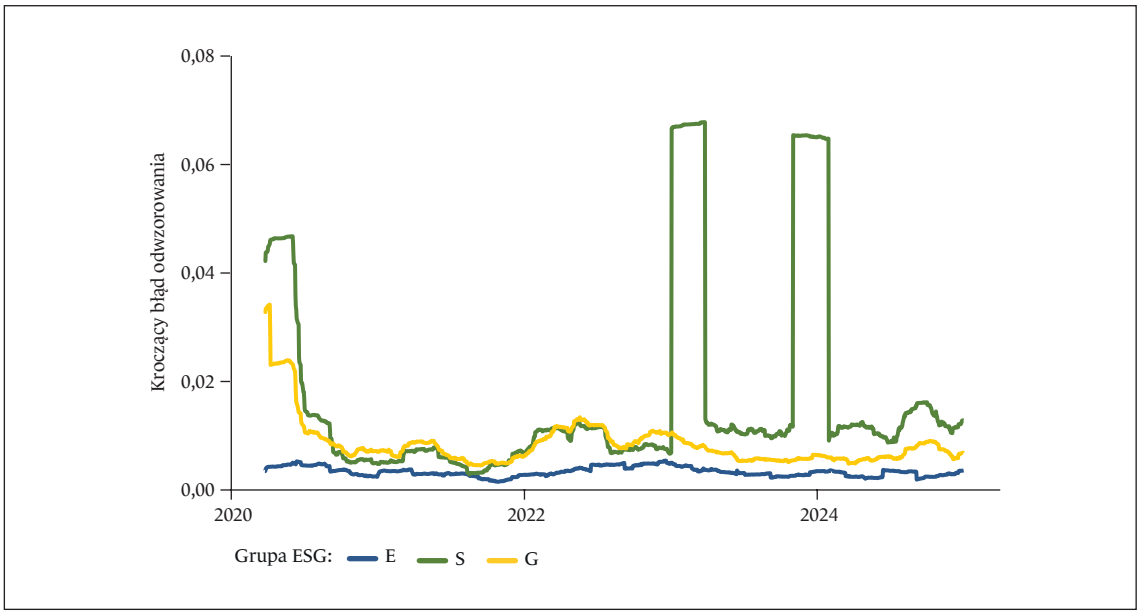
Źródło: opracowanie własne.

Tabela 4
Wyniki analizy wariancji (ANOVA) dla miar efektywności i ryzyka

Miary	F_STAT	P_VALUE	Istotność
TE_PRICE_ANNUAL	0,3544	0,7153	nieistotne statystycznie
TE_NAV_ANNUAL	0,7344	0,5184	nieistotne statystycznie
TD_PRICE	0,5203	0,6189	nieistotne statystycznie
TD_NAV	0,4631	0,6501	nieistotne statystycznie
ALPHA_ANNUAL	1,065	0,402	nieistotne statystycznie
BETA	1,6268	0,2726	nieistotne statystycznie

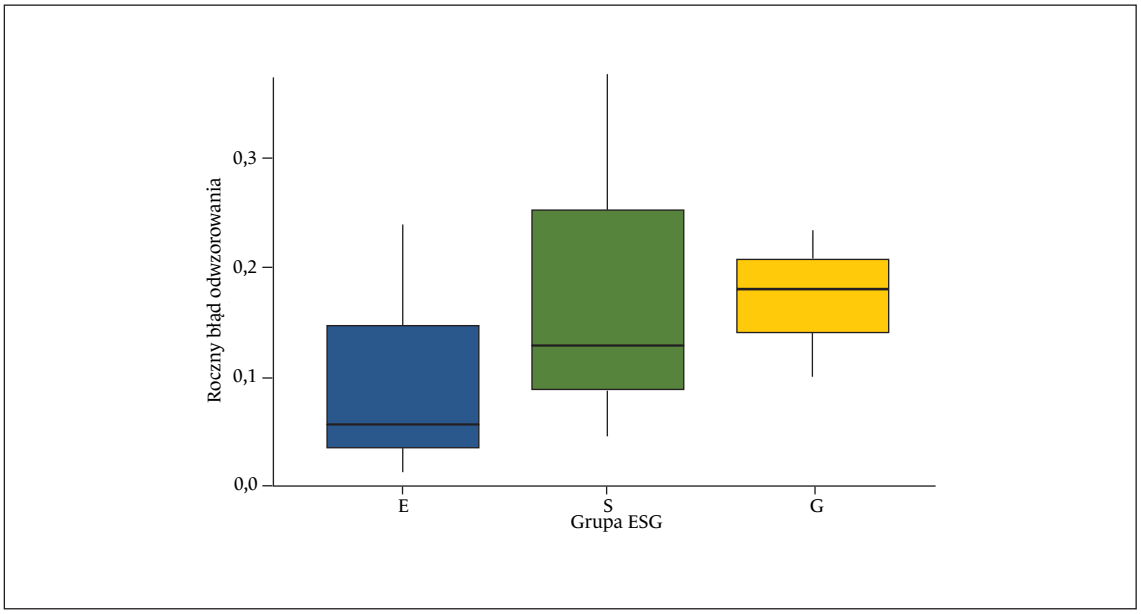
Źródło: opracowanie własne.

Wykres 1
Kroczący błąd odwzorowania z oknami 60-dniowymi



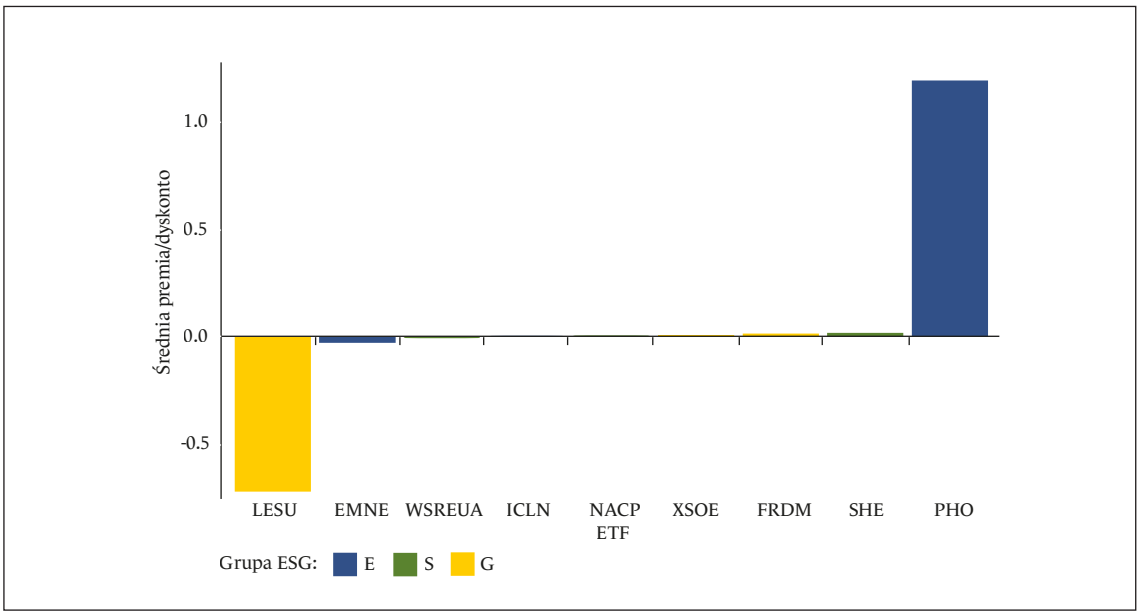
Źródło: opracowanie własne.

Wykres 2
Rozkład błędu odwzorowania w poszczególnych grupach ESG



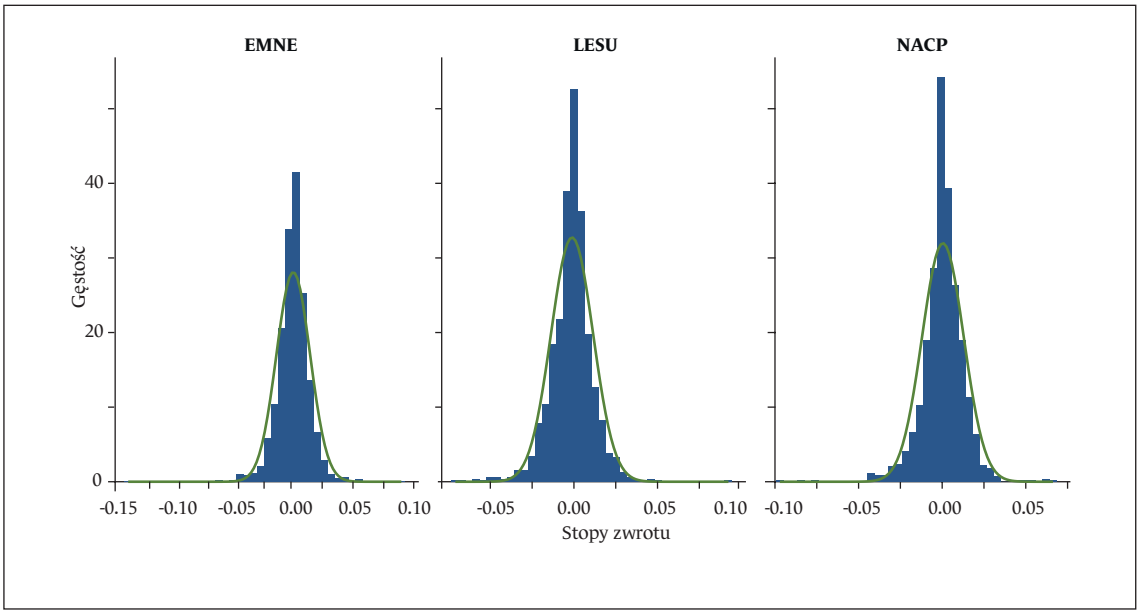
Źródło: opracowanie własne.

Wykres 3
Wykres średniej premii/dyskonta ETF względem NAV



Źródło: opracowanie własne.

Wykres 4
Histogramy – rozkłady stóp zwrotu na przykładzie funduszy EMNE, LESU i NACP



Źródło: opracowanie własne.

Comparative analysis of ETF funds' tracking error by pillars E, S and G

Abstract

This paper investigates the replication quality and risk characteristics of Environmental, Social, and Governance (ESG) exchange-traded funds (ETFs) by examining their ability to track benchmark indices. The analysis is conducted on a sample of nine ESG ETFs grouped into three categories (E, S, G), using daily data on market prices, net asset values (NAV), and corresponding benchmark indices.

The empirical framework combines multiple complementary measures of tracking performance and risk. Tracking accuracy is evaluated using both tracking error (TE), defined as the standard deviation of return differences, and tracking difference (TD), defined as the mean return deviation. In addition, alternative measures are employed, including mean absolute deviation-based tracking error (TE-MAD), rolling tracking error (60-day windows), and downside tracking error conditioned on negative benchmark returns. Regression-based metrics (alpha, beta, R^2, and RMSE) are used to assess the statistical relationship between ETF returns and benchmark returns. Furthermore, distributional properties of returns are examined using skewness, kurtosis, and the Jarque-Bera test.

The results reveal significant heterogeneity across ESG categories. Environmental ETFs exhibit the lowest tracking error and the highest consistency in benchmark replication, despite relatively higher return volatility. Their return distributions are closest to normality, with comparatively lower skewness and kurtosis. In contrast, Governance ETFs demonstrate systematically higher tracking errors and weaker regression fit (lower R^2, higher RMSE), indicating less precise replication, accompanied by pronounced negative skewness, and the highest kurtosis, suggesting elevated exposure to extreme downside risk. Social ETFs present the greatest dispersion of results, with moderate tracking error, but substantial variability across funds and evidence of structural mismatches between ETF returns and benchmark dynamics.

Across all groups, tracking difference is generally negative, indicating that ETFs tend to underperform their benchmarks on average, consistent with the presence of management fees, transaction costs, and portfolio rebalancing frictions documented in the literature. Importantly, all return distributions significantly deviate from normality, as confirmed by the Jarque-Bera test, with strong evidence of leptokurtosis and left-skewness. These findings justify the use of non-parametric statistical methods and highlight the importance of tail-risk-sensitive measures in evaluating ETF performance.

Rolling tracking error analysis further shows that replication quality is time-varying, with episodic increases in tracking deviations, particularly pronounced in Social ETFs, suggesting sensitivity to changing market conditions and index composition dynamics.

Overall, the study contributes to the growing literature on ESG investing by demonstrating that differences across ESG dimensions are not limited to performance levels, but extend to replication efficiency, stability, and risk structure. The findings indicate that Environmental ETFs provide the most reliable benchmark tracking, while Governance and Social ETFs exhibit higher complexity and risk, particularly in extreme market conditions. These results have important implications for investors, portfolio managers, and researchers, emphasizing the need for a multidimensional approach to ETF evaluation that incorporates both tracking metrics and distributional risk characteristics.

Keywords: tracking error, ETF, ESG, stock market