Are the green ETFs really green?

Rita Laura D'Ecclesia* Giacomo Morelli† Kevyn Stefanelli‡

First Version: Oct 29, 2021;
This Version: May 11, 2022

Abstract

Green ETFs have experienced a large increase in volumes and returns in the last years. However, financial agents ask themselves whether these assets are truly green. The large amount of currently available environmental metrics are widely considered as unreliable and their discrepancies generate confusion. The evaluation of the most capitalized green ETFs according to two of the most popular environmental metrics, namely the E Score and the carbon intensity, shows poor green performances of the sampled funds. Adopting a screening process based on the exclusion of the worst-in-class companies, we build synthetic low-carbon funds choosing among the 2021 holdings of these green funds. The synthetic assets show similar, and sometimes better, financial and environmental outcome than the listed ETFs. However, evidences exhibit significative differences in the companies environmental classification, green or brown, according to the two metrics.

Keywords: Green ETF, Environmental Metrics, ESG Score, Carbon Intensity, Portfolio Selection.
JEL Code:

*Department of Statistics, Sapienza University of Rome, Rome, 00185 Rome, Italy. E-mail: rita.decclesia@uniroma1.it

†Department of Statistics, Sapienza University of Rome, Rome, 00197 Rome, Italy. E-mail: giacomo.morelli@uniroma1.it

‡Corresponding author. Department of Economics and Social Sciences, Sapienza University of Rome, Rome, 00185 Rome, Italy. E-mail: kevyn.stefanelli@uniroma1.it

1 Introduction

Among the large variety of low carbon green financial assets, the environmental, social, and governance (ESG) exchange-traded funds (ETFs) boomed in popularity passing from a total value of 6 billion USD in 2013 to 25 in 2019, with an annual growth rate that hit 45% in 2018 and has exceeded 200% over 2020.[1] Moreover, projections on the growth of this market see two out of three ETFs expected to be low-carbon by 2030, according to a 2018 report by Morningstar (2018).

However, the lack of transparency and consistency within these assets raises the question of whether they are really green or simply the result of marketing initiatives (greenwashing). For instance, albeit the large part of the ESG ETFs adopts screening processes based on the exclusion of the sins industries (e.g., tobacco and weapons), they include also companies that are directly involved in fossil fuels business.[2] The way environmental performances of the green funds are evaluated remains unclear. The large set of currently available metrics often presents large contradictions, generating confusion in investors, especially among those retailers. Low-carbon ETFs are solely evaluated using licensed environmental scores defined by the issuers of the same assets, that declare as *green* their products without providing exhaustive information on the funds' selection strategies.

In this paper, we analyze a sample of the ten most capitalized global green ETFs quoted from 1/1/2006 to 21/10/2021 retrieving financial and environmental information relative to all the companies included in the ETFs in 2021. We create a unique dataset of 246 firms listed on global exchanges. We evaluate the environmental performance of each company according to two of the most popular metrics, namely the Environmental (E) pillar of the ESG Score and the carbon intensity, which is a factor proportional to the total net CO_2 emissions. We screen the sample through a stock-picking process based on the two metrics to build equally weighted and global minimum variance synthetic funds considering: (i) only the best-in-class green companies, (ii) exclusively the best half of the sample, and (iii) all the firms except the most polluting ones, that belong to the last quantile of the environmental metrics (EMs) distributions. The resulting synthetic funds are statistically compared with each other and the sample of the ten listed green ETFs, both evaluating their environmental and financial performances. Furthermore, we test the resilience of synthetic funds during periods of market distress.

Our findings are as follows. The E Score and the carbon intensity show large discrepancies, with companies that are considered as green according to a metric and brown for the other. This evidence determines differences in the synthetic funds over the years and, from a financial point

[1]https://unctad.org/system/files/official-document/diae2020d1_en.pdf

[2]https://www.weforum.org/agenda/2021/07/esg-exchange-traded-funds-not-as-green-as-you-think/

of view, significant discrepancies that emerge between the best-in-class equally weighted funds. The comparison between the synthetic and the market securities highlights poor environmental performances of the sample of green ETFs compared to those of the synthetic green funds which also show similar, and sometimes higher Sharpe Ratios. Moreover, the Paris Agreement (December 12, 2015), which is widely considered as a signal of the increase in the investors' environmental concerns, does not entail a variation in the environmental metrics discrepancies.

The paper is organized as follows. Section 2 reviews the literature. Section 3 describes the methodology and reports the results of the empirical analysis and Section 4 concludes. Additional tables and figures are included in the Appendix **??**.

2 Literature Review

The definition of green investments appears blurred and often ambiguous, also considering a large number of environmental metrics available for the large part of public companies along with their dissimilarities and the difficulties in making a comparison among them (Angelakoglou and Gaidajis (2015), Morioka and de Carvalho (2016), and Thomä et al. (2018)). In fact, the divergences among empirical results on the relationship between company environmental performance (CEP) and company financial performances (CFP) are primarily due to the lack of a unique and clear environmental metric (Capelle-Blancard and Monjon (2012), Diaz-Rainey et al. (2017), van Dijk-de Groot and Nijhof (2015)).

Popescu et al. (2021) propose a classification of the most popular environmental metrics, providing an evaluation method based on several factors including coverage, reliability, fungibility, and transparency. Others, like Cabello et al. (2014), Petrillo et al. (2016), and Bender et al. (2019), define their environmental measures based on quantitative and qualitative factors. In June 2020, the European Commission (2020) (EUC) begin the process of the definition of a sustainable investment metric, called EU Taxonomy, aiming to be the gold standard for green finance worldwide. However, a recent report of the European Securities and Markets Authority (2021) (ESMA) shows that only 1-2% of the existing financial assets can be labeled as green according to the new EU Taxonomy.

In the last years, several rating agencies provide their licensed ESG Scores, which became popular because they are easy to use and cover a large part of public companies. Each ESG Score is the result of the aggregation of three different company responsibility disclosures based on the relative environmental (E), social (S), and governance (G) performances. However, the existing ESG Scores have been largely criticized in literature, for the lack of intertemporal coherence, transparency, and the inconsistency among the different rating agencies' methodologies (Avramov et al. (2021), Berg et al. (2019), Brandon et al. (2019), Escrig-Olmedo et al. (2019)). Chapter **??** reports an exhaustive literature review on the ESG scores features and issues.

Alternatively, the companies' total CO_2 emissions are widely used as a proxy to estimate the CEP (Coeslier et al. (2016), Garvey et al. (2018), Huang et al. (2021)). This metric is largely influenced by the company dimension, the industrial sector analyzed and suffers from the ambiguity in the international definitions of the greenhouse gases (GHG) (Jeswiet and Nava (2009), Johnson (2009), Watkins and Durning (2012)). Agents overcome the first issue by using the carbon intensity, evaluating the CEP in terms of the quantity of net CO_2 emissions necessary to produce one dollar of revenues.

Researchers still disagree on the relationship between CFP and CEP. On that note, several authors like Statman and Glushkov (2009), Friede et al. (2015), and Busch and Lewandowski (2016), indicate this association as positive, while Halbritter and Dorfleitner (2015), for instance, excluding any sort of relations, stressing the dependence of results on the time frame considered for the analysis. Others, like Delmas et al. (2015), show a decline in the CFP after a CEP improvement in the short run, but a potential growth in the long term. Differently, Busch et al. (2020) show companies as financially unable to reduce their carbon footprint without the support of the regulators.

Matsumura et al. (2014) using a dataset of S&P500 firms involved in the Carbon Disclosure Project between 2006 and 2008 show an inverse relation between tons of carbon emissions produced and firms' values.[3] Using a larger and more recent data sample, Bolton and Kacperczyk (2021) find a relation between emissions reduction and stocks returns, which turns out to be not significant using the carbon intensity, instead. Moreover, Andersson et al. (2016) propose a long-term passive investing strategy to hedge the climate risk by building a low-carbon index, which reduces by 50% the total investment carbon footprint and successfully tracks the target index performance.[4] Similarly, Capasso et al. (2020) measuring the effect on credit risk show that the distance to default decreases for companies characterized by low emissions intensity, with the effect that increases in mean after policymakers interventions.

The literature on the green ETFs performance is scarce, offering several possible research insights. Several works study the profitability of the energy ETFs sector overall, showing trading strategies that lead to good financial performances (Papailias and Thomakos (2013), Thomakos and Papailias (2013)). Moreover, Alexopoulos and Thomakos (2016) highlight the risk mitigation effect of the energy ETFs for the US market sector, which is solely characterized by high volatility. Malinda and Hui (2016) show the energy ETFs as characterized by long-term volatility and negative asymmetry volatility. Marszk (2019) studies the growth of the green ETFs during the period

[3] https://www.cdp.net/en

[4] The authors also denote the fact that in this way investors effectively get a "free option" on carbon.

2006-2017, observing the EU market became larger than that US only at the end of 2017. The inflow of capital and the consequent increase in the profitability of the green ETFs rose after the Paris Agreement (Fahmy (2021), Fahmy (2022), Lantushenko et al. (2021)). Alexopoulos (2018) exhibits how an energy ETFs portfolio outperforms both fossil fuels and clean energy ETFs based portfolios, because of its larger diversification. Furthermore, in line with Dutta et al. (2020), the clean energy funds are shown as more affected by periods of market uncertainty than those fossil fuels, having suffered more the 2008 global financial crisis. Similarly, Henriques et al. (2022) build an efficient portfolio choosing among 60 mixed energy ETFs observed from 2014 to 2018. Results reveal the natural gas and oil-based funds as the most represented assets in the optimal portfolios, while the renewable energy ETFs are often excluded from the holdings.

3 Data and methodology

We select a sample of the ten most capitalized green ETFs worldwide as representative of their market sector. The list of the sampled ETFs is reported in Table 1. The descriptive analysis of these financial assets is reported in the Chapter **??** of this thesis.

ETF	**Ticker**	**Sector**	**Area**	**ESG**
Lyxor New Energy (DR) UCITS ETF	ENER	Clean Energy	Global	AA
Invesco MSCI Sustainable Future ETF	ERTH	Low Carbon Firms	Global	BBB
First Trust Global Wind Energy ETF	FAN	Wind Energy	Global	AA
First Trust NASDAQ Clean Edge Smart Grid Infrastructure Index ETF	GRID	Smart Grid	USA	AA
iShares Global Clean Energy ETF	INRG	Clean Energy	Global	A
Invesco Global Clean Energy ETF	PBD	Clean Energy	Global	A
Invesco WilderHill Clean Energy ETF	PBW	Clean Energy	USA	A
First Trust NASDAQ Clean Edge Green Energy Index ETF	QCLN	Clean Energy	USA	A
VanEck Vectors Low Carbon Energy ETF	SMOG	Clean Energy	Global	A
Invesco Solar ETF	TAN	Solar Energy	Global	A

Table 1: Description of the green ETFs comprised in the final sample of analysis observed from January 2006 to October 2021.

We collect data from January 2006 to October 2021 regarding all the companies comprised in these funds according to their holdings reports available in 2021: (i) the financial summary which contains the company fundamentals (e.g., gross profits, revenues), (ii) the ESG statement view that contains all the statistics on the CEP, and (iii) the daily market adjusted-closing prices of these firms. Data are retrieved from the Refinitiv workspace.

Table 2 shows some features of these ETFs. Some companies comprised in the green ETFs holdings show missing environmental or financial information for all the sample period on the Refinitiv data source. However, we do not find any systematic reason for the missing data (e.g., specific country or sector), and then we consider these deficiencies as completely at random. For this reason, we decide to exclude companies characterized by missing data from the sample. Nevertheless, the proportions of holdings analyzed for each ETF are considered sufficient to represent the entire fund composition. In particular, the proportion of companies for which we obtain information over the total funds' composition swings between 59,90% (PBW) and 97,90% (GRID).[5] The amount of sampled companies belonging to each ETF spans between 24 (TAN) and 91 (ERTH), while the funds Refinitiv ESG scores are comprised between BBB (ERTH) and AA (ENER, FAN, and GRID).

We observe several differences in the companies included in the green ETFs. In particular, they belong to different sectors, according to The Refinitiv Bussiness Classification (TRBC), and are listed on exchanges all over the world, as Figure 1 and Figure 2 show. Among the most chosen business, we find the "Electric Utilities & IPPs", the "Machinery, Tools, Heavy Vehicles, Trains & Ships" and the "Renewable Energy" sectors. The "Aerospace & Defense", the "Collective Investments", the "Electronic Equipment & Parts", the "Natural Gas Utilities", and the "Transport Infrastructure" sectors are among the less present business sectors among the ETFs holdings. A large part of the companies is US-based. The small, albeit significant, amount of firms belonging to the emerging markets, such as the Asians and the Australian, makes the sample representative of the entire global exchanges. Moreover, a consistent number of companies has been founded, or at least listed, less than twenty years ago.

[5]The negligible amount of cash owned in different currencies is not considered in the computation of proportions.

ETF	Prop Cov	# Comp	ESG rating
ENER	91.00%	32	AA
ERTH	86.60%	91	BBB
FAN	89.90%	40	AA
GRID	97.90%	62	AA
INRG	92.60%	63	A
PBD	61.20%	76	A
PBW	59.90%	41	A
QCLN	76.90%	41	A
SMOG	95.10%	65	A
TAN	77.90%	24	A

Table 2: Summary of the data on ETFs concerning the period January 2006 - October 2021: Prop Cov, the proportion of the sampled companies over the total ETFs holdings, # Comp, the number of these companies, and the ESG rating provided by Refinitiv.

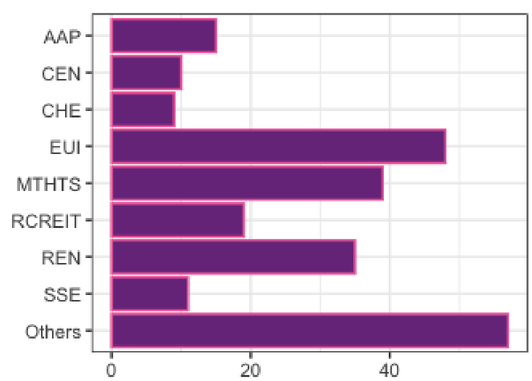


Figure 1: Distribution of companies by industrial sector (TRBC) over the period January 2006 - October 2021. Legend: AAP = "Automobiles & Auto Parts", CEN = "Construction & Engineering", CHE = "Chemicals", EUI = "Electric Utilities & IPPs", MTHTS = "Machinery, Tools, Heavy Vehicles, Trains & Ships", RCREIT = "Residential & Commercial REITs", REN = "Renewable Energy", SSE = "Semiconductors & Semiconductor Equipment", while Others contains the least represented sectors.

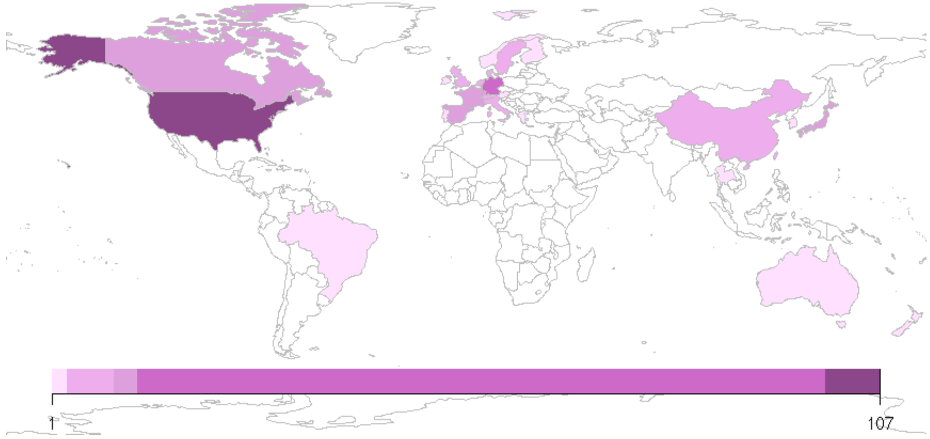


Figure 2: Distribution of sampled companies all over the world between January 2006 and October 2021.

3.1 Metrics comparison

In this paper, we analyze the environmental performance of the green ETFs holding using two of the most popular metrics, namely the E Score (ES) and the carbon intensity (CO_2S). The reasons to use these metrics are: (i) their large coverage of public companies, (ii) their ease of understanding, and (iii) their popularity, which allows us to compare the outcome of this analysis with the results of a wide range of financial papers.

Table 3 shows the descriptive statistics of the ES and the CO_2S, over the sample period.[6] While the ES ranges between 0 and 1 and represents a pure number, the CO_2 is expressed in total net CO_2 emissions over million of revenue in USD. The minima of the ES range from 0.08 (2013) to 3.27 (2015), while for the CO_2S it ranges from 21.60 (2008) to 0.01 (2018). Similarly, we do observe an increase in the yearly average value of the E Score, as well as the relative median, but this trend flattens around 55 after a couple of years of observation. The two distributions heavily differ in shape. While the ES distribution shows slight negative skewness (the mean is moderately lower than the median) we observe a large positive skewness in the CO_2S, with the average value which is definitively influenced by the high values in the right tail of the distribution. In this regard, a distribution characterized by fat tails is considered useful to pursue a screening process that excludes the most polluting firms from the portfolio. The yearly averages of the CO_2S measure do not show a clear trend along the years, with values that span between 15955.73 (2007) and 67338.10 (2010), while the medians range from 577.76 (2008) to 1172.38 (2017). The highest values, respectively for

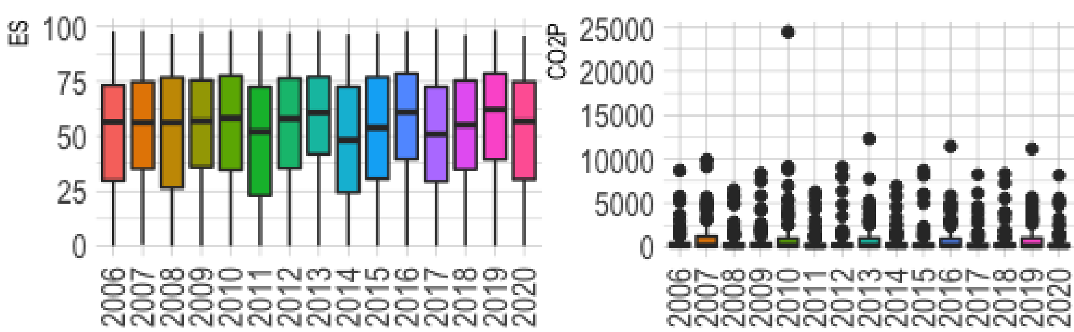
[6]Refinitiv computes the carbon intensity as the ratio between the total net CO_2 emissions and the revenues expressed in million USD.

the E Score and the CO_2S, are reached in 2020 (99.02) and 2010 (1617924.13), assuming values remarkably different from the rest of the distribution. However, the maxima distribution over the years relative to the carbon intensity fluctuates more than that relative to the ES. The distributions of the two metrics are graphically summarized in Figure 3a and Figure 3b, where the box plots indicate the yearly ES distribution as fairly symmetric and characterized by the absence of outliers, which conversely are conspicuous in that of the CO_2S.

Figure 4 shows the correlations obtained by comparing yearly the two environmental scores distributions of firms over the period analyzed as:

$$Cor(ES_t, CO_2S) = \frac{\mathbb{E}[(ES_{i,t} - \bar{ES}_t)(CO_2S_{i,t} - C\bar{O}_2S_t)]}{\sigma_{ES,t}\sigma_{CO_2S,t}}, \tag{1}$$

where for each year t N_t is the number of companies available, $\bar{ES}_t$ and $C\bar{O}_2S_t$ are the two EMs annual sample averages, and $\sigma_{ES,t}$ and $\sigma_{CO_2S,t}$ the relative standard deviations. The magnitude of the association ranges between -0.17 (2007) to 0.05 (2010), assuming values close to zero all over the time horizon considered. This evidence indicates the lack of coherence between the ES and the CO_2S and consequent dependence on the environmental metric in the evaluation of the CEPs. While this yearly comparison is considered as informative to exhibit the green features of the sample, we denote two main related issues: (i) the sample of companies changes each year, and (ii) the ES lacks also intertemporal coherence, with aggregating techniques and the set of environmental variables considered that has varied over the past years.



(a) (b)

Figure 3: ES (3a) and CO_2S (3b) distribution by year (January 2006 - October 2021). Points indicate outliers.

	E Score						CO_2S					
Year	Min	Mean	Med	Max	SD	NComp	Min	Mean	Med	Max	SD	NComp
2006	0.61	39.08	39.19	95.09	24.98	64	20.95	23351.46	1000.82	268082.48	58894.63	36
2007	0.29	48.76	51.98	93.08	24.77	74	18.93	15955.73	793.89	198589.67	41979.94	42
2008	3.21	53.43	58.45	96.75	27.01	84	21.60	48175.25	577.76	1146486.56	175359.78	50
2009	0.19	53.33	59.62	98.17	28.50	93	9.57	50455.47	687.37	1129884.97	170938.17	58
2010	0.63	55.66	58.42	95.85	27.13	101	10.90	67338.10	624.22	1617924.13	230855.99	64
2011	2.51	56.05	59.64	97.47	27.35	106	4.60	62891.85	660.33	1537196.70	208640.18	67
2012	1.23	55.38	59.89	95.93	27.37	108	1.78	55366.55	864.79	1204602.97	172036.54	71
2013	0.08	55.38	57.98	96.47	26.33	112	5.11	50764.13	889.81	942360.65	141686.04	73
2014	2.96	54.34	58.12	98.25	26.93	118	7.18	46983.01	917.15	975236.24	140551.02	81
2015	3.27	52.48	55.97	97.67	27.81	139	0.26	50260.54	892.03	1057566.16	142245.18	90
2016	0.57	51.37	56.03	97.97	28.20	153	0.22	53760.80	995.77	888971.32	139836.59	94
2017	1.70	53.34	59.93	98.39	27.57	167	6.42	50845.85	1172.38	850251.44	136712.38	107
2018	2.29	53.88	57.53	98.16	27.31	192	0.01	45062.39	1045.88	770481.68	123818.45	118
2019	0.45	54.38	56.55	97.18	27.04	224	1.81	42609.42	1099.87	716296.63	119528.95	139
2020	1.65	55.16	57.70	99.02	26.18	242	0.38	39002.99	875.19	776970.15	112858.19	152

Table 3: Descriptive statistics relative to the ES and the CO_2S metrics observed in the sample between January 2006 and October 2021.

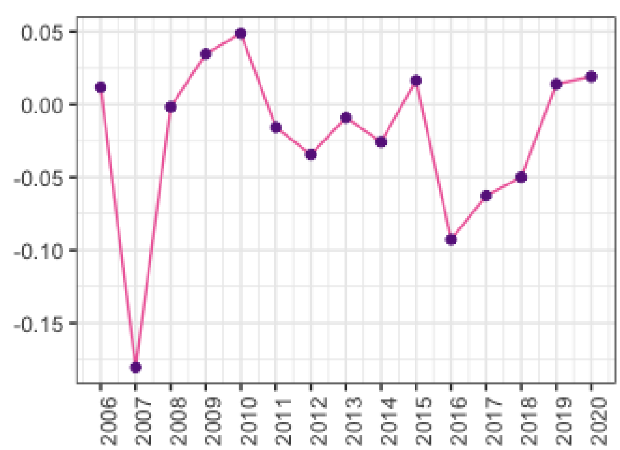


Figure 4: Time varying correlation between the two environmental metrics (January 2006 - October 2021).

3.2 The screening process for a synthetic green fund

We exploit the ES and the CO_2S to discriminate companies in the sample of green ETFs and build two synthetic green funds (SGFs) accordingly. The screening process consists of the exclusion of all the firms which exceed a fixed threshold, used to discriminate between green and brown companies. In particular, we choose as cut-off values the first, the second, and the third quantiles of the EMs distributions, which allows us to conduct a sensitivity analysis, as well. For instance, the best-in-class subset is chosen by selecting only the companies that show ES (CO_2S) values greater (lower) than the third (first) quantile of the distribution. Alternatively, choosing the median as a threshold, we emulate an investment strategy that considers only the best half of the sample. Finally, choosing the first and the third quantile of the distribution, respectively for the ES and the CO_2S, we screen the sample by excluding only the most polluting firms. The first method allows us to obtain a subset that includes only the greenest companies, but its limited number of assets could be a drawback, with the second and the third subsets which permit a selection over a larger sample of companies that potentially offers greater financial performances for the synthetic funds.

We calibrate the weights of the funds only considering the environmental performances relative to the previous financial year. In other words, for each year t, the feasible sample is composed of companies that show the best CEP relative to the year $t-1$. The composition of funds is decided at the beginning of each financial year and the weights remain constant for the entire period.[7] As shown in Equation (2), the holdings of each fund F in the year t are chosen according to the environmental metric (EM) and the quantile q as follows:

$$
\begin{aligned}
F_{EM_{q,t}} &= \theta(B_{EM_{q,t-1},t}) \\
B_{EM_{q,t}} &= \{C_{i,t} | EM_{C_{i,t-1}} \leq EM_{q,t-1}\}, \quad i = 1, 2, \ldots, s_t \leq N_t
\end{aligned} \tag{2}
$$

where θ is a portfolio selection function, $B_{EM_q,t-1}$ is the set of N_t companies, $C_{i,t-1}$, defined as "green" in the previous year $t-1$ according to the EM and the quantile q, with $i = 1, .., s_t$.

We choose θ among popular funds selection strategies: (i) equally weighted funds (EWF), and (ii) global minimum variance funds (GMVF), which assigns portfolio weights for the year t according to the covariance matrix Σ_{t-1} estimated at time $t-1$. Each portfolio weight $\omega_{i,t}$ relative to the asset i at time t according to the EWP strategy is chosen as:

$$
\omega_{i,t} = 1/s_t, \quad i = 1, 2, \ldots, s_t, \tag{3}
$$

[7] The funds' holdings are dynamically update once in a year because of the annual frequency of the environmental data.

while for the GMVF $\omega_{i,t}$ represents the solution of the following minimization problem:

$$\begin{aligned} \min_{\omega_t} \frac{1}{2}\omega_t' \Sigma_{t-1}\omega_t \\ s.t.\;\; \omega_t' 1_n =& 1; \\ \omega_t \geq& 0. \end{aligned} \tag{4}$$

For each metric, we build six synthetic funds as the combinations of the three subsets individuated by the three quantiles of the EMs distributions chosen as cut-off values (Q25, Q50, and Q75) and the two portfolio selection strategies, the EWF and the GMVF. We compare the SGFs with each other measuring the effect on the financial performance entailed by a more (less) severe screening process and from the choice of alternatives EMs. Then, we operate a comparison between the synthetic green funds and the listed green ETFs, in terms of composition, and environmental and financial performances.

3.3 Empirical results

We identify the SGFs by the metric used (S and C, respectively for the ES and the CO_2S), the quantile chosen (25, 50, and 75, respectively for the third (first), the second, and the first (third) quantile of the ES (CO_2S) distribution) and the portfolio selection strategy (E and G, for the EWF and the GMVF). For instance, ${}_E SGF^C_{25}$ refers to the fund built using the CO_2S environmental metric (C), selecting only the companies belonging to the first quantile of the distribution (25), with weights chosen according to the EWP technique (E).

Year	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19	'20
ES															
B_{25}	16	19	21	23	25	27	27	28	30	35	38	42	48	56	61
B_{50}	32	37	42	46	50	53	54	56	59	69	76	83	96	112	121
B_{75}	48	55	63	69	75	79	81	84	88	104	114	125	144	168	181
CO_2P															
B_{25}	9	11	13	15	16	17	18	18	20	23	24	27	30	35	38
B_{50}	18	21	25	29	32	33	35	36	40	45	47	53	59	69	76
B_{75}	27	31	37	43	48	50	53	54	60	67	70	80	88	104	114

Table 4: Number of companies included into each subset $B_{q,t}$ for each year t and quantile q divided by the two metrics, ES and CO_2P (January 2006 - October 2021).

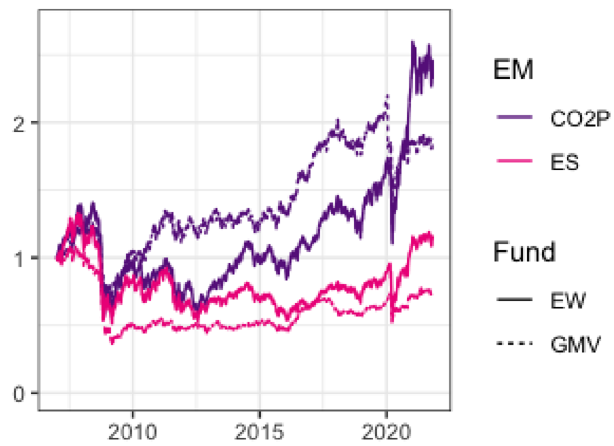
Table 4shows the number of companies selected for each year according to the quantile criteria. The increasing number of companies observed over the sample period is related to the lack of environmental data in the first years of observations for some of the green ETFs holdings. The sample contains companies that have been founded, or at least quoted, in recent years, and hence it is impossible to retrieve any kind of data relative to these firms in the early 2000s. Moreover, the coverage of the two environmental metrics largely increased in the last years, with that of the ES which results larger than the other all over the sample period.

Figure 6, Figure 7, and Figure 8 in the Appendix **??** show the Venn diagrams of the subsets of companies selected for each quantile based on the two environmental metrics by year. The two subsamples of best-in-class companies are strongly different with just a bunch of companies that are considered to be green according to both the ES and the CO_2S metrics. Differently, the holdings of the ${}_{\cdot}SGF^{\cdot}_{50}$ and the ${}_{\cdot}SGF^{\cdot}_{75}$ synthetic funds are more similar to each other, because of the less severe screening processes adopted that reduces the divergences in the EMs.

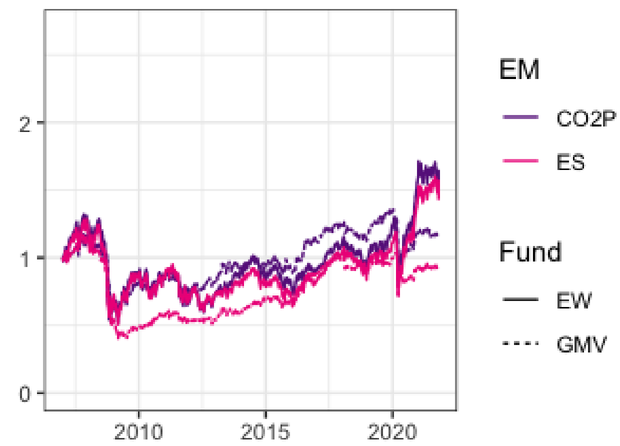
Figure 5 shows that both the ${}_{\cdot}SGF^{C}_{25}$ funds exhibit the largest cumulative returns and the discrepancies between the SGFs built according to the two metrics gradually narrow considering the ${}_{\cdot}SGF^{\cdot}_{50}$ and ${}_{\cdot}SGF^{\cdot}_{75}$ funds, because of their similar compositions. These results point out the importance of the screening process in determining the funds' financial performances. The differences between the two environmental metrics arise exclusively considering the best-in-class companies. While the ${}_{\cdot}SGF^{C}_{25}$ funds outperform all the others in terms of cumulative returns, highlighting a clear relation between CEP and CFP, they almost double their ES peers, ${}_{\cdot}SGF^{S}_{25}$, in the last period of observation, indicating also a dependence on the environmental metric chosen.

The four ${}_{\cdot}SGF^{\cdot}_{25}$ are subjected to the same fluctuations along the years, showing similar patterns and variations. Comparing the two best-in-class SGFs by portfolio selection technique, the GMVF profits are larger (lower) than those of its EWF peer for the CO_2S (ES) metric for almost the entire sample period. Moreover, in the last year of analysis, the ${}_{E}SGF^{C}_{25}$ exhibits the largest cumulative returns, outperforming all the SGFs after the COVID-19 global crisis. The ${}_{E}SGF^{S}_{25}$ shows slight profits only in the first and in the last periods of analysis, while the ${}_{E}SGF^{S}_{25}$ financial outcome remains negative. This result highlights the difficulties of these two SGFs in recovering the losses faced during the 2008 global financial crisis, with the ${}_{E}SGF^{S}_{50}$ and ${}_{E}SGF^{S}_{75}$ that show positive financial outcomes, in terms of cumulative returns, determined by a greater diversification.

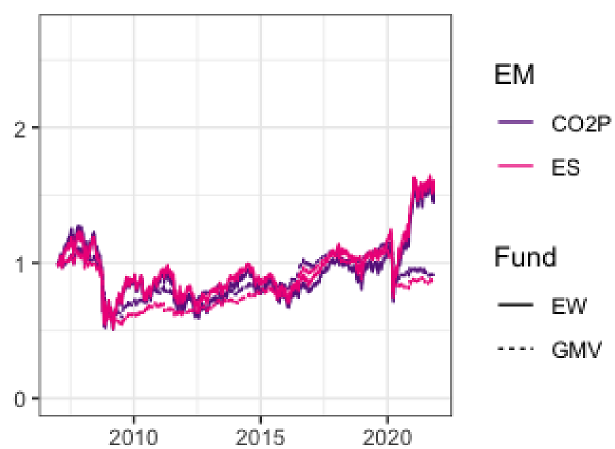
Table 5 shows the results of the tests conducted on the differences between the ES and CO_2S SGFs returns. We test the differences in average returns (t-test), volatilities (F-test), and Sharpe Ratios (SRs) as shown in Section **??**. Results point out significant differences in the financial performances exclusively for the SGFs built according to the most strict screening processes, with



(a)



(b)



(c)

Figure 5: Comparison among the cumulative returns of the SGFs estimated according to the two selection criteria (EW and GMV) over the period Jan 2007 - Oct 2021 divided by quantile: ${}_{\cdot}SGF^{\cdot}_{25}$ in (5a); ${}_{\cdot}SGF^{\cdot}_{50}$ in (5b) and ${}_{\cdot}SGF^{\cdot}_{75}$ in (5c).

the ${}_{E}SGF^{C}_{25}$ that shows a SR greater than its ES peer. Moreover, all the CO_2 SGFs exhibit volatilities that are statistically larger than those relative to the ES funds, except the ${}_{G}SGF^{C}_{50}$, which suggests a significative impact on the financial performances depending on the chosen EM.

We study whether the growth in the investors' sensitivity to environmental themes observed in the last years entails possible variations in the way companies are environmentally evaluated and the relative impact on the financial outcome. In other words, conscious of the temporal incoherence of some EMs, we are searching for potential discrepancies generated by the demand for a more accurate

SGF	t-test	F-test	HAC-test
$_{E}SGF^{\cdot}_{25}$	0.0001	0.8756 ***	-0.013 **
	(0.000)	(–)	(0.006)
$_{E}SGF^{\cdot}_{50}$	0.0001	0.9390 *	-0.001
	(0.000)	(–)	(0.004)
$_{E}SGF^{\cdot}_{75}$	0.0001	0.9098 ***	0.002
	(0.000)	(–)	(0.003)
$_{G}SGF^{\cdot}_{25}$	-0.0005	0.7976 ***	-0.021
	(0.000)	(–)	(0.015)
$_{G}SGF^{\cdot}_{50}$	0.0000	0.9694	-0.014
	(0.000)	(–)	(0.012)
$_{G}SGF^{\cdot}_{75}$	0.0000	0.8475 ***	0.002
	(0.000)	(–)	(0.008)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 5: Results of the tests, statistics and relative standard errors (in brackets), conducted on the differences in financial performances between each couple of SGFs built according the two EMs and the same subset of companies over the period Jan 2007 - Oct 2021.

evaluation of the company's carbon footprint. In particular, we study if after the Paris Agreement the divergences between the two metrics have increased, along with the financial performances of the SGFs built accordingly. Table 7 exhibits the results of the tests conducted exclusively on the subset of data after December 12, 2015. The findings are as follows. The EW SGFs show significative discrepancies in the average returns, the $_{\cdot}SGF^{\cdot}_{25}$ reject the hypothesis of null differences in volatilities, while all the SRs tests result not statistically significative. These mixed results are too weak to conclude through a significant impact of the Paris Agreement on the EM discrepancies.

We investigate the dependence between the SGFs financial outcomes and the EMs during different phases of the market. Table 8, Table 9, and Table 10 show the comparison between each pair of SGFs, ES and CO_2S, separated by year. Results suggest that the EMs discrepancies largely emerge for almost all the SGFs during phases of tumultuous markets, as during the GFC years (2008-2009), and in periods of market expansion, as in 2019 where the green ETFs experienced a large growth in volumes inflows and returns. However, only the $_{G}SGF^{\cdot}_{25}$ and the $_{G}SGF^{\cdot}_{50}$ show differences in SRs during these periods, where the CO_2S funds outperform the others. Therefore, the relation between the way we evaluate the CEP and the variation in the CFP is more evident during phases of large fluctuations of the market, highlighting the role of the environmental screening process as a tool to ensure price stability and resilience to green financial assets.

We conclude this study by evaluating the ten ETFs sampled both in terms of environmental and financial performances, comparing them with those of the SGFs. Figure 9 shows how the holdings of the best-in-class SGFs in 2021 massively differ from those of the green ETFs considered. For instance, we have only one company that is included in both the ${}_{\cdot}SGF_{25}^{\cdot}$ and in the PBW, while the ERTH and the PBD holdings have only eight companies in common with those of the best-in-class SGFs, with which constitute the largest intersections in the sample. The huge discrepancies observed in the SGFs and in the listed funds compositions suggest that green ETFs managers use different EMs to evaluate the CEP and make investment decisions accordingly. On the other hand, one can think that green ETFs issuers follow a set of criteria in their financial choices, including that environment. However, this hypothetical strategy would exclude the large part of the best-in-class green companies according to two of the most popular EMs, raising doubts on the real carbon footprint of these green ETFs.

The environmental performance of the ten green ETFs and the SGFs are estimated by computing the weighted average of each EM relative to companies comprised in each of them. In other words, for the fund f and the year t, the optimal weight ω fixed for the company i is multiplied by the respective EM (ES or CO_2S) value as follows:

$$EM_{f,t} = \sum_{i=1}^{s_t} \omega_{i,t} \cdot EM_{i,t}. \tag{5}$$

Table 6 reports the results of the comparison relative to the year 2021. The highest ES among the ETFs is 70.72, relative to the FAN, but this value is close to that of the ${}_{E}SGF_{75}^{S}$, which is the lowest ES rated among the synthetic funds (67.06). The comparison between the ES values relative to the two ${}_{\cdot}SGF_{25}^{S}$, respectively 86.10 (EW) and 84.41 (GMV), with those of the PBW and the QCLN, respectively 33.81 and 33.98, increases the doubts on whether the green ETFs are green. Moreover, the carbon intensities of the ETFs are definitively greater than those of the SGFs. In particular, the FAN and the SMOG show values that exceed 63000, while the highest value among the SGFs is slightly over 1000 (${}_{E}SGF_{75}^{C}$) and the minimum is just 37.82 (${}_{G}SGF_{25}^{C}$). The results of this comparison suggest that the green ETFs are necessarily built according to (one or more) different environmental metrics or, as we conclude before, their managers do not consider the CEP as the first investment decision criteria.

Several contradictions emerge by the evaluation of the ETFs according to the two EMs studied. For instance, the FAN ETF shows the highest values for both the EMs, while the PBW shows the second-lowest score of ES and the minimum value of CO_2S. This evidence raises the question of why an environmental rating like the ES does not penalize (reward) assets characterized by a high (low) carbon intensity.

ETF	ES	CO_2S
ENER	59.90	6464.61
ERTH	58.52	4335.39
FAN	70.72	66109.28
GRID	68.11	21634.90
INRG	52.22	44146.08
PBD	51.81	3311.34
PBW	33.98	1641.19
QCLN	33.81	22471.11
SMOG	64.35	63137.71
TAN	54.55	12509.53

SGFs	ES	CO_2S
${}_ESGF^{\cdot}_{25}$	86.10	53.04
${}_ESGF^{\cdot}_{50}$	77.16	209.32
${}_ESGF^{\cdot}_{75}$	67.06	1018.91
${}_GSGF^{\cdot}_{25}$	84.41	37.82
${}_GSGF^{\cdot}_{50}$	75.73	245.64
${}_GSGF^{\cdot}_{75}$	69.46	715.68

Table 6: Environmental performances relative to the 2021 compositions of the ten ETFs sampled and the SGFs evaluated through the ES and the CO_2S metrics.

Table 11, Table 12, Table 13, Table 14, Table 15, and Table 16 present the results of the tests conducted on the average returns, the volatilities, and the SRs to evaluate the differences between the green ETFs and the SGFs from a financial perspective. A large part of the t-tests is not significant to assess differences in the returns of the ETFs and the synthetic funds. Interestingly, only the SGFs built according to the CO_2S metric exceed the ETFs performances in terms of SR. In particular: (i) the ${}_ESGF^C_{25}$ statistically outperforms the FAN, the INRG, the ENER, and the TAN; (ii) the ${}_ESGF^C_{50}$ beats the PBW, the FAN, the INRG, the PBD, and the TAN; while both (iii) the ${}_GSGF^C_{25}$ and (iv) the ${}_ESGF^C_{75}$ show SRs statistically greater than the INRG. Conversely, the ${}_ESGF^E_{25}$ is the only SGF that shows a SR statistically lower than an ETF (the ERTH). This evidence highlights the importance of the metric used to evaluate the CEP. Moreover, all the volatilities of the green ETFs result significative larger than those of the SGFs. However, this result is strongly influenced by the large growth of this class of assets in the period 2019-2020. The greater financial and environmental performances of these synthetic green funds compared to the listed green ETFs exhibit clear evidence of an existing relationship between CEP and CFP.

4 Conclusion

The last decade has been characterized by the large expansion of the green ETFs market, also influenced by the increase in the investors' environmental concerns. These ETFs show good financial performances and represent a valid alternative to green stocks for ethical investors. However, the lack of transparency in the way managers select the funds' holdings and in the environmental screening methodologies adopted begs the question of whether these financial assets are truly green.

In this paper, we analyze the environmental performance of 246 companies, which represent the 2021 holdings of the ten most capitalized green ETFs, according to two of the most popular environmental metrics, namely the E Score and the carbon intensity. We operate a yearly screening process based on the quantiles of the EMs distributions obtaining three different subsets that contain: (i) only the most environmentally sustainable companies, (ii) the best half of the sample, and (iii) all the firms except the worst-in-class. Given these subsets, we build a list of synthetic funds adopting an annual holdings recalibration on the bases of the previous year's environmental performances of the firms and two different portfolio selection criteria: (i) equally weighted and (ii) global minimum variance. Then, we compare the environmental and financial performances of the SGFs with the ten ETFs sampled to assess: (i) whether the green ETFs are really green according to the two metrics used, and (ii) if choosing companies exclusively according to their environmental features affects the funds' financial outcome. Moreover, we evaluate the potential discrepancies generated by the use of alternative EMs to determine the SGFs compositions and how it consequently impacts the financial result, also analyzing different phases of the market.

The analysis highlights poor environmental performances of all the green ETFs considered, according to both the metrics studied. For instance, the QCLN exhibits the lowest ES, albeit it is A-rated according to the ESG score. It shows an estimated ES value of 33.81/100 in 2021, which makes this ETF far to can be considered green. Overall, the highest estimated ES value among the ETFs is 70.72 (FAN) which is close to the minimum value within the SGFs (67.06 of the ${}_{E}SGF^{\cdot}_{75}$) and definitively lower than the relative maximum, that peaks at 86.10 (${}_{E}SGF^{\cdot}_{25}$). On the same line, the SGFs show CO_2S massively lower than the green ETFs, with values that amount to just a couple of dozens and several thousand in the two groups, respectively. This result confirms the large difference in terms of environmental performances between the listed funds and the synthetic assets built with the only purpose of being green.

The comparison between the funds' ES and the respective carbon intensities points towards a lack of coherence between the two metrics analyzed. The FAN ETF presents the highest values for both two measures, highlighting a contradiction between them and a strong dependence on the environmental metric used to evaluate the environmental performances.

The construction of the SGFs through a severe green screening process of companies assures high environmental scores according to the two metrics. However, we denote large dissimilarities in the composition of the ES and the CO_2S based funds, especially in those comprised of the best-in-class green companies. In particular, the large range of values assumed along with the fat tails that characterized the carbon intensity distribution allows us to individuate more adequately the low-carbon companies. On the contrary, the high concentration around the average value that characterizes the ES distribution represents an issue in discriminating green and brown firms.

From a financial perspective, we observe that the performances of the SGFs roughly coincide with, and sometimes exceed, those of the listed green ETFs. For instance, the ${}_{E}SGF^{C}_{25}$ shows a SR statistically greater than four out of ten ETFs (namely the FAN, the INRG, the ENER, and the TAN). Comparing the ES and the CO_2S funds we observe a significative difference in terms of SR only relatively to those most screened equally weighted funds, due to the discrepancies in the composition of these SGFs. In this case, the EW fund built according to the CO_2S metric assures also greater financial performances than its ES peer.

The results on the relation between the effect of the Paris Agreement on the SGFs returns and the environmental metric used to individuate the green companies results inconsistent, with no statistical differences found in the performance of the SGFs built according to the two EMs after 12 December 2015. Differently, some significative dissimilarities emerge from the yearly juxtapositions of the ES and the CO_2S funds during phases of tumultuous markets (e.g., 2008, 2009, and 2019), revealing a dependence of the well-known resilience of green assets on the environmental metric used.

The lack of transparency in the way companies are selected into the green ETFs raises the question of to what extent the ETFs managers consider the firms' environmental footprints as a decision criterion. The analysis has shown that the green performances of these funds are inadequate according to two of the most popular metrics currently available. However, the discrepancies detected between the ES and the CO_2S increase the confusion on how agents should measure the CEP. The evaluation of the environmental performances of the green ETFs according to the forthcoming EU Taxonomy rules and the definition of a more accurate, transparent, and granular green metric will be the object of future analysis.

5 Appendix

5.1 Venn diagrams to compare the funds holdings.

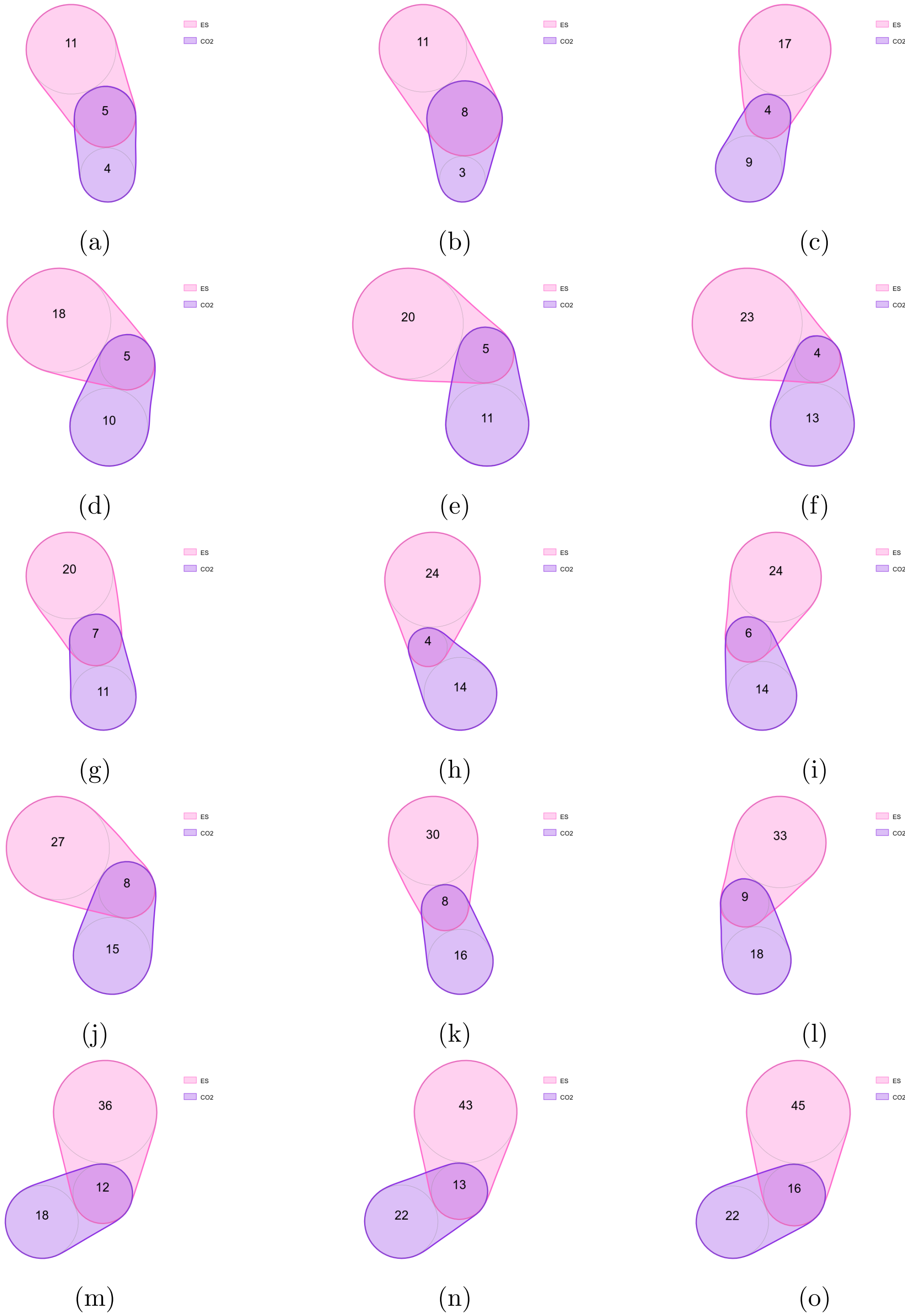


Figure 6: Venn diagrams per year of companies selected in the Q25 subsets according to the two environmental metrics from 2006 (6a) to 2020 (6o).

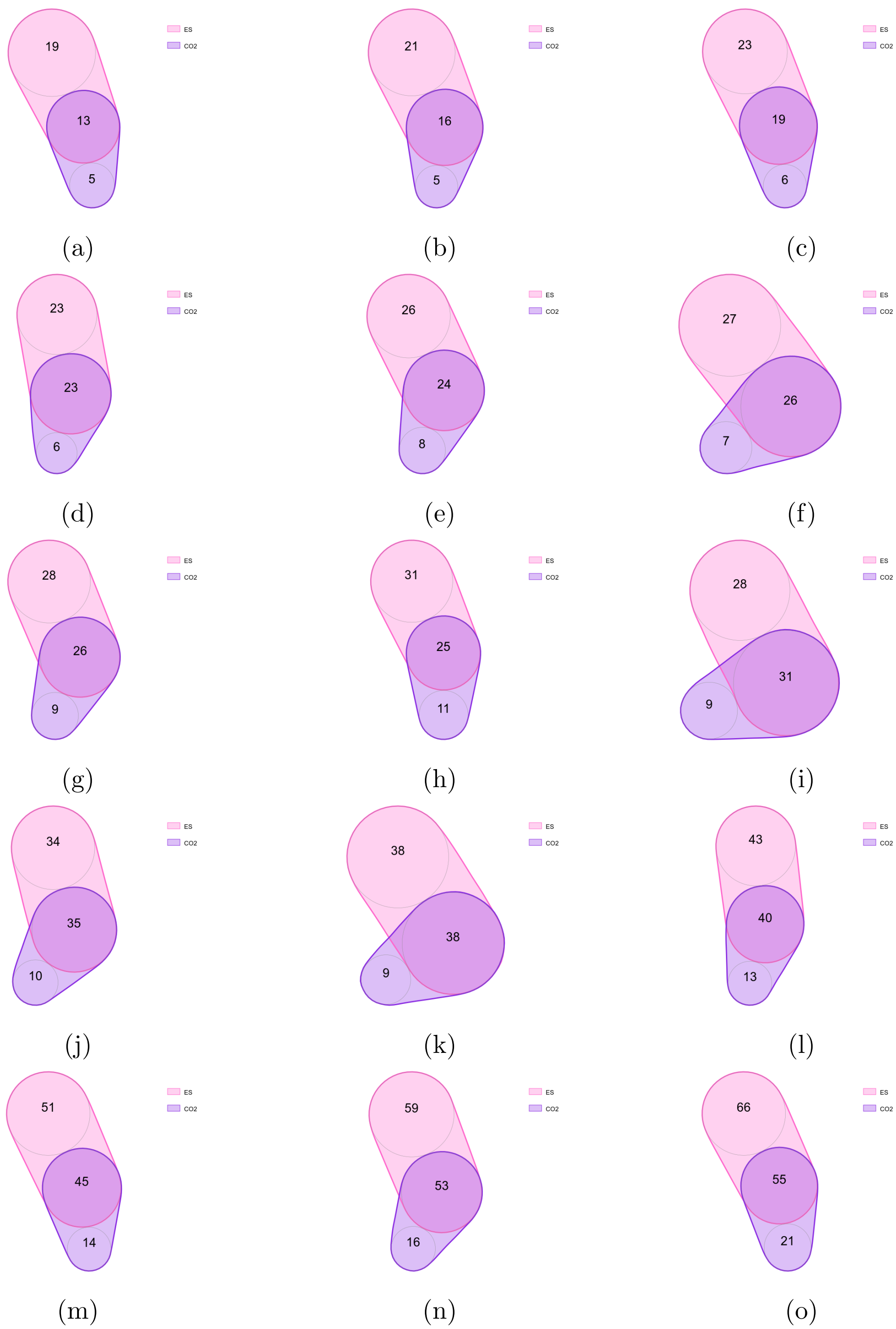
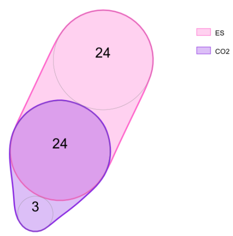
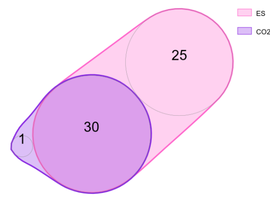


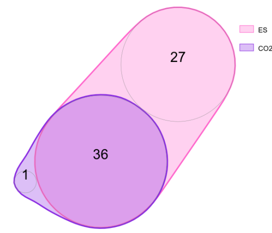
Figure 7: Venn diagrams per year of companies selected in the Q50 subsets according to the two environmental metrics from 2006 (7a) to 2020 (7o).



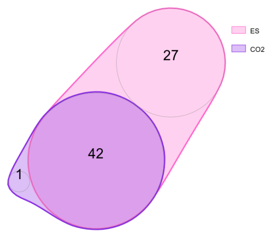
(a)



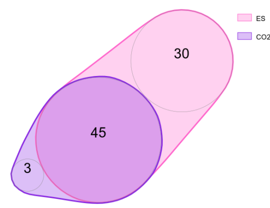
(b)



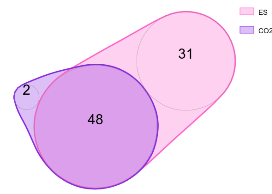
(c)



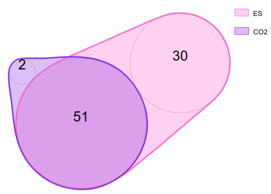
(d)



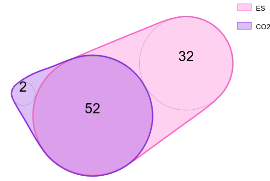
(e)



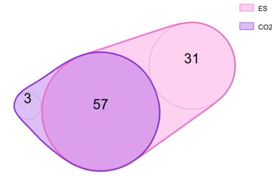
(f)



(g)

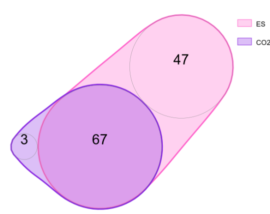


(h)

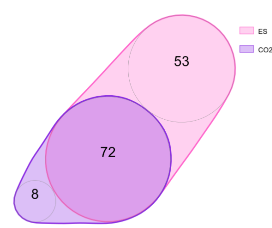


(i)

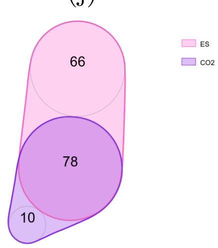
(j)



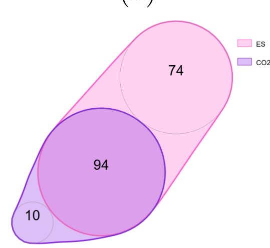
(k)



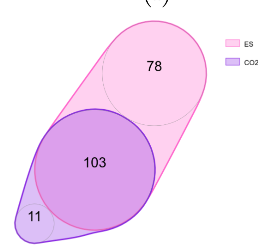
(l)



(m)



(n)



(o)

Figure 8: Venn diagrams per year of companies selected in the Q75 subsets according to the two environmental metrics from 2006 (8a) to 2020 (8o).

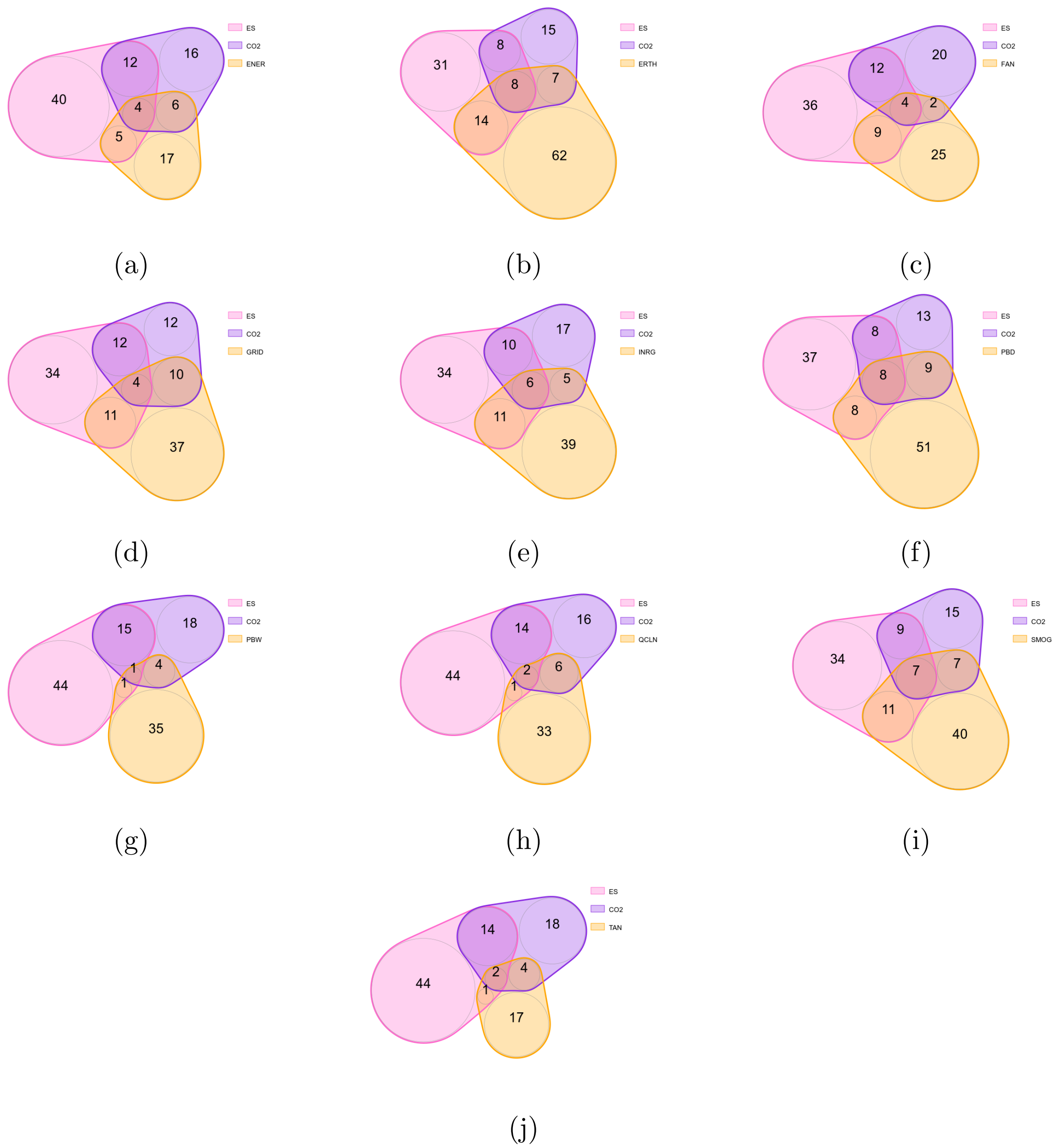


Figure 9: Venn diagrams of companies belonging to the first quantile of the ES and CO_2S distributions relatively to the year 2021 compared to the ETFs holdings.

5.2 Synthetic funds financial performances comparison after the Paris Agreement.

SGF	t-test	F-test	HAC-test
${}_ESGF^{\cdot}_{25}$	0.00053 **	0.87299 ***	-0.012
	(0)	(–)	(0.012)
${}_ESGF^{\cdot}_{50}$	0.00044 **	0.9965	0.003
	(0)	(–)	(0.007)
${}_ESGF^{\cdot}_{75}$	0.00045 **	0.97932	0.001
	(0)	(–)	(0.006)
${}_GSGF^{\cdot}_{25}$	0.00036	0.77505 ***	0.002
	(0)	(–)	(0.034)
${}_GSGF^{\cdot}_{50}$	0.00033	1.11441	-0.003
	(0)	(–)	(0.022)
${}_GSGF^{\cdot}_{75}$	0.00014	0.92626	0.011
	(0)	(–)	(0.015)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 7: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the SGFs built according the two EMs after the Paris Agreement (December 12, 2015 - October 21, 2021).

5.3 Synthetic funds financial performances comparison by year.

	EW			GMV		
Year	t-test	F-test	HAC-test	t-test	F-test	HAC-test
2007	0.00105 **	0.84278	-0.015	0.0004	0.87652	-0.065
	(0.001)	(–)	(0.026)	(0)	(–)	(0.06)
2008	-0.00231 *	0.79769 *	-0.009	-0.00373 ***	0.73599 **	-0.051 *
	(0.001)	(–)	(0.015)	(0.001)	(–)	(0.028)
2009	0.00081	0.91813	-0.017	0.00063	1.09691	-0.051
	(0.001)	(–)	(0.029)	(0.001)	(–)	(0.058)
2010	-0.00014	0.96169	0.015	0.0005	0.76748 *	-0.042
	(0.001)	(–)	(0.019)	(0)	(–)	(0.053)
2011	-0.00103	1.01729	-0.003	-0.00045	0.78323 *	-0.031
	(0.001)	(–)	(0.016)	(0.001)	(–)	(0.048)
2012	0.00021	0.86092	-0.001	-0.00007	0.74959 **	-0.019
	(0.001)	(–)	(0.028)	(0)	(–)	(0.079)
2013	0.0007 *	0.80914 *	-0.057 **	0.00001	0.64736 ***	0.015
	(0)	(–)	(0.029)	(0)	(–)	(0.065)
2014	0.00001	0.73601 **	-0.056 **	0.00036	0.64901 ***	0.035
	(0)	(–)	(0.028)	(0)	(–)	(0.065)
2015	-0.0004	0.9895	-0.042	0.00003	0.83479	-0.071
	(0)	(–)	(0.027)	(0)	(–)	(0.051)
2016	0.0004	1.08313	-0.021	0.00173 ***	0.86056	0.02
	(0)	(–)	(0.023)	(0)	(–)	(0.063)
2017	0.00082 ***	0.80282 *	-0.008	0.0003	1.02925	-0.153 *
	(0)	(–)	(0.034)	(0)	(–)	(0.087)
2018	-0.00038	0.79679 *	0.002	-0.00044	0.67462 **	0.002
	(0)	(–)	(0.029)	(0)	(–)	(0.067)
2019	0.00109 ***	0.67737 ***	-0.024	0.0011 ***	0.72678 **	-0.01
	(0)	(–)	(0.038)	(0)	(–)	(0.07)
2020	0.00108	1.0604	-0.029	-0.0006	0.72091 *	0.039
	(0.001)	(–)	(0.022)	(0.001)	(–)	(0.065)
2021	0.00003	0.39523 ***	0.034	-0.00016	1.24495	0.012
	(0.001)	(–)	(0.037)	(0)	(–)	(0.077)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 8: Results of the tests, statistics and standard errors (in brackets), conducted yearly on the differences in financial performances between the $.SGF_{25}^{S}$ and the $.SGF_{25}^{C}$, separated by fund selection strategy (January 2007 - October 2021).

	EW			GMV		
Year	t-test	F-test	HAC-test	t-test	F-test	HAC-test
2007	0.00088 *	0.82775	-0.005	0.0008 **	0.84326	-0.027
	(0)	(–)	(0.025)	(0)	(–)	(0.046)
2008	-0.00232 **	0.87461	0.001	-0.00317 ***	0.84931	-0.026
	(0.001)	(–)	(0.011)	(0.001)	(–)	(0.034)
2009	0.00089	0.94857	-0.003	0.00035	1.29895 *	-0.169 ***
	(0.001)	(–)	(0.012)	(0.001)	(–)	(0.057)
2010	-0.00004	0.99184	0.000	0.00064	0.67444 ***	0.062
	(0.001)	(–)	(0.012)	(0)	(–)	(0.038)
2011	-0.00086	0.98026	0.006	-0.00049	1.08937	0.017
	(0.001)	(–)	(0.009)	(0.001)	(–)	(0.038)
2012	0.00012	0.87831	-0.008	0.00021	0.82585	-0.058
	(0)	(–)	(0.016)	(0)	(–)	(0.07)
2013	0.00078 **	0.86785	-0.037 *	0.00058	0.65579 ***	0.001
	(0)	(–)	(0.022)	(0)	(–)	(0.062)
2014	-0.00009	0.8773	0.009	0.00066 **	0.99758	0.084
	(0)	(–)	(0.02)	(0)	(–)	(0.065)
2015	-0.00034	1.05331	-0.018	-0.00019	1.00745	0.02
	(0)	(–)	(0.016)	(0)	(–)	(0.038)
2016	0.00041	1.03132	0.011	0.00157 ***	0.88856	0.022
	(0)	(–)	(0.017)	(0)	(–)	(0.064)
2017	0.00081 ***	0.88596	0.032	0.00089 ***	0.86344	0.019
	(0)	(–)	(0.024)	(0)	(–)	(0.082)
2018	-0.00053	0.91843	-0.006	-0.00042	1.17905	-0.046
	(0)	(–)	(0.021)	(0)	(–)	(0.055)
2019	0.00094 ***	0.81081 *	-0.01	0.00093 ***	0.91835	-0.123 *
	(0)	(–)	(0.02)	(0)	(–)	(0.064)
2020	0.00101	1.06766	0.001	-0.00113	1.1936	0.046
	(0.001)	(–)	(0.011)	(0.001)	(–)	(0.041)
2021	-0.00015	0.91547	0.017	-0.00003	1.16607	0.004
	(0.001)	(–)	(0.017)	(0)	(–)	(0.062)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 9: Results of the tests, statistics and standard errors (in brackets), conducted yearly on the differences in financial performances between the $.SGF_{50}^{S}$ and the $.SGF_{50}^{C}$, separated by fund selection strategy (January 2007 - October 2021).

	EW			GMV		
Year	t-test	F-test	HAC-test	t-test	F-test	HAC-test
2007	0.00071 *	0.84047	-0.01	0.00037	0.8877	-0.018
	(0)	(–)	(0.017)	(0)	(–)	(0.046)
2008	-0.00217 **	0.84311	-0.001	-0.00286 ***	0.82577	-0.006
	(0.001)	(–)	(0.009)	(0.001)	(–)	(0.023)
2009	0.00092	0.96431	0.004	0.00025	0.67506 **	-0.024
	(0.001)	(–)	(0.01)	(0.001)	(–)	(0.038)
2010	0.00003	0.90946	0.007	0.00055	0.74022 *	0.007
	(0.001)	(–)	(0.007)	(0)	(–)	(0.031)
2011	-0.00088	0.88522	0.013 *	-0.0005	0.9029	0.011
	(0.001)	(–)	(0.007)	(0.001)	(–)	(0.023)
2012	0.00014	0.82924	-0.001	0.00031	0.86459	-0.015
	(0)	(–)	(0.012)	(0)	(–)	(0.044)
2013	0.00076 **	0.85684	-0.011	0.0003	0.72205 **	0.058
	(0)	(–)	(0.013)	(0)	(–)	(0.045)
2014	-0.00022	0.85434	0.012	0.00059 **	0.98391	0.01
	(0)	(–)	(0.011)	(0)	(–)	(0.024)
2015	-0.00029	1.00995	-0.015	-0.00013	0.96796	-0.004
	(0)	(–)	(0.01)	(0)	(–)	(0.031)
2016	0.00028	0.97749	0.01	0.00137 ***	0.95486	0.006
	(0)	(–)	(0.01)	(0)	(–)	(0.059)
2017	0.00092 ***	0.84166	0.009	0.00086 ***	1.11099	0.023
	(0)	(–)	(0.019)	(0)	(–)	(0.047)
2018	-0.0006	0.91323	0.002	-0.00048	1.22883	-0.017
	(0)	(–)	(0.015)	(0)	(–)	(0.033)
2019	0.00091 ***	0.76328 **	0.007	0.00102 ***	0.91424	-0.039
	(0)	(–)	(0.018)	(0)	(–)	(0.047)
2020	0.00106	1.08844	-0.005	-0.00209 *	0.88142	0.015
	(0.001)	(–)	(0.009)	(0.001)	(–)	(0.03)
2021	0	0.8135	0.005	-0.00035	0.95446	0.066
	(0.001)	(–)	(0.018)	(0)	(–)	(0.075)

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 10: Results of the tests, statistics and standard errors (in brackets), conducted yearly on the differences in financial performances between the $.SGF_{75}^{S}$ and the $.SGF_{75}^{C}$, separated by fund selection strategy (January 2007 - October 2021).

5.4 Results of the comparison between the financial performances of the ETFs and the synthetic funds.

$_{E}SGF^{\cdot}_{25}$						
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	0.00001	2.86698 ***	-0.003	0.0004	2.51079 ***	-0.016
	(0)	(–)	(0.012)	(0)	(–)	(0.011)
ERTH	0.00016	1.60026 ***	0.014 *	0.00025	1.40051 ***	0.000
	(0)	(–)	(0.008)	(0)	(–)	(0.008)
FAN	-0.00006	1.73174 ***	-0.005	0.00002	1.51603 ***	-0.017 **
	(0)	(–)	(0.009)	(0)	(–)	(0.008)
GRID	0.00015	1.64374 ***	0.014	0.00023	1.48297 ***	0.002
	(0)	(–)	(0.01)	(0)	(–)	(0.01)
INRG	-0.00013	2.51176 ***	-0.009	-0.00003	2.21149 ***	-0.023 **
	(0)	(–)	(0.011)	(0)	(–)	(0.01)
ENER	-0.00009	1.741 ***	-0.001	0.00001	1.52839 ***	-0.014 *
	(0)	(–)	(0.01)	(0)	(–)	(0.009)
PBD	0.00002	2.03559 ***	0.003	0.00012	1.78387 ***	-0.011
	(0)	(–)	(0.009)	(0)	(–)	(0.008)
QCLN	0.00019	2.60386 ***	0.014	0.00029	2.28446 ***	0.001
	(0)	(–)	(0.012)	(0)	(–)	(0.011)
SMOG	0.00004	2.52371 ***	0.004	0.00014	2.21171 ***	-0.009
	(0)	(–)	(0.01)	(0)	(–)	(0.009)
TAN	-0.00015	4.70746 ***	-0.011	-0.00007	4.11853 ***	-0.023 *
	(0)	(–)	(0.014)	(0)	(–)	(0.013)

Note: $\alpha = 0.05$, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 11: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_{E}SGF^{\cdot}_{25}$ (January 2007 - October 2021).

$_{E}SGF_{50}^{:}$						
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	-0.00013	4.78522 ***	0.004	-0.00004	3.91692 ***	-0.030 *
	(0)	(–)	(0.017)	(0)	(–)	(0.016)
ERTH	0.00002	2.72605 ***	0.020	0.00002	2.23341 ***	-0.007
	(0)	(–)	(0.014)	(0)	(–)	(0.014)
FAN	-0.00022	2.94499 ***	-0.010	-0.00008	2.46806 ***	-0.031 **
	(0)	(–)	(0.016)	(0)	(–)	(0.015)
GRID	0.00017	2.77829 ***	0.01700	0.0002	2.17782 ***	0.000
	(0)	(–)	(0.018)	(0)	(–)	(0.018)
INRG	-0.00042 *	4.21750 ***	-0.018	-0.00013	3.53615 ***	-0.035 **
	(0)	(–)	(0.017)	(0)	(–)	(0.016)
ENER	-0.00025	2.86791 ***	-0.006	0.00003	2.39006 ***	-0.019
	(0)	(–)	(0.016)	(0)	(–)	(0.015)
PBD	-0.00017	3.45776 ***	0.002	0.00003	2.85369 ***	-0.025 *
	(0)	(–)	(0.015)	(0)	(–)	(0.014)
QCLN	0.00002	4.29810 ***	0.017	0.00018	3.56411 ***	-0.012
	(0)	(–)	(0.017)	(0)	(–)	(0.016)
SMOG	-0.00016	4.27044 ***	0.002	0.00003	3.54389 ***	-0.023
	(0)	(–)	(0.016)	(0)	(–)	(0.015)
TAN	-0.00045	7.79430 ***	-0.018	-0.00025	6.48360 ***	-0.037 **
	(0)	(–)	(0.02)	(0)	(–)	(0.018)

Note: $\alpha = 0.05$, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 12: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_{E}SGF_{50}^{:}$ (January 2007 - October 2021).

	$_ESGF_{75}^{\cdot}$					
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	0.00004	3.07238 ***	-0.008	0.00005	2.88511 ***	-0.009
	(0)	(–)	(0.011)	(0)	(–)	(0.011)
ERTH	0.00019	1.71429 ***	0.009	0.0002	1.60937 ***	0.008
	(0)	(–)	(0.008)	(0)	(–)	(0.007)
FAN	-0.00003	1.84316 ***	-0.01	-0.00003	1.74096 ***	-0.01
	(0)	(–)	(0.009)	(0)	(–)	(0.009)
GRID	0.00019	1.80286 ***	0.009	0.00019	1.75783 ***	0.009
	(0)	(–)	(0.01)	(0)	(–)	(0.01)
INRG	-0.0001	2.69192 ***	-0.014	-0.0001	2.53442 ***	-0.014
	(0)	(–)	(0.011)	(0)	(–)	(0.011)
ENER	-0.00006	1.86238 ***	-0.006	-0.00006	1.75433 ***	-0.005
	(0)	(–)	(0.009)	(0)	(–)	(0.009)
PBD	0.00006	2.17904 ***	-0.003	0.00006	2.04899 ***	-0.003
	(0)	(–)	(0.009)	(0)	(–)	(0.008)
QCLN	0.00022	2.79132 ***	0.009	0.00023	2.62391 ***	0.009
	(0)	(–)	(0.011)	(0)	(–)	(0.011)
SMOG	0.00007	2.70106 ***	-0.001	0.00008	2.54198 ***	-0.002
	(0)	(–)	(0.01)	(0)	(–)	(0.009)
TAN	-0.00012	5.01692 ***	-0.016	-0.00012	4.73632 ***	-0.016
	(0)	(–)	(0.014)	(0)	(–)	(0.013)

Note: $\alpha = 0.05$, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 13: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_ESGF_{75}^{\cdot}$ (January 2007 - October 2021).

$_{G}SGF^{;}_{25}$						
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	-0.00012	5.69834 ***	-0.005	-0.00014	5.54478 ***	-0.021
	(0)	(–)	(0.017)	(0)	(–)	(0.018)
ERTH	0.00008	3.17972 ***	0.014	0.00011	3.07941 ***	0.002
	(0)	(–)	(0.014)	(0)	(–)	(0.015)
FAN	-0.00018	3.45416 ***	-0.011	-0.00013	3.467 ***	-0.020
	(0)	(–)	(0.016)	(0)	(–)	(0.017)
GRID	0.00026	3.33185 ***	-0.002	0.00010	3.37335 ***	0.006
	(0)	(–)	(0.018)	(0)	(–)	(0.019)
INRG	-0.00031	4.95652 ***	-0.018	-0.00031	4.85546 ***	-0.032 *
	(0)	(–)	(0.016)	(0)	(–)	(0.017)
ENER	-0.00017	3.34108 ***	-0.005	-0.00012	3.24102 ***	-0.016
	(0)	(–)	(0.016)	(0)	(–)	(0.017)
PBD	-0.00009	4.05401 ***	-0.001	-0.00010	3.97973 ***	-0.017
	(0)	(–)	(0.015)	(0)	(–)	(0.016)
QCLN	0.00003	5.11834 ***	0.007	0	4.92888 ***	-0.010
	(0)	(–)	(0.017)	(0)	(–)	(0.018)
SMOG	-0.00009	5.01579 ***	-0.001	-0.00009	4.87131 ***	-0.015
	(0)	(–)	(0.016)	(0)	(–)	(0.017)
TAN	-0.00041	9.26281 ***	-0.021	-0.00041	9.21505 ***	-0.033
	(0)	(–)	(0.021)	(0)	(–)	(0.021)

Note: $\alpha = 0.05$, $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Table 14: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_{G}SGF^{;}_{25}$ (January 2007 - October 2021).

$_{G}SGF_{50}^{:}$						
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	0.00005	3.31737 ***	-0.010	0.00004	3.01829 ***	-0.008
	(0)	(–)	(0.011)	(0)	(–)	(0.011)
ERTH	0.00020	1.85056 ***	0.007	0.00019	1.68365 ***	0.009
	(0)	(–)	(0.007)	(0)	(–)	(0.007)
FAN	-0.00001	1.98372 ***	-0.013	-0.00003	1.81959 ***	-0.010
	(0)	(–)	(0.009)	(0)	(–)	(0.009)
GRID	0.00020	1.8968 ***	0.006	0.00019	1.77012 ***	0.008
	(0)	(–)	(0.01)	(0)	(–)	(0.01)
INRG	-0.00008	2.90942 ***	-0.016	-0.00011	2.65073 ***	-0.012
	(0)	(–)	(0.01)	(0)	(–)	(0.01)
ENER	-0.00005	2.01027 ***	-0.007	-0.00007	1.83134 ***	-0.004
	(0)	(–)	(0.009)	(0)	(–)	(0.009)
PBD	0.00007	2.35077 ***	-0.005	0.00005	2.14005 ***	-0.002
	(0)	(–)	(0.008)	(0)	(–)	(0.008)
QCLN	0.00023	3.01329 ***	0.00700	0.00022	2.74293 ***	0.009
	(0)	(–)	(0.011)	(0)	(–)	(0.011)
SMOG	0.00009	2.91536 ***	-0.003	0.00007	2.65523 ***	-0.001
	(0)	(–)	(0.009)	(0)	(–)	(0.009)
TAN	-0.00011	5.40146 ***	-0.019	-0.00012	4.951 ***	-0.017
	(0)	(–)	(0.013)	(0)	(–)	(0.013)

Note: $\alpha = 0.05$, $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Table 15: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_{G}SGF_{50}^{:}$ (January 2007 - October 2021).

$_{G}SGF_{75}^{\cdot}$						
	ES			CO_2S		
ETF	t-test	F-test	HAC-test	t-test	F-test	HAC-test
PBW	-0.00022	6.97544 ***	-0.010	-0.00021	5.78297 ***	-0.013
	(0)	(–)	(0.019)	(0)	(–)	(0.019)
ERTH	0.00003	3.81678 ***	0.014	0.00003	3.20919 ***	0.009
	(0)	(–)	(0.016)	(0)	(–)	(0.016)
FAN	-0.00024	4.36905 ***	-0.015	-0.00022	3.61576 ***	-0.015
	(0)	(–)	(0.017)	(0)	(–)	(0.017)
GRID	0.00006	4.12104 ***	-0.003	0.00006	3.54349 ***	0.007
	(0)	(–)	(0.02)	(0)	(–)	(0.02)
INRG	-0.00046 *	6.24199 ***	-0.027	-0.00046 *	5.16967 ***	-0.031 *
	(0)	(–)	(0.018)	(0)	(–)	(0.018)
ENER	-0.00032	4.24211 ***	-0.01800	-0.00022	3.44827 ***	-0.013
	(0)	(–)	(0.017)	(0)	(–)	(0.017)
PBD	-0.00020	5.08525 ***	-0.007	-2e-04	4.16974 ***	-0.013
	(0)	(–)	(0.017)	(0)	(–)	(0.017)
QCLN	-0.00009	6.15477 ***	0.000	-0.00011	5.15119 ***	-0.007
	(0)	(–)	(0.019)	(0)	(–)	(0.019)
SMOG	-0.00020	6.23022 ***	-0.006	-0.00019	5.10844 ***	-0.010
	(0)	(–)	(0.017)	(0)	(–)	(0.018)
TAN	-0.00051	11.54572 ***	-0.026	-0.00055	9.42908 ***	-0.031
	(0)	(–)	(0.022)	(0)	(–)	(0.022)

Note: $\alpha = 0.05$, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 16: Results of the tests, statistics and standard errors (in brackets), conducted on the differences in financial performances between the sampled green ETFs and the $_{G}SGF_{75}^{\cdot}$ (January 2007 - October 2021).

References

Alexopoulos, T. A. (2018). To trust or not to trust? a comparative study of conventional and clean energy exchange-traded funds. *Energy Economics*, 72:97–107.

Alexopoulos, T. A. and Thomakos, D. D. (2016). Functional smoothing for risk management of energy assets. *International Journal of Energy and Statistics*, 4(02):1650006.

Andersson, M., Bolton, P., and Samama, F. (2016). Hedging climate risk. *Financial Analysts Journal*, 72(3):13–32.

Angelakoglou, K. and Gaidajis, G. (2015). A review of methods contributing to the assessment of the environmental sustainability of industrial systems. *Journal of Cleaner Production*, 108:725–747.

Avramov, D., Cheng, S., Lioui, A., and Tarelli, A. (2021). Sustainable investing with esg rating uncertainty. *Journal of Financial Economics*.

Bender, J., Bridges, T. A., and Shah, K. (2019). Reinventing climate investing: building equity portfolios for climate risk mitigation and adaptation. *Journal of Sustainable Finance & Investment*, 9(3):191–213.

Berg, F., Koelbel, J. F., and Rigobon, R. (2019). *Aggregate confusion: The divergence of ESG ratings*. MIT Sloan School of Management.

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*.

Brandon, R. G., Krueger, P., Riand, N., and Schmidt, P. S. (2019). ESG rating disagreement and stock returns. *Research Paper*, (19-67).

Busch, T., Johnson, M., and Pioch, T. (2020). Corporate carbon performance data: Quo vadis? *Journal of Industrial Ecology*.

Busch, T. and Lewandowski, S. (2016). Corporate carbon and financial performance: A meta-analysis. In *Academy of Management Proceedings*, volume 2016, page 11657. Academy of Management Briarcliff Manor, NY 10510.

Cabello, J. M., Ruiz, F., Pérez-Gladish, B., and Méndez-Rodríguez, P. (2014). Synthetic indicators of mutual funds' environmental responsibility: An application of the reference point method. *European Journal of Operational Research*, 236(1):313–325.

Capasso, G., Gianfrate, G., and Spinelli, M. (2020). Climate change and credit risk. *Journal of Cleaner Production*, 266:121634.

Capelle-Blancard, G. and Monjon, S. (2012). Trends in the literature on socially responsible investment: Looking for the keys under the lamppost. *Business Ethics: a European Review*, 21(3):239–250.

Coeslier, M., Louche, C., and Hétet, J.-F. (2016). On the relevance of low-carbon stock indices to tackle climate change. *Journal of Sustainable Finance & Investment*, 6(4):247–262.

Delmas, M. A., Nairn-Birch, N., and Lim, J. (2015). Dynamics of environmental and financial performance: The case of greenhouse gas emissions. *Organization & Environment*, 28(4):374–393.

Diaz-Rainey, I., Robertson, B., and Wilson, C. (2017). Stranded research? leading finance journals are silent on climate change. *Climatic Change*, 143(1):243–260.

Dutta, A., Bouri, E., Saeed, T., and Vo, X. V. (2020). Impact of energy sector volatility on clean energy assets. *Energy*, 212:118657.

Escrig-Olmedo, E., Fernández-Izquierdo, M. Á., Ferrero-Ferrero, I., Rivera-Lirio, J. M., and Muñoz-Torres, M. J. (2019). Rating the raters: Evaluating how esg rating agencies integrate sustainability principles. *Sustainability*, 11(3):915.

European Commission (2020). EU taxonomy for sustainable activities. <<ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en>>.

European Securities and Markets Authority (2021). Final Report - Advice on Article 8 of the Taxonomy Regulation. <<ec.europa.eu/info/business-economy-euro/banking-and-finance/sustainable-finance/eu-taxonomy-sustainable-activities_en>>.

Fahmy, H. (2021). The rise in investors' awareness of climate risks after the Paris Agreement and the clean energy-oil-technology prices nexus. *Energy Economics*, page 105738.

Fahmy, H. (2022). Clean energy deserves to be an asset class: A volatility-reward analysis. *Economic Modelling*, 106:105696.

Friede, G., Busch, T., and Bassen, A. (2015). Esg and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4):210–233.

Garvey, G. T., Iyer, M., and Nash, J. (2018). Carbon footprint and productivity: does the "E" in ESG capture efficiency as well as environment. *Journal of Investment Management*, 16(1):59–69.

Halbritter, G. and Dorfleitner, G. (2015). The wages of social responsibility - where are they? a critical review of esg investing. *Review of Financial Economics*, 26:25–35.

Henriques, C. O., Neves, M. E., Castelão, L., and Nguyen, D. K. (2022). Assessing the performance of exchange traded funds in the energy sector: a hybrid dea multiobjective linear programming approach. *Annals of Operations Research*, pages 1–26.

Huang, J., Xiang, S., Wang, Y., and Chen, X. (2021). Energy-saving r&d and carbon intensity in china. *Energy Economics*, 98:105240.

Jeswiet, J. and Nava, P. (2009). Applying ces to assembly and comparing carbon footprints. *International Journal of Sustainable Engineering*, 2(4):232–240.

Johnson, E. (2009). Goodbye to carbon neutral: Getting biomass footprints right. *Environmental impact assessment review*, 29(3):165–168.

Lantushenko, V., Schellhorn, C., and Zaynutdinova, G. R. (2021). Climate change concerns meet return-chasing: Evidence from energy exchange-traded funds. *The Financial Review*.

Malinda, M. and Hui, C. J. (2016). The study of the long memory in volatility of renewable energy exchange-traded funds (etfs). *Journal of Economics, Business and Management*, 4(4):252–257.

Marszk, A. (2019). Sustainable investing exchange-traded funds: Us and european market. *Journal of Economics & Management*, 37:69–86.

Matsumura, E. M., Prakash, R., and Vera-Munoz, S. C. (2014). Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Review*, 89(2):695–724.

Morioka, S. N. and de Carvalho, M. M. (2016). A systematic literature review towards a conceptual framework for integrating sustainability performance into business. *Journal of Cleaner Production*, 136:134–146.

Morningstar (2018). Sustainable Funds U.S. Landscape Report 2018. <<morningstar.com/insights/2019/02/19/esg-landscape>>.

Papailias, F. and Thomakos, D. D. (2013). Trading energy etfs with an improved moving average strategy. *International Journal of Energy and Statistics*, 1(01):31–43.

Petrillo, A., De Felice, F., García-Melón, M., and Pérez-Gladish, B. (2016). Investing in socially responsible mutual funds: Proposal of non-financial ranking in italian market. *Research in International Business and Finance*, 37:541–555.

Popescu, I.-S., Hitaj, C., and Benetto, E. (2021). Measuring the sustainability of investment funds: A critical review of methods and frameworks in sustainable finance. *Journal of Cleaner Production*, page 128016.

Statman, M. and Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4):33–46.

Thomä, J., Dupré, S., and Hayne, M. (2018). A taxonomy of climate accounting principles for financial portfolios. *Sustainability*, 10(2):328.

Thomakos, D. D. and Papailias, F. (2013). Momentum trading in new york stock exchange (nyse) energy stocks. *International Journal of Energy and Statistics*, 1(04):243–256.

van Dijk-de Groot, M. and Nijhof, A. H. (2015). Socially responsible investment funds: a review of research priorities and strategic options. *Journal of Sustainable Finance & Investment*, 5(3):178–204.

Watkins, J. and Durning, B. (2012). Carbon definitions and typologies in environmental impact assessment: greenhouse gas confusion? *Impact Assessment and Project Appraisal*, 30(4):296–301.