Tracking-based green portfolio optimization
PRELIMINARY DRAFT

Diana Barro* d.barro@unive.it
Marco Corazza* corazza@unive.it
Gianni Filograsso* gianni.filograsso@unive.it

August 26, 2024

Abstract

In this contribution, we discuss how to handle financial and sustainable investment goals, focusing on greenness and ESG features. Sustainable investing has attracted increasing interest with an associated growing commitment to take an active part in investment choices. Among thematic investments, green and energy-related ones have emerged, capturing investors' attention. Non-optimized strategies and traditional portfolio allocation models cannot guarantee the necessary flexibility. To answer this demand, ESG tailored-made allocations should be provided, with the aim of representing the preferences and commitments of investors adequately. This contribution introduces a novel ESG-focused tracking error model to optimize portfolio allocation. We consider two reference benchmarks, accounting for a financial target and an ESG one, respectively. The objective function results in a convex linear combination of the two goals where the combining parameter accounts for the investor's financial and ESG preferences. A symmetric tracking error measure is proposed to replicate the financial benchmark passively, while an asymmetric measure is used to track and possibly outperform the thematic ESG benchmark. Identifying the benchmarks for the two components represents a crucial step and, jointly with the choice of the combining parameter, accounts for the portfolio's overall risk-return and ESG profiles. In the model, the sustainability feature is handled not only with the presence of the ESG benchmark but also with the introduction of dedicated constraints. Namely, a desired minimum level of greenness and a maximum amount of carbon intensity can be accounted for. An application to the EUROSTOXX 600 equity market is presented and discussed for different choices of the combining parameter, representing different sustainability preferences and risk-return profiles. Furthermore, a discussion on the choice of the benchmarks is provided.

Keywords: Tracking Error, Portfolio optimization, Green sustainability, ESG, EUROSTOXX 600

---

*Dept. of Economics, Ca' Foscari University of Venice

1 Introduction

Sustainability-themed funds seek to capture long-term trends that have the potential to reshape structurally the global economy, by shifting the priorities of societies, such as climate change, circular economy and biodiversity. According to a recently published survey (Eurosif, 2023), asset managers implementing sustainability in their portfolios still rely heavily on positive and negative screening strategies, but a non-negligible 55% of respondents implement green investment policies through a thematic approach. In this way, portfolio managers can intercept the evolving investors' taste for green securities w.r.t. a specific topic, although capturing preferences for sustainability is not straightforward, as the attitude of investors towards green products is dynamic and heterogeneous (Assaf et al., 2024). For example, building upon the seminal contribution of Pástor et al. (2021), in which investors are allowed to have different green tastes, Rojo-Suárez and Alonso-Conde (2024) consider shifts in green tastes, and find that ESG preferences are increasingly taken into account, pushing the greenness level of the market portfolio. Moreover, different investors have different green preferences: for instance, on the basis of survey data, Bauer et al. (2021) document that pension funds are specifically focused on the United Nation Sustainable development goals.

Furthermore, from the perspective of a portfolio manager, difficulties in capturing heterogeneous green preferences are compounded by further complications, i.e. their measurement and integration into portfolio models (Billio et al., 2021).

In general, as stated above, since the pioneering contribution of Cox et al. (2004), evidence in the literature has shown that the standard practice for implementing sustainability in portfolio management consists in placing exclusion filters, despite being discouraged for being an inefficient practice (Utz et al., 2015). Sustainability-themed portfolios, either discretionary or systematic, are designed to capture long-term trends by getting exposure to theme-related activities, through the identification of an investment cluster. The portfolio construction process is where differences lie and optimized proposals have an edge, due to higher versatility in the definition of the desired allocation (Methling and von Nitzsch, 2019).

Several attempts to incorporate green preferences into the (optimized) portfolio construction process have been made. Utz et al. (2015) incorporate ESG scores into the optimization model and find that, based on expected returns and variances estimates of sustainable mutual funds, the projected tricriterion efficient surface can yield much better results from a sustainability perspective, without affecting the financial performance of such funds. Escrig-Olmedo et al. (2017) address the challenge of integrating the preference of heterogeneous investors by developing a fuzzy multicriteria approach for decision making. Moreover, the implications of green preferences for the shape of investors' efficient frontier are similarly deemed critical and have been extensively considered in the literature (Pedersen et al., 2021; Steuer and Utz, 2023), where the problems are respectively characterized as a bicriterion maximum Sharpe Ratio-ESG and a tricriterion one. Differently, Cesarone et al. (2022) characterize the problem through the ϵ-constraint method, transforming the ESG tri-objective problem into a single-objective one, while Xidonas and Essner (2022) formulate the problem as a multi-objective minimax and incorporate the controversy dimension into the portfolio optimization model. Another natural way to proceed is to maximize the ESG quality of the portfolio, while keeping its tracking error and turnover below the desired thresholds, along with constraints capping the exposure to regions, sectors, risk factors and individual stocks (Alessandrini and Jondeau, 2021).

Green portfolio tilts might also be based on the portfolio carbon footprint: several contributions in the literature have shown that portfolio carbon emissions can be drastically reduced while managing simultaneously the tracking error, which can be kept at reasonable levels, even when the portfolio is aimed at reaching the net zero target by 2050 (Andersson et al., 2016; Bolton and Kacperczyk, 2021; Le Guenedal and Roncalli, 2022).

Among thematic optimization approaches, a possible strategy could consist in the separate optimization of a conventional portfolio targeting financial criteria, and a stand-alone thematic portfolio focusing on other criteria, which is also referred to as core-satellite approach. More recently, a formulation controlling for inter-portfolio correlation effects has been proposed (Methling and von Nitzsch, 2019), allowing to deal flexibly with any characterization of green preferences in the form of a themed portfolio, including e.g. carbon emissions, ESG scores or renewable energy generation.

We are interested in modeling the optimal portfolio allocations so that they account for the decision maker preferences expressed as goals with respect to financial and thematic performances. In this contribution, we focus on an holistic approach that jointly optimizes the goals and, through a parametric approach, allows for the flexibility to handle a spectrum of preferences. We assume that the decision maker set goals with reference to benchmarks that are used as representative risk-return profiles. In more detail, we consider a benchmark, that we identify as *broad market*, with the aim of accounting for diversification opportunities across sectors and assets and for the financial performance goal, and a *thematic benchmark* that accounts for the environmental performance goal. Furthermore, through the introduction of constraints the decision maker can enforce specific thematic requirements, such as, for example, limits to carbon emissions. The resulting model is a double tracking-error problem formulated as a single-period stochastic programming problem, where the uncertainty faced by the decision maker is modeled through a scenario approach. To generate the scenarios we rely on deep learning techniques, specifically a Generative Adversarial Network (GAN).

The rest of the paper is organized as follows. In Section 2 we present the model, in Section 3 we discuss its application to data from the European stock market, and finally in Section 4 we present the conclusions and future research directions, Appendix A presents the scenario generation method used.

2 A scenario-based double tracking error model

Let us consider an investor that is interested in pursuing financial and environmental sustainability goals, simultaneously. To integrate them into the portfolio construction process we rely on a double tracking error formulation of the problem. To this aim, two benchmarks are identified as reference targets to be tracked, i.e. a broad market benchmark that can account for risk/return preferences whilst a green thematic benchmark that allows to emphasize the environmental performance.

Let z and w be the broad market benchmark and the thematic benchmark, respectively. Furthermore, let $x = (x_1, \ldots, x_I)$ be the vector of assets used to track the broad market, and $y = (y_1, \ldots, y_{I_E})$ the vector of assets used to compose the portfolio to track the thematic benchmark, where with I and I_E we denote the set of assets used to track each corresponding benchmark. These two set are meant to be general and may or may not present intersections. Let ϕ^M and ϕ^G be the distance measures used to track the market and green benchmarks, respectively. As distance measures to quantify the tracking errors we resort to the Mean Absolute Deviation (MAD) and its asymmetric version, the Mean Absolute Downside Deviation (MADD). These choices allow us to consider both a passive replication strategy (symmetric distance) and a more flexible active one (asymmetric distance).

In this contribution we consider a static setting, and formulate a single period scenario-based portfolio selection problem (Dembo and Rosen, 1999; Gilli et al., 2019).

Let $s = (1, \ldots, S)$ denote a finite set of scenarios and π_s their probabilities of occurrence. With $x_{i,s}$ we denote the weight of stock $i \in I$ in scenario s, and, similarly, $y_{j,s}$ is the weight of stock $j \in I_E$. Furthermore, z_s and w_s are the returns of the benchmarks in scenario s, and r_s is the vector of returns in scenario s, for assets in I and I_E.

Two goals are identified for our problem: the first measure aims at tracking passively the main index, based on the MAD distance

$$\phi_s^M = \sum_{i=1}^{I} |r_{i,s} x_{i,s} - z_s|$$

while, for the second goal we introduce an asymmetric tracking formulation based on the MADD distance

$$\phi_s^G = \sum_{i=1}^{I_E} |r_{j,s} y_{j,s} - z_s|^-$$

.

As for the constraints, we assume that the tracking portfolios are managed separately and are kept together only by an overall budget constraint. We do not consider transaction costs and we assume a long-only formulation.

The environmental goal performance can be enhanced including an additional carbon intensity budget for the thematic portfolio. The modelization of green constraints is key, since it allows to pursue different investment policy. For instance, a decarbonization pathway can be introduced by imposing at every reallocation date a stricter constraint on portfolio emissions (Le Guenedal and Roncalli, 2022). Furthermore, a wide array of filters can be applied to remove industries that are deemed undesirable, or by filtering out firms with low ESG scores (Alessandrini and Jondeau, 2021). In the proposed formulation, we require the carbon budget to be lower than the average of the green benchmark (Le Guenedal et al., 2022). A decarbonization rate η, is set to 0.05, with the carbon intensity denoted as CI.

Moreover, a combining parameter, $\lambda \in [0, 1]$, is introduced to deal with different preferences for the green thematic component. Note that $\lambda = 1$ denotes respectively the case of a passive tracking portfolio, that reflects the preferences of an unaware investor. Whilst, $\lambda = 0$ identifies an investor entirely focused on the thematic performance of the asymmetric tracking portfolio, allowing for a more active management of the green portfolio.

The resulting double-tracking portfolio optimization problem is:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \quad & \sum_{s=1}^{S} \pi_s \left[\lambda\, \phi_s^M + (1-\lambda)\, \phi_s^G \right] \\
& \text{s.t.} \sum_{j=1}^{I_E} CI_j y_j < (1-\eta)\overline{CI} \\
& \sum_{i=1}^{I} x_i + \sum_{j=1}^{I_E} y_j = 1 \\
& x_i \geqslant 0 \quad i = 1, \dots, I \qquad y_j \geqslant 0 \quad j = 1, \dots, I_E
\end{aligned}
\tag{1}
$$

In particular, the objective function tracks the financial market by means of a symmetric deviation measure, and captures deviations from a thematic green index through an asymmetric measure, possibly aiming to outperform it. The problem is formulated through two separate subset of decision variables, for the financial and the thematic components. An additional constraint on the thematic component sets a cap on the allowed carbon intensity of the portfolio. Different preferences for financial performances and for greenness are managed by a combining parameter. To assess the proposed model over a longer horizon we consider a monthly reallocation. The chosen distance measures allow the problem to be formulated as a linear stochastic programming problem (Konno and Yamazaki, 1991):

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \quad & \sum_{s=1}^{S} \pi_s \left[\lambda(\alpha_1 \theta_s^+ + \alpha_2 \theta_s^-) + (1-\lambda) \cdot (\beta_1 \gamma_s^+ + \beta_2 \gamma_s^-) \right] \\
& \text{s.t.} \quad \theta^+ - \theta^- = \sum_{i=1}^{I} r_{i,s} x_{i,s} - z_s \\
& \qquad \gamma^+ - \gamma^- = \sum_{j=1}^{I_E} r_{j,s} y_{j,s} - w_s \\
& \qquad \sum_{j=1}^{I_E} CI_j y_j < (1-\eta)\overline{CI} \\
& \qquad \sum_{i=1}^{I} x_i + \sum_{j=1}^{I_E} y_j = 1 \\
& \qquad \theta_s^+, \theta_s^-, x_i, y_j \geqslant 0
\end{aligned} \tag{2}
$$

Among the many available scenario generation techniques, we rely on a generative adversarial network (GAN) approach (Goodfellow et al., 2014) to construct the scenario tree, in particular a Wasserstein-Gradient Penalty GAN (WGAN-GP) (Arjovsky et al., 2017; Gulrajani et al., 2017; Silva and de Almeida Filho, 2023); GANs constitute a recent emerging trend in the literature, whose effectiveness for the generation of financial time series has been previously highlighted (Dahl and Sørensen, 2022). The methodology and the network architecture are briefly presented in Appendix A. Among the positive properties, we mention the capability of handling several stylized facts in financial markets, including temporal dependencies and fat tails of distributions, whereas among the few cons' of such approach, we mention hyperparameter sensitivity and computational costs (Takahashi et al., 2019; Koshiyama et al., 2021).

To assess our proposal, that we refer to as 'market-thematic' portfolio, assuming $\lambda = 0.5$, we compare it with three other formulations of the problem, each dealing with a specific dimension. In particular we consider two special cases of Problem 2, with $\lambda = 1$ denoting the market portfolio replicating the benchmark z_s, that will be called henceforth the 'market' portfolio. No constraints on carbon intensity are applied. Then, with $\lambda = 0$ we obtain a solution of the problem that will be referred to as the thematic portfolio, with a carbon intensity budget over the period, aiming to replicate the Environmental benchmark. Finally, a third formulation with $\lambda = 1$ and the carbon intensity constraint is considered, so as to 'mimic' the performance of a standard tracking portfolio screening out top emitters, that will be defined 'market-decarbonized'.

In this way, we can compare our proposal with several existing formulations in the practitioner-oriented literature. First, the market-decarbonized formulation is equivalent to integrating sustainability standards through the definition of constraints on the portfolio GHG intensity or E score, which is a widely documented standard practice (Alessandrini and Jondeau, 2021; De Spiegeleer et al., 2023). Second, the thematic portfolio allows us to test a formulation that exclusively focuses on an investment theme, which is partly closer to idea recently challenged by (Methling and von Nitzsch, 2020) of constructing conventional and themed portfolios separately. Our proposal departs from this perspective, since we consider both components jointly, allowing to control for both the financial and green dimensions, and to interpret other widespread existing formulations as a special case of our model.

Finally, formulations with different constraints on environmental performance and carbon emissions are tested to verify whether there is a relationship between the two, and to assess whether additional, stricter constraints on the feasible region have an actual impact on portfolios. In particular, w.r.t. formulation 2, we also consider an additional constraint where the portfolio environmental score is higher than the average of the green index's score:

$$\sum_{j=1}^{I_E} E_j y_j \geqslant \bar{E}$$

3 Data and results

In this section, we present the results, w.r.t. the in-sample and out of-sample performance of the portfolios. In Section 3.1 we briefly present the data, while in Section 3.2 we discuss the results relative to different portfolio models.

3.1 Data and experimental setting

We test our model on the four formulations presented in Section 2 on a dataset of European stock prices. In particular, we perform experiments based on 600 stocks from the EUROSTOXX 600 index, along with the index itself and its green counterpart, the STOXX Europe Environmental leaders, which is composed of 169 stocks, from January 31, 2015 to December 29, 2023. The former covers large, mid and small capitalization companies among 16 European countries, representing more than 90% of the market capitalization of European stock markets. The countries that make up the index are the United Kingdom (composing around 23.3% of the stocks), France (composing around 12.6% of the stocks), Germany (composing around 11.8% of the stocks), as well as Austria, Belgium, Denmark, Finland, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Spain, Sweden, Switzerland. As for the sectors, the Industrials, the Financials and the Consumer Discretionary sectors represent respectively about 24.2%, 19.8%, and the 10.6% of the overall sample. The Environmental index is constructed by filtering out companies based on involvement in incidents with negative environmental implications, or belonging to controversial sectors. Compared to the main index, the composition is not significantly altered, with the United Kingdom, France and Germany representing respectively about 18.9%, 13.0% and 10.6% of the sample, and the Financials and the

	Mean	Std	Skewness	Kurtosis	Normality	#
EUROSTOXX 600	0.0025	0.0425	-0.4817	4.4050	<0.01	600
STOXX Europe Env. Leaders	0.0034	0.0455	-0.5737	5.3761	<0.001	169

Table 1: Descriptive statistics of the benchmarks. The normality column corresponds to the outcome of a Jarque Bera test.

Industrials representing than 54% of the firms in the index. The GHG intensity[1], Environmental and ESG scores are retrieved from Bloomberg. Summary statistics of the two benchmarks are reported in Table 1: note that normality assumption does not hold, the distributions are negatively skewed and have slightly fatter tails compared to a Normal distribution.

The rolling analysis is performed on a PC with an AMD Ryzen 7 4700U (2.0 GHz) CPU and 8 GB of RAM, using MATLAB R2022a. We present the tracking results for the 12-months period from January 31, 2023 to December 29, 2023. The corresponding bootstrap period corresponds to an expanding window ranging from January 31, 2015 to December, 29 2023, where the additional (realized) information is periodically included at the end of each time step. All the tests have been performed by generating 200 scenarios at each step. To ensure that both symmetric and asymmetric deviations are taken into account, we set respectively $\alpha_1 = 1$, $\alpha_2 = 1$, $\beta_1 = 0$, $\beta_2 = 1$.

3.2 Results

Let us now outline the results of our computational tests: in the following we break down the portfolio composition and we highlight how the two components of the portfolio interact. Moreover, we shed light on both in-sample and out-of-sample performance of our proposal, compared to baseline strategies that serve as benchmark.

In Table 2 we break down the weights of the two portfolio components over the 12 periods, in four parts. We consider first the non-overlapping parts of the two components, i.e. the share of assets exclusively held in either the market or the thematic portfolio. Then, we take into account the overlapping portion for both, i.e. the share of assets that are held in both portfolios, separately attributed to each component.

An interesting pattern we observe is that, despite the reference universe for the thematic portfolio being a subset of the main one, there is limited overlapping of the two, always below 10% (Figure 1). Moreover, the overlapping portion of the portfolio components is delimited to a limited subset of sectors.

A careful analysis by sector and by country of the market-thematic portfolios (Figures 2 and 3) shows that the two strategies do not have exceptionally different geographic and sectoral allocations,

---

[1] The GHG intensity is computed by dividing the location-based GHG Scope 1 and 2 emissions by the annual turnover in million dollars.

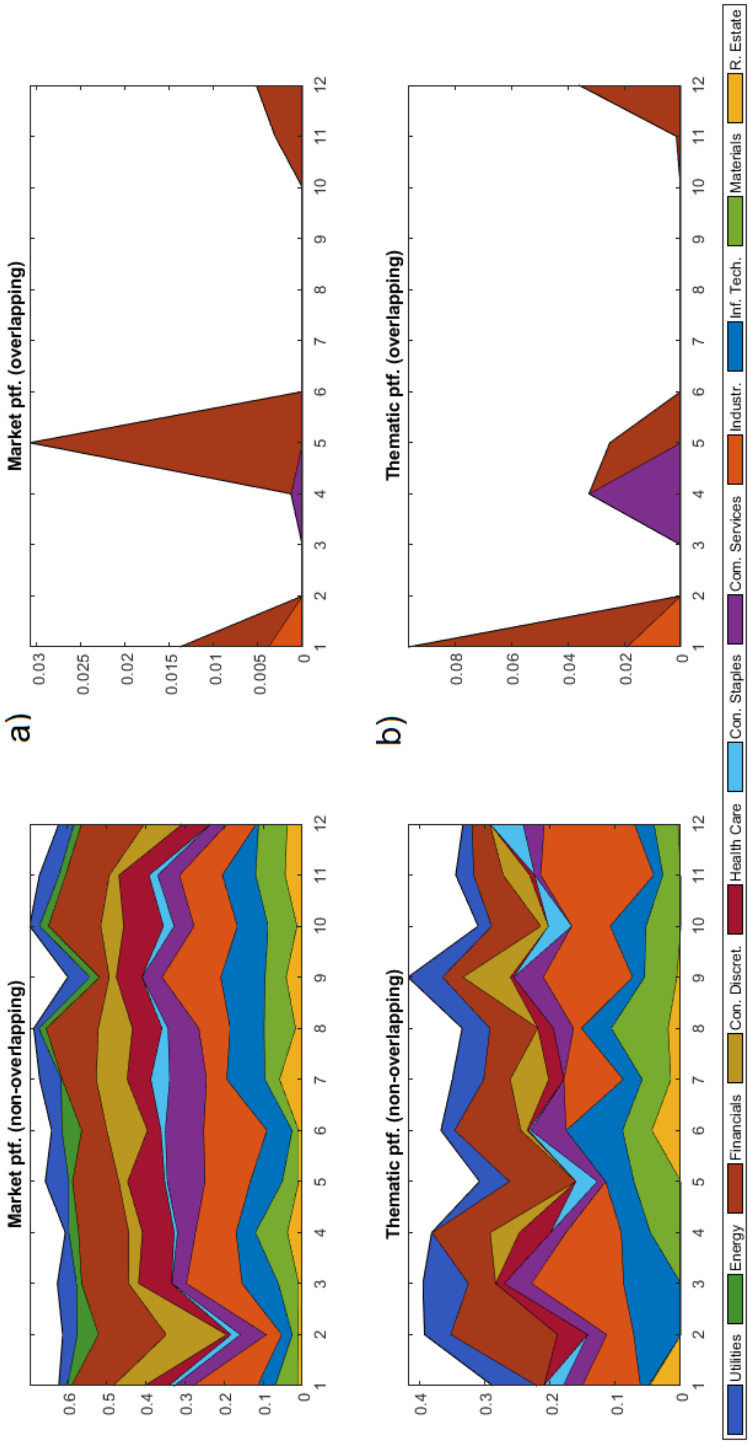

Figure 1: Overlapping and non-overlapping portfolio partitions, defined as the shared investments within the market (upper panel a) and thematic (lower panel b) portfolio components.

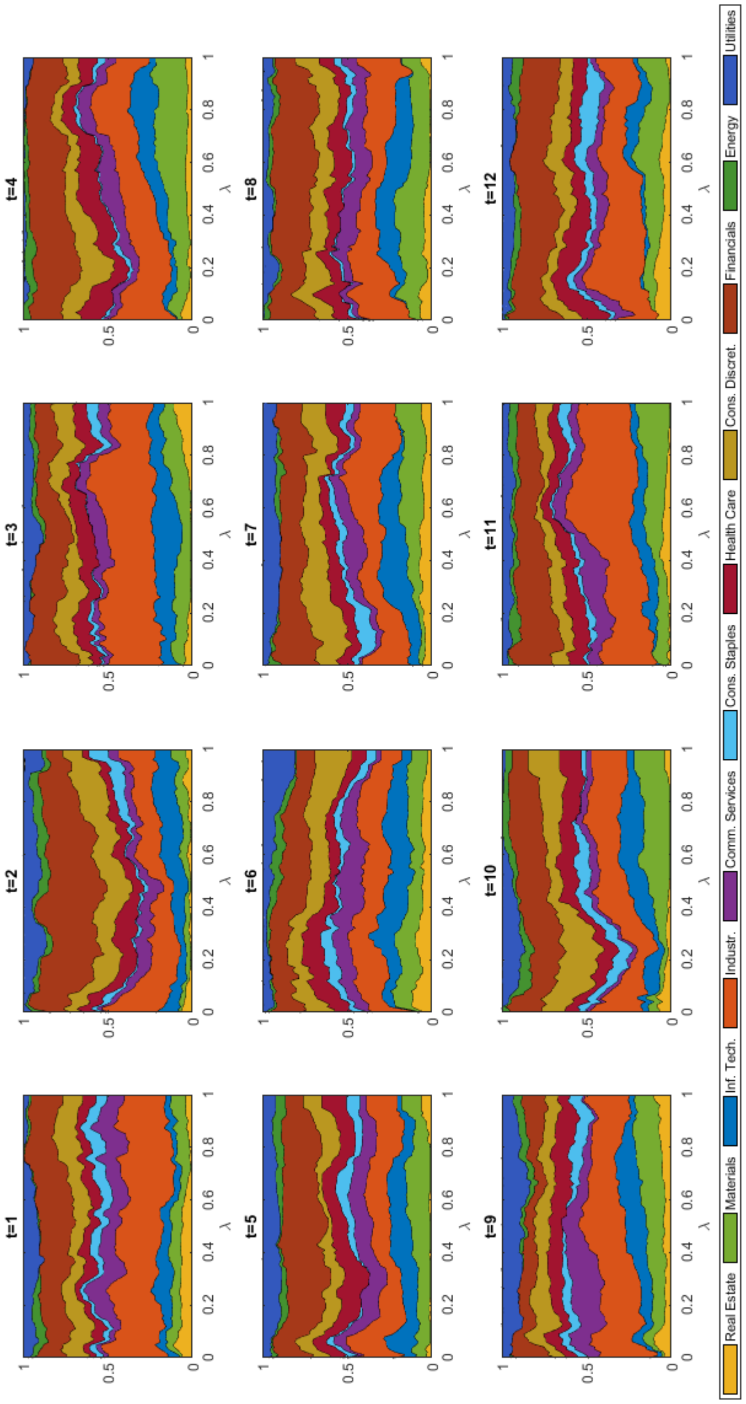

Figure 2: Portfolio breakdown by sector over different windows, over different preferences for green assets $\lambda \in [0, 1]$

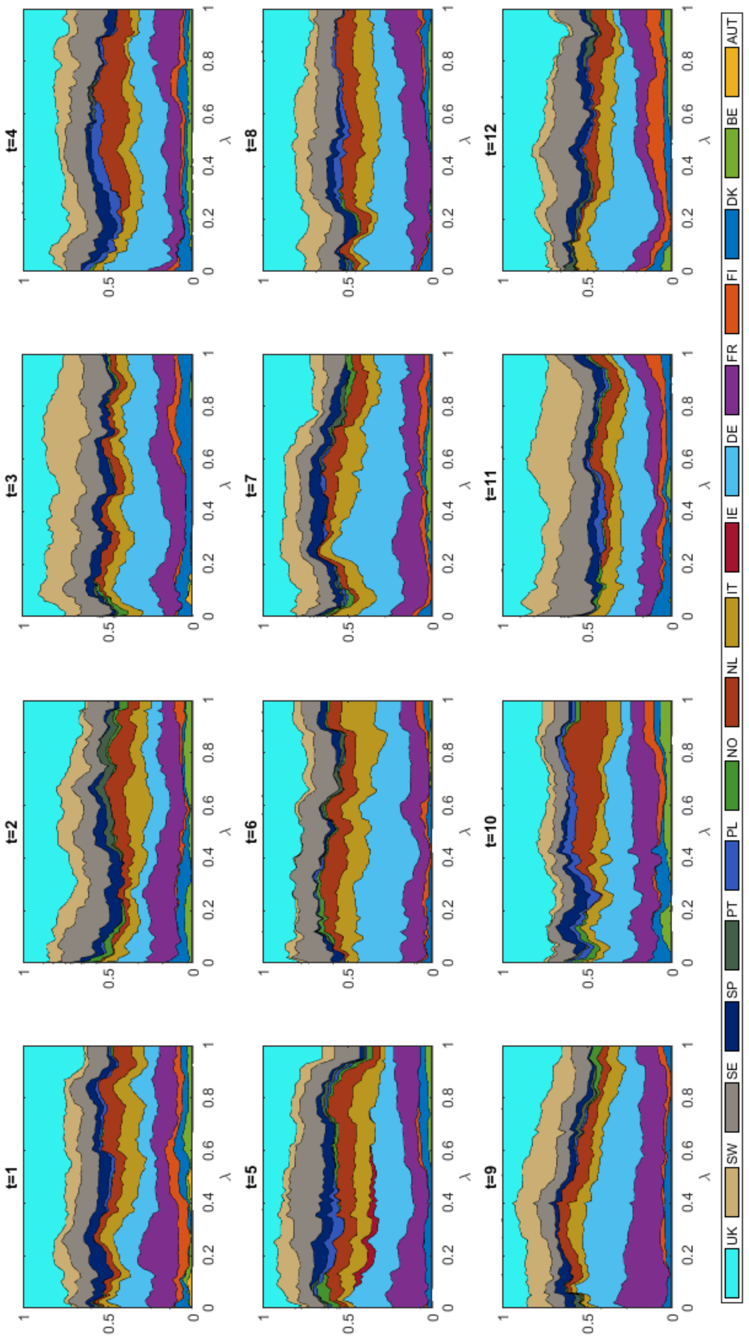

Figure 3: Portfolio breakdown by country over different windows, over different preferences for green assets $\lambda \in [0, 1]$

Weight at time t	1	2	3	4	5	6	7	8	9	10	11	12
Non-overlapping market	0.614	0.612	0.611	0.601	0.656	0.640	0.661	0.666	0.598	0.696	0.661	0.627
Non-overlapping thematic	0.285	0.388	0.389	0.365	0.307	0.360	0.339	0.324	0.412	0.304	0.335	0.332
Overlapping market	0.013	0.000	0.000	0.001	0.030	0.000	0.000	0.000	0.000	0.000	0.003	0.005
Overlapping thematic	0.088	0.000	0.000	0.032	0.025	0.000	0.000	0.000	0.000	0.000	0.001	0.036

Table 2: Portfolio composition for the market and thematic components, based on out-of-sample optimization. Each column, by definition, sums up to one.

despite a few sectors not being included in either the passive or the exclusively thematic version of the portfolio. Note in particular that we let the combining parameter λ vary, to attribute different importance to the two objectives of the problem. The position in the Materials sector tends to be sizeable for $\lambda = 1$ (market portfolio), while assumes less importance in the composition with $\lambda = 0$ (thematic portfolio), while the remaining sectors tend to assume more stable weights for all the values of λ; analogously, Germany is the top country by size of the position for a selection of windows, while it receives a much smaller weight in the passive version of the portfolio. Denmark and Norway similarly have less importance in the thematic portfolio.

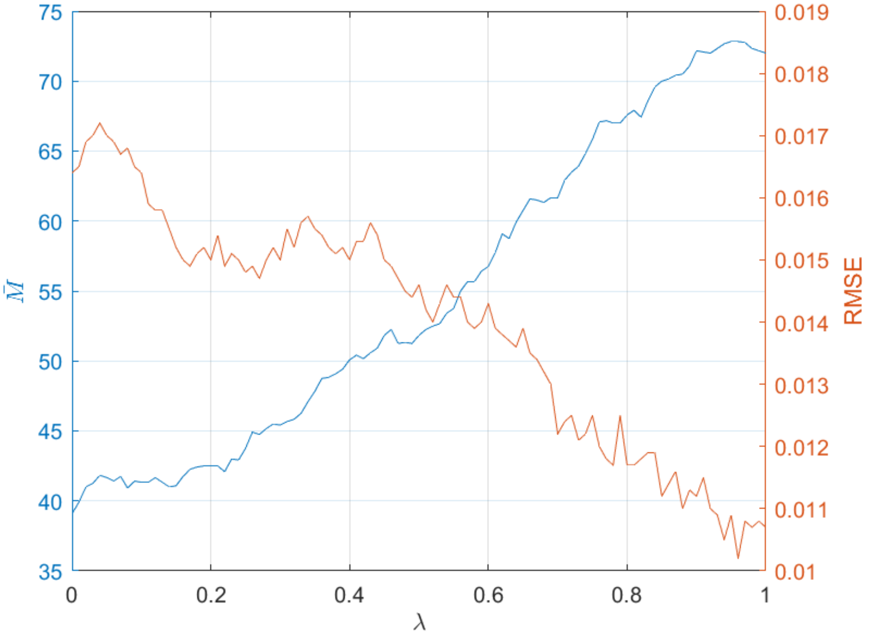

Figure 4: Comparison between the average number of stocks $\bar{M}$ over 12 out-of-sample windows and the Root Mean Square Error (RMSE) of the market-thematic portfolio, where the reference benchmark is the EUROSTOXX 600, for increasing values of λ.

The implications are at least fivefold: a further breakdown of the portfolio in its components, indeed, (for $\lambda = 0.5$) reveals that (i) stock picking is where the two components significantly diverge, as the overall sectoral and geographic exposure is only relatively affected by the presence of a green tilt (ii) a small fraction of the two market-thematic components overlaps, with only four sectors involved in the shared portion of the portfolio, therefore (iii) stocks that belong to the green benchmark are not crucial for replicating the main index, despite being components of the latter. As we show in Figure 4, to achieve higher tracking accuracy, w.r.t. to the main benchmark, (iv) a small subset of assets is sufficient and can effectively replicate the index, although the number required increases for higher values of λ, where the reference universe of assets is the EUROSTOXX 600. By taking into consideration the results in Figures 2 and 3 jointly with Figure 1 it can be concluded that the portfolios are well diversified across sectors and countries, with little variation observed across λ: from an *asset allocation* perspective, however, a marked difference in terms of both portfolio size and tracking error points to a *stock picking* effect. Indeed, besides the presence of limited overlapping between the two portfolio components noted in point (i), note that the inclusion of additional assets improves the tracking performance of the portfolio, but it does not alter the overall allocation. We further explore the presence of such effect by isolating the portfolio inflows and outflows by sector, in each period, in Figure 5: the fact that the turnover of assets entering/leaving the portfolio is high for certain sectors, shows that the underlying asset allocation is stable over time, despite being characterized by the presence of new assets at the end of each period. This is mainly due to the fact most of the variation in each portfolio reoptimization can be attributed to rotation within the same sectors. Finally, (v) the constraint on carbon intensity is crucial in the sense that more importance is given on the carbon budget of the portfolio: what emerges from a visual inspection of Table 4 is that, for different values of λ, notably dissimilar values of carbon intensity can be observed. More precisely, as we show in Figure 6, alternative formulations of the market-thematic portfolio, including the special cases of the thematic one ($\lambda = 0$) or the market portfolio ($\lambda = 1$), have a clear sustainability profile, as the interplays between the E scores and carbon intensity are significant; we conclude that strategies combining both measures are to a certain extent interrelated. The relationship between the two is indeed strong to the point that the two time series display a statistically significant negative correlation of -0.9426. This result will be particularly relevant in the subsequent discussion of several further models in Table 5, where we will compare formulations with different constraints on the environmental performance. Finally, note that by construction the two Figures 4 and 6 can be interpreted respectively as the price to pay in terms of tracking error by deviating from the market formulation ($\lambda = 1$), and the cost in terms of emissions caused by deviating from the thematic portfolio ($\lambda = 0$), driven by a variation of λ.

Let us now consider an array of measures, as reported in the system of equations 3, gauging the portfolio performance from different perspectives. In particular, we employ three tracking

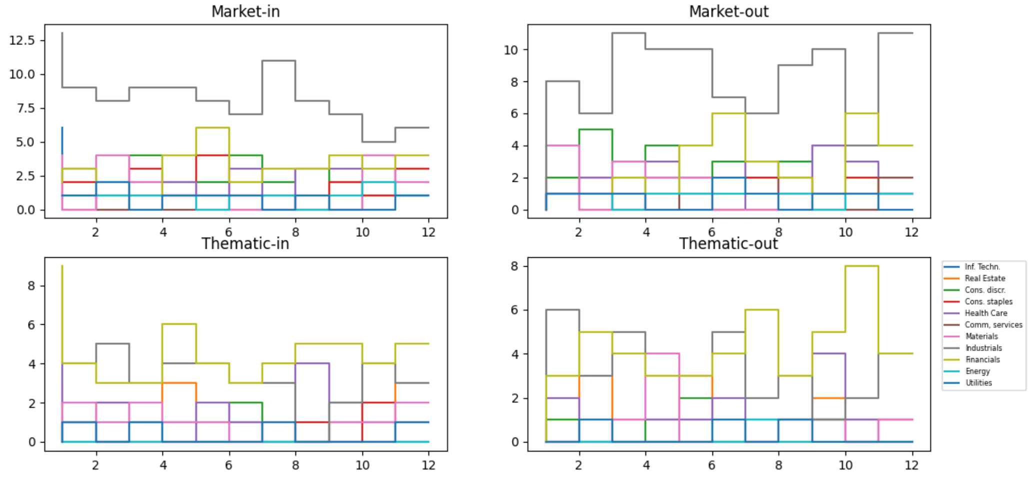

Figure 5: Number of assets entering (in) and leaving the market and thematic portfolio components (out) by sector, over the period January 2023 to December 2023.

measures, i.e. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Then we consider two absolute and two environmental performance measures: Sharpe Ratio (SR), Omega Ratio (OR), Average Environmental Score ($\bar{E}$) and Average Carbon Intensity ($\overline{CI}$). With r_t we denote the true returns, with $\bar{r}_t$ the estimated returns, with $\bar{r}_t$ the average portfolio returns and with σ the portfolio volatility.

$$\begin{cases} MAE & = \frac{1}{T}\sum_{t=1}^{T} |\widehat{r}_t - r_t| \\ RMSE & = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (\widehat{r}_t - r_t)^2} \\ MAPE & = \frac{1}{T}\sum_{t=1}^{T} |\frac{\widehat{r_t} - r_t}{r_t}| \\ SR & = \frac{\bar{r}}{\sigma} \\ OR & = \frac{\frac{1}{T}\sum_{t=1}^{T}[(\widehat{r}_t - \tau)^+]}{\frac{1}{T}\sum_{t=1}^{T}[(\tau - \widehat{r}_t)^+]} \\ \bar{E} & = \frac{\sum_{t=1}^{T} E_t}{T} \\ \overline{CI} & = \frac{\sum_{t=1}^{T} CI_t}{T} \end{cases} \tag{3}$$

In Tables 3 and 4 we present respectively the in-sample and out-of-sample tracking and absolute performance metrics of the four portfolio formulations discussed above. As for the in-sample results, the passive approach unsurprisingly leads to a close fit to the benchmark returns. The market-thematic approach that we introduce in this work displays a slightly higher tracking error, compared to a passive strategy or a passive approach with carbon constraints, although lower

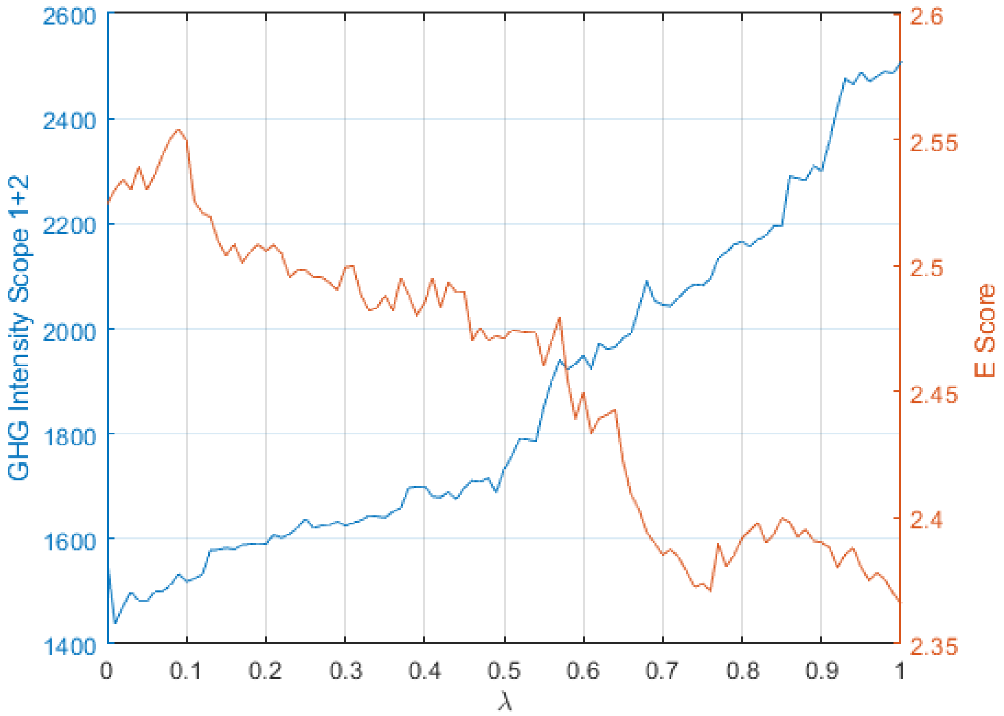

Figure 6: Comparison between the average portfolio carbon intensity GHG over 12 out-of-sample windows and the E score of the market-thematic portfolio, where the reference benchmark is the EUROSTOXX 600, for increasing values of λ.

	MAE	RMSE	MAPE	$\bar{r}$	σ	SR	OR	$\bar{E}$	$\overline{CI}$
Market	0.0007	0.0009	0.0001	0.0091	0.0395	0.1008	1.2726	2.0551	2593.3
Market-thematic	0.0159	0.0180	0.0518	0.0069	0.0239	0.1698	1.5442	1.9911	2283.8
Market-decarbonized	0.0046	0.0077	0.0165	0.0130	0.0409	0.0908	1.2650	2.1226	1880.3
Thematic	0.0307	0.03621	0.1290	0.0139	0.0823	0.0564	1.1566	2.1534	1870.3

Table 3: Comparison between in-sample error and performance statistics of market, market-thematic, market-decarbonized and thematic portfolios.

	MAE	RMSE	MAPE	$\bar{r}$	σ	SR	OR	$\bar{E}$	$\overline{CI}$
Market	0.0092	0.0105	0.0040	0.0106	0.0400	0.2643	1.8790	2.3660	2507.3
Market-thematic	0.0119	0.0144	0.0069	0.0087	0.0438	0.1977	1.6172	2.4732	1687.7
Market-decarbonized	0.0099	0.0118	0.0042	0.0139	0.0391	0.3550	2.4807	2.4338	2395.3
Thematic	0.0143	0.0164	0.0076	0.0034	0.0373	0.0901	1.2515	2.5243	1558.7

Table 4: Comparison between out-of-sample error and performance statistics of market, market-thematic, market-decarbonized and thematic portfolios over a 12 months period, from January 2023 to December 2023.

than the thematic one. Altogether, the strategy shows a good balance between tracking error, and performance on financial and sustainability indicators.

As for the out-of-sample results, we note that the market-thematic strategy is aligned with competing ones in terms of tracking performance, including the thematic portfolio, which displays also poor financial performance, but outperforms other allocations from the sustainability point of view. In this formulation, important differences between a market-thematic approach and a standard strategy replicating the index with a lower bound on portfolio carbon intensity can be also observed. More in general, although it is possible to construct combinations of assets that are more efficient than others for a given level of carbon emissions, implementing sustainability rules involves finding a trade-off with financial and tracking performance: hence the better thematic or market-thematic sustainability profiles come at a price, in terms of tracking error, even when compared with a green version of the market portfolio.

In Figure 7 we report the cumulative wealth paths for all the strategies and benchmarks, assuming the initial wealth is equal to 1000: note that the proposed market-thematic and the two alternative strategies tracking the main index perform similarly.

Now let us evaluate the market-thematic approach separately, by considering an array of constraints to be imposed on the set of feasible solutions, assuming different formulations of the problem. We want to verify whether there exists a nexus between the environmental performance, as measured by Bloomberg Environmental scores, and the firm's carbon intensity. In other words, due to the fact that the constraint is imposed on pre-screened firms belonging to the environmental index, and the environmental scores already incorporate an assessment of carbon emissions, the inclusion of a constraint on greenhouse gases intensity might have little relevance in our analysis. We consider four different settings, in which the portfolio tracking and sustainability performance is evaluated assuming (i) no constraints, (ii) a constraint on Environmental scores, (iii) a constraint on carbon intensity, or (iv) both, where the latter formulation allows us to identify whether the constraint on carbon emissions is active.

Based upon the results in Table 5, several relevant conclusions can be derived. First, all the solutions are characterized by similar tracking and financial performance, suggesting that the in-

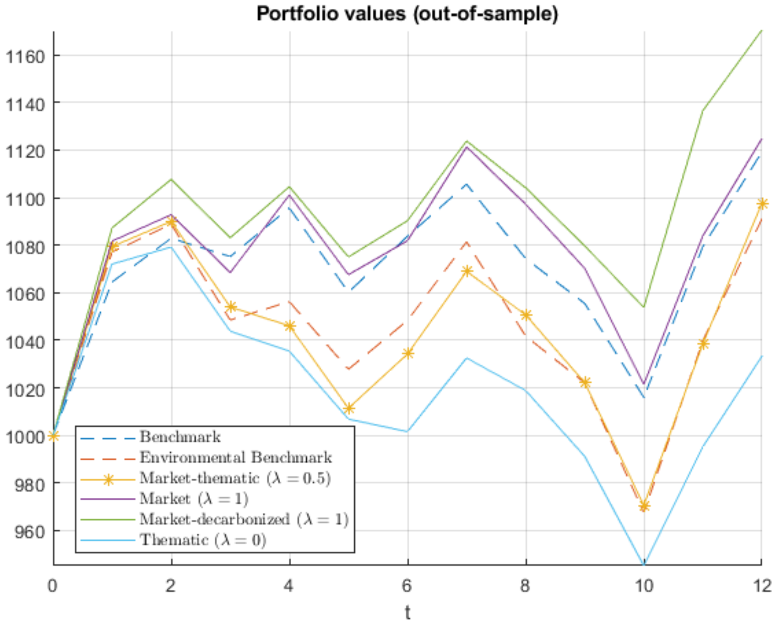

Figure 7: Out-of-sample cumulative wealth trajectory of 1000 invested at time 0 for different portfolio strategies, over the period January 2023-December 2023.

	MAE	RMSE	MAPE	$\bar{r}$	σ	SR	OR	$\bar{E}$	$\overline{CI}$
GHG+E constraint	0.0145	0.0186	0.0067	0.0143	0.0454	0.3145	2.2666	2.2004	1855.3
No constraints	0.0220	0.0250	0.0079	0.0156	0.0524	0.2982	2.2642	1.9164	2469.8
GHG constraint	0.0187	0.0233	0.0085	0.0133	0.0524	0.2543	1.8764	2.1135	1731.1
E constraint	0.0117	0.0143	0.0069	0.0084	0.0440	0.1900	1.5848	2.4677	1971.3

Table 5: Comparison between out-of-sample error and performance statistics of the market-thematic portfolio assuming different constraints over a 12 months period, from January 2023 to December 2023.

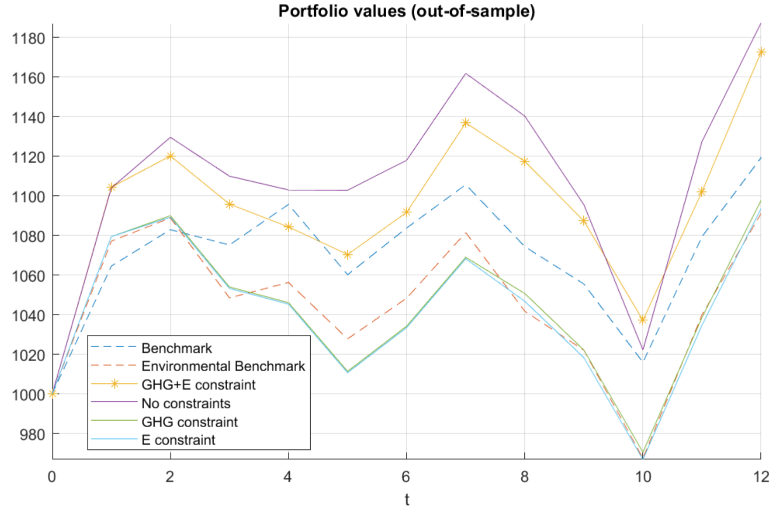

Figure 8: Out-of-sample cumulative wealth trajectory of 1000 invested at time 0 for the market-thematic strategy with different constraints, over the period January 2023-December 2023.

troduction of constraints on environmental performance has a negligible impact, regardless of the precise formulation. Second, all the constraints are active: despite a clear interaction between the two, as higher scores limit the portfolio's carbon intensity, better results in terms of GHG intensity emerge when specific requirements are imposed. The reference descriptive statistics of the E and GHG scores, along with the constraints thresholds on the full dataset and on the filtered set within the green benchmark are reported in Table 6.

In Figure 8 we report the cumulative wealth trajectories across four formulations of the problem, where different constraints are taken into account. We note that the four proposals do not deviate significantly from the baseline trajectory, for which a minimum target of emissions is specified. A further careful examination of the main statistics of the E scores and carbon intensity, shows that high dispersion across firms can be observed, although the average scores of the main benchmark do not deviate too significantly from those of the green index. Our results suggest that the two constraints are inherently related, and possibly only a highly restrictive constraint on scores can counterbalance this interaction effect.

4 Conclusions

In this contribution we have dealt with the problem of constructing tracking portfolios with a green mandate, where the formulation simultaneously incorporates exposure to a broad index and a thematic one. A biobjective formulation is designed to take into account the exposure to both,

	Mean	Std	Skewness	Kurtosis	Threshold (I)	Threshold (I_E)
E	2.42	2.11	0.49	2.21	2.42	2.75
GHG	1517.08	1115.70	1.18	2.87	1441.12	938.61

Table 6: Descriptive statistics and constraint thresholds for the E score and carbon intensity, where I denotes the global sample and I_E the subsample containing the assets within the environmental benchmark.

respectively in terms of a symmetric risk measure for the former component and an asymmetric risk measure for the latter.

Results highlight that our proposal can flexibly capture different preferences of investors, while taking exposure to green-themed assets and keeping the overall tracking error low. We also show that our approach generalizes other previously proposed models and widely adopted in practical asset management to incorporate an ESG mandate or a decarbonization target.

Future work should focus on (i) the definition of a multistage tracking strategy controlling for portfolio rebalancing and associated transaction costs, dealing with the limitations of a static formulation or (ii) the integration of data-driven scenario generation based on generative machine learning.

A Wasserstein GAN

Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) are a class of machine learning frameworks where two neural networks, a generator G and a discriminator D, are trained simultaneously through adversarial processes. The generator G creates synthetic data samples, while the discriminator D evaluates them against real data samples. The goal of the generator is to produce data that is distributionally indistinguishable from the real data, while the discriminator aims to correctly distinguish between real and generated data.

The core idea of GANs can be formalized using the following minimax game with the value function $V(G, D)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

Here, $p_{data}(x)$ is the distribution of the real data, and $p_z(z)$ is the distribution of the input noise z to the generator. The discriminator D aims to maximize the probability of correctly identifying real and fake data samples, while the generator G tries to minimize the probability that the discriminator correctly identifies the fake samples.

During training, the generator learns to map from a noise distribution $p_z(z)$ to the data space $G(z)$, improving over iterations to create more realistic data. Conversely, the discriminator gets better at distinguishing real from fake samples. This adversarial process continues until the generator produces samples that are indistinguishable from real data to the discriminator.

The Wasserstein GAN with Gradient Penalty (WGAN-GP) (Arjovsky et al., 2017; Gulrajani et al., 2017) addresses significant limitations in the traditional GAN framework, particularly the issues related to training stability and mode collapse. The key innovation in WGAN-GP is the use of the Wasserstein distance (Earth Mover's distance) instead of the Jensen-Shannon divergence used in standard GANs. The Wasserstein distance offers a more meaningful gradient signal, which stabilizes the training of GANs.

The loss functions for the WGAN are derived from the Wasserstein distance $W(p_r, p_g)$ between the real data distribution p_r and the generated data distribution p_g as follows:

$$W(p_r, p_g) = \sup_{\|f\|_L \leqslant 1} \mathbb{E}_{x \sim p_r}[f(x)] - \mathbb{E}_{x \sim p_g}[f(x)]$$

Here, f represents a 1-Lipschitz function, and $\|f\|_L \leqslant 1$ denotes that f satisfies the Lipschitz constraint. In practice, the discriminator (also called the critic in WGAN) approximates this function.

To enforce the Lipschitz constraint, the original WGAN uses weight clipping, which often leads to undesirable side effects. WGAN-GP improves upon this by introducing a gradient penalty. The modified loss function for the critic in WGAN-GP is:

$$L_D = \mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})] - \mathbb{E}_{x \sim p_r}[D(x)] + \lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

where D is the critic, λ is the penalty coefficient, and $\hat{x}$ are samples interpolated between real and generated data:

$$\hat{x} = \epsilon x + (1 - \epsilon)\tilde{x}, \quad \epsilon \sim \mathcal{U}(0, 1)$$

The gradient penalty term $\lambda \mathbb{E}_{\hat{x} \sim p_{\hat{x}}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$ ensures that the gradient norm $\|\nabla_{\hat{x}} D(\hat{x})\|_2$ remains close to 1, thereby maintaining the Lipschitz constraint without the drawbacks of weight clipping.

The generator's loss function remains the same as in the original WGAN, aiming to minimize the critic's evaluation of the generated data:

$$L_G = -\mathbb{E}_{\tilde{x} \sim p_g}[D(\tilde{x})]$$

The combined effect of these loss functions results in more stable training dynamics and reduces mode collapse. The critic provides a more informative gradient to the generator, leading to more effective learning and better quality of generated samples.

The Wasserstein distance used in WGAN-GP offers several advantages over the traditional GAN approach, including better convergence properties and more stable gradients. This makes WGAN-GP particularly suitable for complex generative tasks where maintaining the diversity and quality of generated samples is crucial.

From a network architecture point of view, we aim to capture the overall correlation structure between columns. Prior to this, we test extensively the model to understand (i) which architecture can be used to generate the scenarios and (ii) whether the methodology can adequately capture the distribution of the data. Based on standard grid search, we find that a relatively general and shallow architecture with two fully-connected hidden layers in both generator and critic is suitable for our problem. We use batch-normalization and ReLU activation function for the generator and leaky ReLU and dropout on each hidden layer for the critic, with 256 neurons each. The model is trained with Adam optimizer and learning rate $2 \cdot 10^{-4}$. Moreover, the simulated time series are found to characterize the original data reasonably well. We verify in particular that, according to standard 2-sample Kolmogorov-Smirnov tests, the distributions of simulated data are not statistically different from the original ones; subsequently, we find that the correlation structure between columns is preserved. Limited autocorrelation emerges from the data.

Disclaimer

This paper was developed within the projects funded by: Next Generation EU - "GRINS - Growing Resilient, INclusive and Sustainable" project (PE0000018), National Recovery and Resilience Plan (NRRP) – PE9 - Mission 4, C2, Intervention 1.3; Next Generation EU - "Just Energy Transition – JET: Stochastic and machine learning methods for the evaluation, mitigation and geographical hedging of involved natural risks (with climate in view)" project (P2022XTLM2), National Recovery and Resilience Plan (NRRP) - Mission 4, C2, Intervention 1.1. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission or the Italian Ministry of University and Research. Neither the European Union, the European Commission nor the Italian Ministry of University and Research can be held responsible for them.

Acknowledgments

We would like to thank participants of the 69th meeting of the EURO Working Group on Commodity and Financial Modeling for useful comments.

References

Alessandrini, F. and Jondeau, E. (2021). Optimal strategies for ESG portfolios. *The Journal of Portfolio Management*, 47(6):114–138.

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

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Assaf, C., Monne, J., Harriet, L., and Meunier, L. (2024). ESG investing: Does one score fit all investors' preferences? *Journal of Cleaner Production*, page 141094.

Bauer, R., Ruof, T., and Smeets, P. (2021). Get real! Individuals prefer more sustainable investments. *The Review of Financial Studies*, 34(8):3976–4043.

Billio, M., Costola, M., Hristova, I., Latino, C., and Pelizzon, L. (2021). Inside the ESG ratings:(Dis) agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5):1426–1445.

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2):517–549.

Cesarone, F., Martino, M. L., and Carleo, A. (2022). Does ESG impact really enhance portfolio profitability? *Sustainability*, 14(4):2050.

Cox, P., Brammer, S., and Millington, A. (2004). An empirical examination of institutional investor preferences for corporate social performance. *Journal of business ethics*, 52:27–43.

Dahl, C. M. and Sørensen, E. N. (2022). Time series (re) sampling using generative adversarial networks. *Neural Networks*, 156:95–107.

De Spiegeleer, J., Höcht, S., Jakubowski, D., Reyners, S., and Schoutens, W. (2023). ESG: A new dimension in portfolio allocation. *Journal of Sustainable Finance & Investment*, 13(2):827–867.

Dembo, R. and Rosen, D. (1999). The practice of portfolio replication. A practical overview of forward and inverse problems. *Annals of Operations Research*, 85(0):267–284.

Escrig-Olmedo, E., Rivera-Lirio, J. M., Muñoz-Torres, M. J., and Fernández-Izquierdo, M. Á. (2017). Integrating multiple esg investors' preferences into sustainable investment: A fuzzy multicriteria methodological approach. *Journal of cleaner production*, 162:1334–1345.

Eurosif (2023). Eurosif Report 2023. Available at `https://www.eurosif.org/news/eurosif-report-may-2023/`. Accessed: 2024-03-02.

Gilli, M., Maringer, D., and Schumann, E. (2019). *Numerical methods and optimization in finance*. Academic Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of Wasserstein GANs. *Advances in neural information processing systems*, 30.

Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management science*, 37(5):519–531.

Koshiyama, A., Firoozye, N., and Treleaven, P. (2021). Generative adversarial networks for financial trading strategies fine-tuning and combination. *Quantitative Finance*, 21(5):797–813.

Le Guenedal, T., Lombard, F., Roncalli, T., and Sekine, T. (2022). Net zero carbon metrics. *Available at SSRN 4033686*.

Le Guenedal, T. and Roncalli, T. (2022). Portfolio construction with Climate Risk Measures. *Available at SSRN 3999971*.

Methling, F. and von Nitzsch, R. (2019). Thematic portfolio optimization: challenging the core satellite approach. *Financial markets and portfolio management*, 33(2):133–154.

Methling, F. and von Nitzsch, R. (2020). Tailor-made thematic portfolios: a core satellite optimization. *Journal of Global Optimization*, 76(2):317–331.

Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2):572–597.

Rojo-Suárez, J. and Alonso-Conde, A. B. (2024). Have shifts in investor tastes led the market portfolio to capture ESG preferences? *International Review of Financial Analysis*, 91:103019.

Silva, J. C. S. and de Almeida Filho, A. T. (2023). Using GAN-generated market simulations to guide genetic algorithms in index tracking optimization. *Applied Soft Computing*, 145:110587.

Steuer, R. E. and Utz, S. (2023). Non-contour efficient fronts for identifying most preferred portfolios in sustainability investing. *European Journal of Operational Research*, 306(2):742–753.

Takahashi, S., Chen, Y., and Tanaka-Ishii, K. (2019). Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications*, 527:121261.

Utz, S., Wimmer, M., and Steuer, R. E. (2015). Tri-criterion modeling for constructing more-sustainable mutual funds. *European Journal of Operational Research*, 246(1):331–338.

Xidonas, P. and Essner, E. (2022). On ESG Portfolio Construction: A Multi-Objective Optimization Approach. *Computational Economics*, pages 1–25.