Computational Management Science (2026) 23:5
https://doi.org/10.1007/s10287-025-00548-z

ORIGINAL PAPER

Mean-CVaR portfolio optimization under ESG disagreement

Davide Lauria[1,2] · Marco Bonomelli[1] · Gabriele Torri[1] · Rosella Giacometti[1]

Received: 15 April 2025 / Accepted: 4 November 2025

Abstract
The ESG score of a company is a measure of its commitment to environmental, social and governance investing standards. ESG scores are produced by rating agencies using unique and proprietary methodologies. The complexity of measurement and the lack of widely accepted standards contribute to inconsistencies across agencies. Discrepancies in ratings issued by multiple data providers are particularly relevant in portfolio optimization problems that integrate ESG objectives into the classical risk-reward framework. In this work, we specifically study the impact on portfolio composition by examining Mean-CVaR-ESG optimal portfolios, where the objective function incorporates the portfolio's ESG score. To address ESG score discrepancies, we introduce a Distributionally Robust Optimization (DRO) reformulation of the Mean-CVaR-ESG model and assess its potential benefits. Our findings reveal a persistent divergence in optimal strategies across the investment horizon when ESG values from different rating agencies are used. We then apply the DRO approach by replacing a single provider's ESG score with a statistic derived from the scores of five different agencies. Our results show that, in this case, the DRO approach effectively mitigates score discrepancies by significantly reducing optimal portfolio concentration while enhancing the ESG evaluation of optimal portfolios across all rating agencies.

Keywords Distributionally robust optimization · ESG investing · Mean-CVaR portfolio optimization

1 Introduction

Environmental, social and governance (ESG) issues have become a significant factor in investors' choices in recent years. Classic risk and return preferences are now joined by new investment profiles in terms of environmental and social sustainability.

Extended author information available on the last page of the article

Published online: 27 November 2025

This trend is already recognized by large financial institutions, which have integrated ESG based products, and regulators. As a result, sustainable investments have seen an increase, both through the selection of securities issued by entities with good ESG performance and through the emergence of new securities linked to the achievement of sustainability goals.

ESG Optimal Portfolios refer to financial strategies in which portfolio weights are optimized by integrating ESG values into the classical framework of maximizing the trade-off between expected return and risk. Various examples and optimization methods have been proposed, analyzed, and discussed in recent literature. One common approach is to define the objective function as a linear combination of three components: expected return, a risk measure, and a portfolio-level ESG score. Typically, the risk measure is the classical variance, following the original framework introduced by Markowitz (1952). The assumption of additivity between financial returns and ESG scores, a common simplification in ESG-aware portfolio optimization, appears both in theoretical models analyzing the equilibrium effects of ESG investing, e.g., Pedersen et al. (2021), Avramov et al. (2022) and in applied portfolio optimization approaches, e.g., Cesarone et al. (2022), Cesarone et al. (2024), Utz et al. (2015), Varmaz et al. (2024). This assumption implies that financial risk is treated independently from ESG considerations, effectively framing the investor's problem as a tri-criteria optimization. In such formulations, the coefficient λ functions as a scaling parameter governing the trade-off between financial and non-financial (ESG) objectives. While this structure offers tractability and interpretability, it may overlook potentially nonlinear interactions between ESG characteristics and asset returns. Moreover, since ESG optimal portfolio models are typically determined for a single period, they do not account for intertemporal coherence, leading to potential time inconsistency: decisions that are optimal today may not remain optimal in the future. Multi-period approaches can, in principle, address these limitations and allow conditional risk measures to be incorporated directly into the objective function (see, e.g., de Mello and Pagnoncelli 2016; Shapiro 2021; Pichler and Shapiro 2021, and references therein), albeit at the cost of greater estimation and analytical complexity.

Regardless of the chosen portfolio model, investors face two main types of risks associated with the ESG dimension. The first risk, that we can call *temporal uncertainty*, arises from the fact that ESG values are revealed at low frequencies, typically annually, quarterly, or monthly.[1] As a result, the decision-maker will only discover the most recent ESG score of each company after making the investment decision. This risk is analogous to the risk investors face regarding the financial performance of their portfolio, where the realized value becomes observable only after the investment decision has been made. The second risk, called *ESG Disagreement*, unique to the ESG dimension, stems from the absence of a standardized methodology for assessing the ESG score of a company. Different rating agencies use proprietary methodologies, which involve varying measurements, statistical approaches for aggregation, and even different scales or ordering. Moreover, this risk is linked to the first: the time variation of ESG scores is statistically different across providers.

[1] One exception is the ESG index produced by Truvalue, which has a daily frequency and is unique in its methodology.

The ambiguity of ESG ratings has been studied in several studies. Billio et al. (2021) pointed out that heterogeneity in rating criteria can lead agencies to hold opposing opinions about the same evaluated companies and that agreement among those providers is substantially low. Furthermore, the disagreement among the scores provided by the rating agencies disperses the impact of preferences of ESG investors on asset prices. On the extreme case, even when there is agreement, it has no impact on financial performance. The analysis is further extended in Billio et al. (2024), confirming the general disagreement of the providers on the ESG evaluation of a company.

Berg et al. (2022), pointed out that ESG ratings divergence is shaped by three sources. The first source of divergence, called scope, arises from the fact that agencies measure different ESG attributes. For instance, one agency may consider the release of certain chemicals, while others do not. The second source, measurement divergence, occurs when agencies assess the same attributes using different metrics. Finally, weight divergence stems from variations in the importance assigned to the same attributes, as agencies prioritize different ESG components. They find that most differences result from measurement and scope divergence, while weight divergence plays a minor role. This result is crucial as it highlights that ESG ratings are not fully comparable. Therefore, any decision-making process that relies, even partially, on a company's ESG score should be robust to variations across different rating providers. Indeed, since agencies measure, to some extent, different features using varying methodologies, we cannot directly compare their ESG values or the order in which they rank companies. Thus, the analysis of the gap in ESG valuation expressed by different raters represents an element of strong interest for decision makers.

Although several authors have studied the causes and implications of ESG score discrepancies, few have examined their consequences for portfolio optimization. Most research on ESG-optimal portfolios has focused on integrating the ESG component into the optimization process, without examining the selection of ESG scores themselves. Among the few exceptions is the study by Cesarone et al. (2024), which addresses ESG disagreement by applying the k-sum operator across providers at the portfolio level, yielding a conservative ESG measure that is incorporated into the optimization alongside expected return and risk.

Our paper, with the aim of filling this gap, contributes to the literature on ESG-optimal portfolio methods that are robust to ESG discrepancies. In particular, we consider Mean-CVaR-ESG optimal portfolios that incorporate ESG preferences in the objective function by linearly aggregating financial returns and ESG scores. We then introduce its DRO reformulation within a data-driven framework that handles uncertainty by directly employing the empirical distribution function, without the need for fitting models to predict the uncertainty vectors. We believe this approach is particularly suited for modeling ESG uncertainty, especially over time. Furthermore, it avoids the need to model the joint distribution of ESG data and financial returns. Indeed, as most providers offer quarterly or annual data, it is generally challenging to produce reliable estimates of ESG score evolution and model them alongside financial returns. The choice of CVaR is based on its well-established properties as a coherent risk measure with appealing theoretical and practical benefits. In particular, we rely on Esfahani and Kuhn (2018), who presents a method for reformulating the DRO portfolio problem with piecewise linear objective functions as a linear programming problem. Our tests are conducted using ESG scores from five of the most widely used ESG rating agencies, applied to a subset of 303 assets from the Eurostoxx Europe 600 index as of June 29, 2024, offering a focused perspective on the European market.

The scope of this paper is twofold. The first objective is to assess the impact of ESG ambiguity on portfolio selection problems, focusing on the actual portfolio strategies implemented by investors. Specifically, we aim to quantify, out-of-sample, the extent to which two optimal strategies differ when computed using ESG values from various agencies. We examine these differences from multiple perspectives and introduce uncertainty into the empirical distribution of ESG scores provided by each individual agency. We adopt this approach because, if discrepancies between scores from different providers were minor, a robust reformulation might suffice to manage the ambiguity. However, we demonstrate that this is not the case. Therefore, we consider the various providers as simultaneous observers of a company's ESG characteristics. This framework provides an ideal setting for constructing the jointly observed empirical distribution, capturing the potential dependency between ESG scores and financial returns without imposing specific distributional assumptions.

As an initial attempt to address ESG discrepancies, we propose two simple aggregation measures: the arithmetic average of all providers' scores and a linear combination of this average with the standard deviation, capturing both central tendency and variability in ESG ratings. However, applying such statistics to traditional Mean-CVaR-ESG portfolios, without accounting for ESG uncertainty, results in highly concentrated portfolios, even with relatively low ESG preferences. Therefore, our second objective is to test whether the DRO approach, which constructs an uncertainty ball around the jointly observed empirical distribution of returns and ESG values from multiple providers, can address the ambiguity problem and deliver better performance than each individual provider.

Our findings show that the DRO approach effectively addresses this issue by significantly reducing optimal portfolio concentration, while preserving the financial and ESG performance across all rating agencies, as achieved by the traditional Mean-CVaR model without distributional uncertainty.

The remainder of the paper is organized as follows. Section 2 introduces the general optimization problem, along with a discussion of the optimization method we have selected. Section 3 presents the datasets and provides extensive out-of-sample results on the impact of ESG ambiguity on optimal solutions and portfolio performance when individual providers are considered. In Sect. 4, we test the DRO portfolio model when the ESG score used in the optimization is a statistic based on the scores of five providers. Finally, Sect. 5 presents final comments and suggestions for future research.

Notation We denote by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ the inner product between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. The expected value of random variable under the probability measure $\mathbb{Q}$ is denoted by $\mathbb{E}^{\mathbb{Q}}[\cdot]$. The vector of ones in $\mathbb{R}^n$ is be denoted by $\mathbf{1}_n := [1, 1, \ldots, 1]^\top$. The p-norm of $\boldsymbol{x} \in \mathbb{R}^n$ is defined as $||\boldsymbol{x}||_p := \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$ and the maximum (infinity) norm as $\|\boldsymbol{x}\|_\infty := \max(|x_1|, \ldots, |x_n|)$. The Dirac measure, which assigns a unit mass at x

and zero elsewhere, is denoted by δ_x. The dual norm of a vector y with respect to the norm $||\cdot||$ is defined as $||\cdot||_*$.

2 DRO-mean-CVaR portfolio optimization with ESG preferences

Consider the case of an investor who seeks the best portfolio allocation among I public companies quoted in a public market. The stock return of the i-th company between two dates $t-1$ and t will be denoted by $r_{i,t}$, for $i \in \mathcal{I} = \{1, \ldots, I\}$, an element of the random vector $\boldsymbol{r_t} = [r_{1,t}, \ldots, r_{I,t}]$. We then introduce J random vectors $\boldsymbol{z_t^{(j)}} = \left[z_{1,t}^{(j)}, \ldots, z_{I,t}^{(j)}\right]$, for $j \in \mathcal{J} = \{1, \ldots, J\}$, each of which containing ESG scores, possibly modified accordingly to some normalization,[2] assigned by the j-th rating agency to each company $i \in \mathcal{I}$. If we consider an investor who selects one of the J agencies to incorporate the ESG scores into its financial decision, its uncertainty will be fully represented by one of the J vectors $\boldsymbol{\xi^{(j)}}_t = \left[\boldsymbol{r}_t, \boldsymbol{z}_t^{(j)}\right]^\top$ of dimension $(2I \times 1)$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The investor problem can be stated as the selection of the vector of portfolio weights $\boldsymbol{w_t}$ among the I companies at time t in the feasible set defined as $\mathcal{W}_t := \left\{\boldsymbol{w}_t \in \mathbb{R}^I : \boldsymbol{w}_t^\top \mathbf{1}_I = 1; \boldsymbol{w}_t \geq 0\right\}$, which ensures that the portfolio avoids short-selling positions and fully invests the available wealth in the risky assets. Since the optimization problem involves only one decision at time t, we will omit the subscript t from now on for simplicity.

The decision maker seeks the minimization of a loss function $l(\boldsymbol{w}, \boldsymbol{\xi})$ which depends on the decision vector $\boldsymbol{w}$ and the uncertainty vector $\boldsymbol{\xi}$. So in general we would be able to solve the stochastic optimization problem:

$$J_N := \inf_{w \in \mathcal{W}} \mathbb{E}^{\mathbb{P}} \left[l(\boldsymbol{w}, \boldsymbol{\xi})\right]. \tag{1}$$

For what concerns the loss function $l(\boldsymbol{w}, \boldsymbol{\xi})$, we assume that the investor has ESG preferences, meaning that, in general, they are willing to tolerate a deterioration in their reward-risk profile in order to achieve a better ESG score for the portfolio. In this work, we consider a reward-risk measure defined as a linear combination of the expected reward of the portfolio and the Conditional Value-at-Risk[3] ($\text{CVaR}_\beta^{\mathbb{P}}$). Here, the expected reward itself is a linear combination of the expected financial return and the expected ESG score of the portfolio, forming a nested linear structure that allows to incorporate both financial and sustainability preferences while maintaining a tractable optimization framework. This configuration leads, for some $j \in \mathcal{J}$, i.e., for a given choice of the rating agency, to the following expected loss function:

[2] The normalization ensures that the ESG scores provided by different rating agencies lie in the same interval, for instance in the set $[0, 1]$.

[3] The $CVaR_\beta$ is a risk measure introduced by Rockafellar and Uryasev (2002) that quantifies the expected loss given that the loss exceeds the Value at Risk (VaR) at a significance level β.

$$\mathbb{E}^{\mathbb{P}}\left[l\left(\boldsymbol{w},\boldsymbol{\xi}\right)\right]= \ -\alpha\left[\ \lambda\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{z}^{(j)}\right]+(1-\lambda)\,\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{r}\right]\ \right]+(1-\alpha)\mathrm{CVaR}_{\beta}^{\mathbb{P}}\left(-\boldsymbol{w}^{\top}\boldsymbol{r}\right), \tag{2}$$

where the parameter $\alpha \in \mathbb{R} \cap [0, 1]$ represents the weight the optimizer assigns to the expected reward, which is itself a combination of the expected financial returns and the expected ESG scores, weighted by the parameter $\lambda \in \mathbb{R} \cap [0, 1]$. In other words, the parameter λ balances the weights that the investor assigns to the utility between financial returns and the ESG value of the portfolio. The CVaR term can be expressed using the representation introduced in Rockafellar and Uryasev (2002) as:

$$\mathrm{CVaR}_{\beta}^{\mathbb{P}}\left(-\boldsymbol{w}^{\top}\boldsymbol{r}\right)=\min_{\tau\in\mathbb{R}}\left\{\tau+\frac{1}{1-\beta}\mathbb{E}^{\mathbb{P}}\left[\left(-\boldsymbol{w}^{\top}\boldsymbol{r}-\tau\right)_{+}\right]\right\}, \tag{3}$$

where $(x)_+ = \max(x, 0)$ denotes the positive part function, and the optimal solution τ^* corresponds to the Value-at-Risk (VaR) at significance level β. The ESG score vector $\boldsymbol{z}^{(j)}$ depends on the selected ESG rating agency $j \in \mathcal{J}$, reflecting the variability in ESG assessments across providers. Consequently, the optimization outcome may be sensitive to the choice of rating agency, which is an important consideration when evaluating the robustness of ESG-integrated portfolios.

The probability measure $\mathbb{P}$ is not directly observable and must be estimated using the previous N realizations of each uncertainty process $\xi^{(j)}$, for $j = 1, \ldots, J$. The n-th vector of observations at a given time, $\hat{\boldsymbol{\xi}}_n^{(j)} = \left[\hat{\boldsymbol{r}}_n,\ \hat{\boldsymbol{z}}_n^{(j)}\right]$, for $n \in \mathcal{N} = \{1, \ldots, N\}$, constitutes the rows of the historical data matrix $\hat{\Xi}_N^{(j)}$, for $j \in \mathcal{J}$. We can assume that each $\hat{\Xi}_N^{(j)}$ is drawn from a probability measure $\mathbb{P}$, which is not observable and difficult to estimate, of the random vector $\boldsymbol{\xi}$ defined on some support $\Xi^{(j)}$. By assuming that all the ESG scores are normalized on the same interval, we can simplify the notation by assuming that the support is the same for every rating agency, so that Ξ is used for all $j \in \mathcal{J}$. The empirical probability measure implied by each of these data matrices will be then defined as $\hat{\mathbb{P}}^{(j)} = \frac{1}{N}\sum_{n=1}^{N}\delta_{\hat{\boldsymbol{\xi}}^{(j)}_n}$.

As stated in the introduction, and motivated by the difficulty of forecasting ESG scores at low frequencies and estimating a joint probability measure for both financial returns and ESG values, we consider the case in which the investor adopts a data-driven approach by replacing the general probability measure $\mathbb{P}$ with its empirical counterpart $\hat{\mathbb{P}}^{(j)}$. Under such assumption, the optimal portfolio problem (2) can be reformulated as a linear programming problem by introducing the positive slack variables $\{s_n\}_{n=1}^{N}$:

$$\begin{aligned}
&\min_{\boldsymbol{w}\in\mathcal{W},\tau,\boldsymbol{s}}\left\{-\frac{\alpha}{N}\sum_{n=1}^{N}\left[\lambda\boldsymbol{w}^{\top}\hat{\boldsymbol{z}}_n^{(j)}+(1-\lambda)\boldsymbol{w}^{\top}\hat{\boldsymbol{r}}_n\right]+(1-\alpha)\left(\tau+\frac{1}{1-\beta}\sum_{n=1}^{N}\frac{1}{N}s_n\right)\right\}\\
&\text{subject to:}\\
&-\boldsymbol{w}^{\top}\hat{\boldsymbol{r}}_n-\tau-s_n\leq 0,\quad s_n\geq 0,\quad n=1,\ldots,N.
\end{aligned} \tag{4}$$

From a practical standpoint, for any decision regarding the rating agency j at a given time t, we must define the length N, which represents the number of previous observations needed to estimate the empirical measure. This choice, in turn, influences the optimal solution. In general, increasing the sample size N improves the estimation of the unobservable distribution of variables. However, at the same time, it may lead to the loss of information about specific market features that characterize the period in which the optimization is performed.

The Distributionally Robust Optimization (DRO) approach consists of solving the problem under the worst possible probability measure, among a set of alternatives that lie within a certain distance of the observed empirical measure, and that share the same support Ξ. Consider, for instance, the set $P(\Xi)$ of probability measures defined on the support Ξ, and its subset $\mathcal{M}(\Xi) := \left\{\mathbb{Q} \in P(\Xi) : \int_{\Xi} |\xi| \; d\mathbb{Q}(\xi) < \infty\right\}$ of $\mathcal{L}^1$ -integrable measures. The Wasserstein Distance[4] between two distributions $\mathbb{Q}_1, \mathbb{Q}_2$ with support Ξ, under the norm $||\cdot||$, is defined as:

$$d_W(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\Pi \in \mathcal{P}} \left\{ \int_{\Xi \times \Xi} ||\xi_1 - \xi_2|| d\Pi(\xi_1, \xi_2) ; \mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{M}(\Xi) \right\} \tag{5}$$

where $\mathcal{P}$ is the set of joint distributions with marginals $\mathbb{Q}_1, \mathbb{Q}_2$. Consider now the set $B_\epsilon\left(\hat{\mathbb{P}}_N\right) := \left\{\mathbb{Q} \in \mathcal{M}(\Xi) : d_W\left(\mathbb{Q}, \hat{\mathbb{P}}_N\right) \leq \epsilon\right\}$ of distributions that belong to the ball of radius ϵ around $\hat{\mathbb{P}}_N$.

The DRO problem is stated as:

$$\inf_{\boldsymbol{w} \in \mathcal{W}} \sup_{\mathbb{Q} \in B_\epsilon\left(\hat{\mathbb{P}}_N\right)} \mathbb{E}^{\mathbb{Q}}\left[\ell(\boldsymbol{w}, \boldsymbol{\xi})\right]. \tag{6}$$

As shown in Esfahani and Kuhn (2018), problem (6) admits a tractable reformulation under the following assumptions:

i) *Light-tailed distributions:* The ambiguity set $\mathbb{Q}$ is restricted to the class of light-tailed distributions. That is, there exists a constant $a > 1$ such that $\mathbb{E}^{\mathbb{Q}}\left[\exp\left(\|\boldsymbol{\xi}\|^a\right)\right] < \infty$. As noted in Esfahani and Kuhn (2018), this condition is trivially satisfied when the support Ξ is compact. In our setting, where short-selling is not allowed, daily returns can be bounded within the interval $[-1, 1]$, and ESG scores lie in [0, 1]. Hence, Ξ is compact, and the light-tail condition holds.[5]
ii) *Piecewise linear loss function:* The function $\ell(\boldsymbol{w}, \boldsymbol{\xi})$ is piecewise linear, i.e.,

[4] While it is common in the scientific literature to refer to this metric as the Wasserstein distance, the foundational formulation is due to Kantorovich. In this context, it is sometimes referred to as the Kantorovich or Kantorovich–Rubinstein metric. See, for instance, Rachev (1991) for a thorough exposition of this argument.

[5] We assume that daily returns lie in $[-1, 1]$, acknowledging that while returns exceeding 100% are theoretically possible in some markets, they are extremely rare in our application. For instance, in the dataset we use, which covers 303 assets from January 1, 2014, to June 30, 2024, the observed range of daily returns lies within $[-0.41, 0.34]$.

$$\ell(\boldsymbol{w}, \boldsymbol{\xi}) := \max_{k \in \mathcal{K}} \left[a_{1,k}\langle \boldsymbol{w}, \boldsymbol{r}_n\rangle + a_{2,k}\langle \boldsymbol{w}, \boldsymbol{z}_n\rangle + b_k\right],$$

for some finite index set $\mathcal{K} \subset \mathbb{N}$.

iii) *Polytopic support:* The uncertainty set is a polytope, $\Xi = \left\{\boldsymbol{\xi} \in \mathbb{R}^{2I} : C\boldsymbol{\xi} \le \boldsymbol{d}\right\}$, for some matrix $C \in \mathbb{R}^{p \times 2I}$ and vector $\boldsymbol{d} \in \mathbb{R}^p$, where p denotes the number of linear constraints.

Under these assumptions, problem (6) can be equivalently reformulated as the following convex optimization problem:

$$\inf_{\nu, \boldsymbol{s}, \boldsymbol{\Gamma}, \tau, \boldsymbol{w}} \quad \nu\epsilon + \frac{1}{N}\sum_{n=1}^{N} s_n \tag{7}$$

$$\text{s.t.} \quad \boldsymbol{w} \in \mathcal{W}, \tag{8}$$

$$b_k\tau + a_{1,k}\langle \boldsymbol{w}, \boldsymbol{r}_n\rangle + a_{2,k}\langle \boldsymbol{w}, \boldsymbol{z}_n\rangle + \langle \boldsymbol{\gamma}_{n,k}, \boldsymbol{d} - C\hat{\boldsymbol{\xi}}_n\rangle \le s_n, \quad \forall n, k, \tag{9}$$

$$\|C^\top \boldsymbol{\gamma}_{n,k} - (a_{1,k} + a_{2,k})\boldsymbol{w}\|_* \le \nu, \quad \forall n, k, \tag{10}$$

$$\boldsymbol{\gamma}_{n,k} \ge 0, \quad \forall n, k. \tag{11}$$

The set $\boldsymbol{\Gamma} := \{\boldsymbol{\gamma}_{n,k}\}_{n \in \mathcal{N}, k \in \mathcal{K}}$ denotes the collection of vectors $\boldsymbol{\gamma}_{n,k}$, each representing the cost of transporting probability mass from the empirical data point $\hat{\boldsymbol{\xi}}_n$ to adversarial locations within the Wasserstein ball. The non-negativity constraint (11) ensures that this mass redistribution can only increase the worst-case loss, reflecting the adversarial nature of the problem. The scalar variable τ serves as a baseline or threshold in the CVaR representation of the piecewise-linear loss, ensuring the proper evaluation of the worst-case loss across all terms $k \in \mathcal{K}$, rather than directly scaling b_k. The inequalities (9) enforce an upper bound on the worst-case expected loss for every empirical observation $n \in \mathcal{N}$ and term $k \in \mathcal{K}$, while the dual norm constraints, $\|C^\top \boldsymbol{\gamma}_{n,k} - (a_{1,k} + a_{2,k})\boldsymbol{w}\|_* \le \nu$, guarantee that the perturbations remain within the prescribed Wasserstein radius. Here, ν quantifies the maximum sensitivity of the loss to worst-case perturbations, directly linking the robustness of the solution to the Wasserstein ball size. In our case, we can transform problem (4), as explained in Appendix A, into its DRO reformulation by setting $K = 2$ and defining: $a_{1,1} = \left(-\alpha(1-\lambda) - \frac{1-\alpha}{1-\beta}\right)$, $a_{2,1} = (-\alpha\lambda)$, $a_{1,2} = (-\alpha(1-\lambda))$, $a_{2,2} = (-\alpha\lambda)$, $b_1 = (1-\alpha) - \frac{1}{1-\beta}$, $b_2 = (1-\alpha)$, $C = \left[\boldsymbol{I}_I, -\boldsymbol{I}_I, \boldsymbol{I}_I, -\boldsymbol{I}_I\right]^\top$, $\boldsymbol{d} = \left[\boldsymbol{1}_{3I}, \boldsymbol{0}_I\right]^\top$.

Furthermore, following Esfahani and Kuhn (2018) and using the 1-norm to define the Wasserstein distance, the dual norm of a vector $\boldsymbol{y} \in \mathbb{R}^n$ will be computed as

$$||\boldsymbol{y}||_* := \max_{\boldsymbol{x}} \{\langle \boldsymbol{x}, \boldsymbol{y}\rangle \;:\; ||\boldsymbol{x}|| \le 1\} = ||\boldsymbol{y}||_\infty. \tag{12}$$

Under this assumption, the optimization problem (7)-(11) is formulated as a linear programming problem, which can be solved efficiently. From an empirical perspective, the optimization problem is repeatedly solved in a rolling window out-of-sample setting, which allows for the assessment of performance stability over time. At each time t, where the interval between two dates is determined by the investors (e.g., days or months), the optimal solution is computed, and the corresponding strategy is implemented in the real market by adjusting the portfolio. After T periods, the realized returns of the optimal portfolio can be observed, enabling an out-of-sample analysis. We can expect that, for a given combination of parameters, the optimal solutions will vary as we select a different rating agency, and so the observed values $\{z_n\}$. We aim to evaluate how the optimal solution changes when switching from one agency to another. To achieve this, we apply the optimization problem to real market data in the next section and examine the out-of-sample sensitivity of the optimal solution to the chosen parameters.

3 The impact of the choice of the rating agency on optimal portfolios

In this section, we evaluate optimal portfolios using information from different providers considered individually and test the effect of introducing DRO. Hence, we apply the DRO model using ESG data from each rating agency separately. Our goal is to explore the extent of differences in optimal strategies depending on which provider is used to determine the optimal approach. We choose this approach because, if discrepancies between scores from different providers were mild, a robust reformulation might be sufficient to manage the ambiguity.

3.1 Implementation setup

We apply the optimization approach described in the previous section on a subset of constituents of the Eurostoxx Europe 600 index composed by 303 companies which serves as representative of the European stock market. The limitation to 303 companies out of the 600 in the Eurostoxx Europe 600 index is due to the lack of ESG score data for all the rating agencies we have considered. We obtained stock price and ESG data from all five agencies, covering the period from January 1, 2014, to June 28, 2024, with a total of 3,032 daily observations for each asset. We consider five rating agencies[6]: Bloomberg (BL), Refinitiv Eikon (RF), Morningstar Sustainalytics (MN), S&P Global (SP), and Truvalue Labs (TV). The first four entities are well-known in the literature and provide composite ESG scores, which combine the three ESG pillars (E, S, and G). These scores are expressed as a number on a scale within a specific interval, typically[7] $[0, 100]$, where 0 represents poor performance and 100 indicates

[6] Data from Bloomberg and S&P Global were retrieved through the Bloomberg Terminal, Refinitiv Eikon data through the London Stock Exchange Group (LSEG) Workspace, Truvalue Labs data via FactSet, and Morningstar Sustainalytics data were manually collected from Yahoo over the years.

[7] Bloomberg used to use a scale of [0, 10] until recently, but it has transitioned to the [0, 100] scale, like the other providers.

good practice. All these companies derive ESG data from a combination of company disclosures (e.g., annual reports, sustainability reports), third-party sources (news, research), and direct engagement with companies. The ESG score of a company is often normalized by considering the specific industry or sector in which the company operates, benchmarking its performance against peers within the same sector. These companies update their scores on a quarterly basis, with some exceptions due to methodological changes and acquisitions (Fig. 1).

The last agency, Truvalue Labs, uses a unique approach based on artificial intelligence and natural language processing to assess ESG-related insights from publicly available information and news. This process is conducted in real-time, providing continuous updates of ESG data. In particular, Truvalue Labs provides two indicators: The Pulse Score and the Insight Score. The Pulse Score aims to provide a short-term measure of ESG sentiment, while the Insight Score serves as a long-term indicator of ESG performance and controversy management. In this research we use the Insight Score score as it provides a more stable metrics on the long term evolution of sustainability reflecting its overall sustainability performance over time.

The first four rating companies have been the focus of detailed studies, as mentioned in the introduction of this work, highlighting the relative inconsistency of their ratings. It is well known that the distribution of ESG scores across companies varies between providers in terms of range, location, and shape (see Billio et al. 2024 and the references therein) To mitigate these differences, we apply min-max normalization[8] to rescale the ESG scores from each rating agency. This linear transformation preserves the optimal solution of problem (7–11) while standardizing each distribution within the interval [0, 1] based on its minimum and maximum values. This

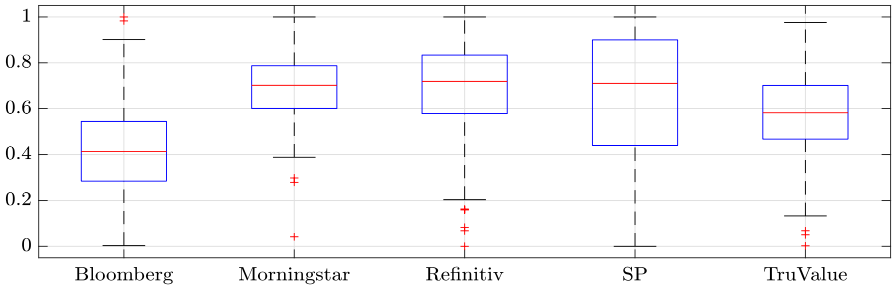

Fig. 1 Box plot of normalized ESG scores for 303 securities in the Eurostoxx Europe 600 index, as rated by the five agencies considered in this study, as of June 1, 2017

[8] Let $E^{(j)}$ be the matrix where the rows represent temporal observations and the columns correspond to the ESG scores of I individual companies provided by the j-th rating agency. Each entry in $E^{(j)}$ contains the observed ESG score for a specific company at a given time. The i-th column of $E^{(j)}$, denoted by the vector $e_i^{(j)}$, represents the time series of ESG scores for the i-th company in the dataset, while $\min E^{(j)}$ and $\max E^{(j)}$ denote the smallest and largest values of the matrix, respectively. The min-max scaling is defined according to the following formula:

$$z_{i,n}^{(j)} := \frac{e_{i,n}^{(j)} - \min E^{(j)}}{\max E^{(j)} - \min E^{(j)}}, \quad i = 1, \ldots, I,$$

approach enhances the comparability of results for a given choice of λ, ensuring that the ESG score matrices from different providers share the same range while preserving the relative ranking of company scores within the same agency.

To obtain a single measure of the distance between the normalized scores from providers i and j over the entire period, we compute the time-averaged Euclidean distance between the two vectors $\hat{z}_n^{(i)}$ and $\hat{z}_n^{(j)}$, normalized by its maximum possible value $\sqrt{I}$:

$$D_{i,j} := \frac{1}{N\sqrt{I}} \sum_{n=1}^{N} \left\| \hat{z}_n^{(i)} - \hat{z}_n^{(j)} \right\|_2, \tag{13}$$

where N is the number of data points over the entire dataset and $\|\cdot\|_2$ denotes the Euclidean norm. Dividing by the maximum possible value $\sqrt{I}$ ensures that the metric is unitless and bounded in [0, 1].

Similarly, we provide a synthetic measure across all providers at a given time n, offering an overview of the dynamics of ESG score dispersion. For each date n, we compute the average pairwise Euclidean distance between normalized ESG score vectors across K providers over I assets:

$$D_n = \frac{1}{\binom{K}{2}\sqrt{I}} \sum_{i<j} \left\| \hat{z}_n^{(i)} - \hat{z}_n^{(j)} \right\|_2, \tag{14}$$

where $\binom{K}{2}$ is the number of unique provider pairs. D_n captures cross-provider disagreement across all assets in the dataset. As before, the distance is normalized by the theoretical maximum Euclidean distance, producing a unitless measure bounded between 0 and 1. A value of 0 indicates perfect agreement among providers, while a value of 1 represents the maximum possible disagreement.

The evolution of D_n, shown in Fig. 2, indicates a gradual decline over time, from approximately 46% to 38%, suggesting a slow convergence of ESG ratings. Both the time-averaged pairwise distance $D_{i,j}$ and the time-specific aggregate distance D_n remain in the range $0.38-0.46$, indicating a substantial and persistent level of disagreement among ESG score providers.

In the next subsection, we provide a general overview of the dataset and explain the ESG data normalization process implemented for portfolio optimization. We adopt a moving window optimization approach with daily rebalancing: on each trading date from June 1, 2017, to June 28, 2024, we solve the portfolio optimization problem, implement the optimal strategy, and measure the out-of-sample results the following day. The decision to trade daily is guided by several key considerations. Firstly, daily trading ensures more reliable out-of-sample data. Given the limited his-

where $z_{i,n}^{(j)}$ represents the normalized ESG score of company i at time n. Note that normalizing the scores to any other interval of real numbers leads to an equivalent optimization, provided the trade-off parameter λ is rescaled accordingly.

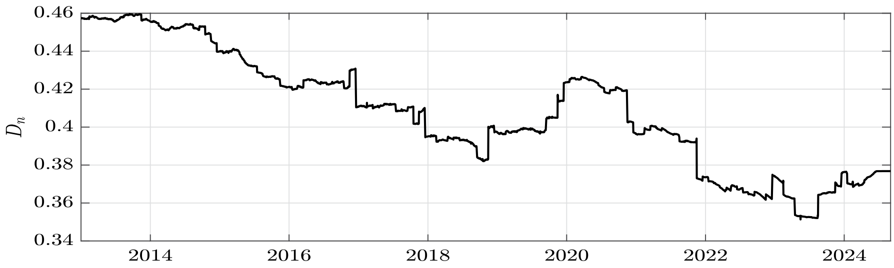

Fig. 2 The distance D_n computed from the ESG scores of the five providers over the entire time horizon of the dataset. Large jumps typically occur when new scores are released

torical data available for ESG metrics, daily optimization helps assess the stability of optimal solutions when both financial and ESG data interact. While most providers update their ESG scores quarterly, stock returns can fluctuate significantly, leading to considerable changes in the objective function, particularly when the parameter λ is set to higher values. Another important factor is the potential impact of the daily ESG indicator provided by Truvalue Labs. This daily update could be crucial for responding swiftly to stock price fluctuations triggered by unexpected news or events related to ESG issues. We also hypothesize that during periods of crisis, characterized by high volatility, investors may prioritize portfolio risk over ESG factors, accepting a trade-off in ESG value for greater stability in financial returns. In such scenarios, prompt portfolio adjustments would be beneficial. Lastly, we argue that daily rebalancing will stress-test the optimization model, especially since we do not impose ad hoc turnover constraints. To evaluate the out-of-sample performance of each strategy, we incorporate transaction costs of 2 basis points for both purchases and sales, which implies 4 basis points for a round-trip trade. We test the model on each trading day within the period from June 1, 2017, to June 28, 2024, for each rating agency and different combination of parameters:

- α, which is inversely related to risk aversion, in the set $\{0, 0.05, 0.1, 0.2, 0.3\}$,
- λ, representing the investor's ESG awareness, in the set $\{0, 0.1, 0.2, 0.3, 0.4\}$,
- ϵ, the radius of the uncertainty ball, in the set $\{0, 0.00025, 0.0005, 0.001\}$,
- β, the CVaR confidence level, in the set $\{0.9, 0.95, 0.99\}$.

We also examine the influence of the time window used to estimate the empirical probability measure by testing three different durations: one, two, and three years, corresponding to $N = 252$, 504, and 756 trading days, respectively. Given the substantial number of parameter combinations, we present results for the case with $\beta = 0.9$ and $\alpha = 0.2$, as we consider this a representative example of the trade-off between expected rewards from both financial and ESG perspectives. Additional results for $\alpha = 0.05$ are provided in Appendix B as an alternative example; the complete set of results is available upon request.

Regarding the time window, the results shown have been obtained using a one-year moving estimation window ($N = 252$) to ensure greater responsiveness to market changes. We include the equally weighted (EW) deterministic strategy as a general benchmark. We will later discuss the sensitivity of the results to variations in these configurations, including changes in β and the time window length, at the end of each section. However, in general, the main conclusions drawn in the following sections remain valid across the entire set of parameter configurations and do not alter the key findings. We emphasize that the analysis is conducted separately for each of the five rating agencies to assess discrepancies in optimal decisions and out-of-sample performance. In Sect. 3.2, we analyze the traditional Mean-CVaR-ESG model that does not contain distributional uncertainty. However, our goal is to evaluate whether the DRO approach, when applied to individual providers, can mitigate ESG discrepancies, leading to similar optimal portfolios regardless of the selected rating agency. This issue is addressed in Sect. 3.3.

3.2 Case $\epsilon = 0$: classical mean-CVaR-ESG model with individual rating providers

In this section, we discuss the results obtained from the traditional Mean-CVaR-ESG historical optimization, where the radius ϵ of the ball around the empirical distribution is zero, and the risk aversion proxy is $\alpha = 0.2$. Table 2 presents the results of the optimal Mean-CVaR-ESG portfolios for increasing values of the parameter controlling the ESG appetite, λ, for each individual provider. The Table displays the key statistics used to compare and evaluate the performance of portfolio strategies: annualized returns[9] (AnnRet), total cost in percentage of the initial wealth (which is the sum of all expenditures required to rebalance the portfolio over the initial wealth), average turnover (AvgTO) over the initial investment, and the average Herfindahl-Hirschman Index (HHI), along with their respective standard deviations (StdTO and StdHHI). Moreover, it reports three widely used risk-reward ratios: the Sortino-Satchel (SS) ratio (Sortino and Satchell 2001), Stable Tail Adjusted Return (STAR) ratio (Martin et al. 2003), and Rachev ratio at the 95% confidence level (Biglova et al. 2004). All statistics are computed based on the sequence of out-of-sample portfolio realizations over the evaluation period. The choice of the first two measures is motivated by the fact that they satisfy all four properties described in Cheridito and Kromer (2013), while the Rachev Ratio, although it does not satisfy the quasi-concavity property, provides information on extreme values, both positive and negative. It also includes CVaR at the 95% confidence level, as well as the Maximum Drawdown (MDD). Table 10 in the Appendix shows results for the case $\alpha = 0.05$.

The first two rows of Table 2 present the results for the Equally Weighted (EW) portfolio and the traditional Mean-CVaR optimal portfolio with no ESG objectives ($\lambda = 0$). These two strategies serve as reference points for comparing the impact of progressively increasing focus on the ESG dimension. To evaluate differences in performance and risk measures between two strategies, we test whether the SS, STAR, Rachev ratio, CVaR, and MDD differ significantly across paired portfolios. Specifically, we employ a Block Bootstrap (BB) procedure in line with the method applied[10] in Ledoit and Wolf (2008) to test the null hypothesis that the statistic $\mathbb{S}(\cdot)$, computed

[9] All the tables present out-of-sample returns that have already been adjusted for transaction costs.

[10] Ledoit and Wolf (2008) applied the BB procedure to compare Sharpe ratios; its performance for Sortino-Satchel, Rachev, and STAR ratios has not been empirically validated.

on the out-of-sample returns of one strategy, equals that of another. The approach has been designed to preserve temporal dependence and heteroskedasticity in returns by resampling overlapping blocks of consecutive out-of-sample observations.[11] A detailed description of the methodology and corresponding results is provided in Sect. 1 of the online appendix. Tables reporting test outcomes show the difference between statistics, with one, two, or three asterisks denoting rejection of the null hypothesis of equality at the 10%, 5%, and 1% significance levels, respectively.

The optimal portfolio with $\lambda = 0$ outperforms the EW strategy both in terms of Risk Reward Ratios (RRRs) and tail measures, with the only exception of the Rachev ratio, and this is true for all the other choices of the α parameter (proxy of the risk aversion) which we do not report in the table for the sake of brevity. The BB test rejects the null hypothesis that the statistics of the optimal strategies equal those of the EW portfolio, although with difference significance levels; see Table 1 in the online appendix. The EW portfolio has an average daily turnover of 0.26% with standard deviation 0.11% which led, under the hypothesis on constant proportional transaction costs, to a total rebalancing expenditure of around 0.49 euro on an initial investment of 100 euro. The traditional optimal portfolios (without ESG objectives) show a higher average daily turnover with a significantly greater standard deviation compared to the EW portfolio. This turnover increases as α rises (i.e., the risk aversion decreases). In the extreme case $\alpha = 1$, where only the expected return matters, the optimal portfolio will exclusively select the asset with the highest past average return. The higher average turnover is reflected in the increased rebalancing costs. Additionally, the concentration index rises considerably compared to the EW portfolio, from 0.33 to 9.02. The latter value remains similar between all the α levels tested and not reported in the table for brevity.

As we gradually increase the λ level, leaving space to the ESG dimension, for each choice of rating provider, we observe a steady deterioration in the performance measures of each rating agency chosen, except for the strategies based on Truvalue scores, which instead shows a consistent increment of the annualized return and RRRs. This result could either be a coincidence or stem from the fact that Truvalue is the only rating agency providing daily updates. As a result, it can offer new insights into the evolution of a company's ESG level before these changes are reflected in the market price. The risk metrics, namely CVaR and MDD, increase under all choices of the rating agency compared to the case of no ESG preferences ($\lambda = 0$). However, they are generally lower than in the EW case, except for MDD in three specific strategies with

Table 1 Pairwise distances $D_{i,j}$ between providers over the entire time horizon

	Bloomberg	Morningstar	Refinitiv	SP	TruValue
Bloomberg	–	0.4418	0.3907	0.4181	0.3956
Morningstar		–	0.4331	0.4335	0.3844
Refinitiv			–	0.4152	0.3768
SP				–	0.3957

[11] We also tested an alternative method for comparing two STAR ratios, recently proposed by Lotfi et al. (2025), which derives the asymptotic null distribution using empirical process theory and the delta method. The results do not differ substantially from those obtained with the Block Bootstrap procedure.

high λ values. At the same time, increasing λ leads to higher values of the Rachev ratio. The BB test confirms the increase in CVaR as λ rises for every rating agency, but it yields mixed results for the Rachev ratio and MDD (see Table 2 in the online appendix). For example, the test indicates statistically significant increases in MDD as λ grows when the Bloomberg and Morningstar ESG scores are used, but not when the Refinitiv and S&P scores are considered. When α is strictly positive, we observe that as ESG appetite λ increases, both the average turnover of optimal strategies and its standard deviation consistently decrease. This, in turn, leads to a reduction in total transaction costs. In contrast, increasing λ results in a higher concentration index, as the optimizer considers both the financial and ESG properties of each asset, focusing on the few assets that perform well in both metrics. In general, for fixed values of α and λ, the choice of rating agency consistently influences the out-of-sample optimal

Table 2 Summary statistics for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024

	AnnRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	Std-HHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
λ = 0.0											
No ESG	8.95	5.97	5.62	9.79	9.02	2.81	6.67	2.03	82.21	1.77	24.01
λ = 0.1											
BL	5.44	2.95	3.24	6.25	24.06	9.89	3.94	1.19	79.08	2.03	32.31
MN	5.53	4.30	4.79	7.65	13.28	4.18	4.25	1.34	87.02	1.80	25.04
RF	4.92	4.36	5.07	8.11	11.16	2.98	3.85	1.18	83.69	1.85	27.94
SP	5.20	4.26	4.91	7.98	11.38	3.07	4.03	1.23	83.07	1.86	26.41
TV	11.97	5.89	4.52	7.41	13.31	4.41	8.33	2.51	88.82	1.89	25.36
λ = 0.2											
BL	3.85	1.65	1.89	4.52	39.39	14.82	2.80	0.84	83.97	2.27	36.86
MN	4.28	2.40	3.23	6.21	20.08	10.76	3.11	0.97	87.75	2.10	31.65
RF	3.16	2.88	3.59	6.43	15.47	5.18	2.62	0.80	86.10	1.92	27.52
SP	4.08	3.87	4.57	7.73	12.10	3.45	3.30	1.01	84.50	1.84	25.63
TV	13.16	3.54	2.63	4.94	21.58	5.50	8.24	2.52	92.99	2.08	27.20
λ = 0.3											
BL	4.69	1.25	1.40	3.69	47.63	16.79	3.11	0.94	87.32	2.47	39.63
MN	4.69	1.75	2.45	4.84	26.98	14.41	3.21	1.00	88.18	2.26	38.37
RF	3.12	2.46	3.10	5.85	17.18	5.44	2.55	0.79	89.00	1.96	27.29
SP	3.06	3.38	4.26	7.26	12.87	4.05	2.60	0.79	85.57	1.87	25.56
TV	13.99	2.83	1.97	4.11	28.87	9.09	8.13	2.51	96.44	2.24	28.01
λ = 0.4											
BL	4.60	0.95	1.07	3.17	52.70	18.83	3.01	0.91	88.33	2.56	40.57
MN	3.60	1.31	2.03	4.46	33.67	15.07	2.50	0.78	86.23	2.46	43.14
RF	3.10	2.33	2.91	5.67	18.75	5.36	2.50	0.78	88.55	2.02	27.43
SP	3.47	3.20	4.08	7.18	13.45	3.49	2.86	0.87	87.14	1.90	25.63
TV	13.46	2.57	1.86	4.03	35.46	12.05	7.54	2.33	98.62	2.36	33.26

Results are presented for each choice of rating agency used for the ESG value in the optimization and for four values of the parameter λ. All optimizations in the table follow the same parameter configuration: $\alpha = 0.2$, $\beta = 0.90$, $\epsilon = 0$, and $N = 252$. When $\lambda = 0$, the optimizer does not consider the ESG value, resulting in a single identical row across all rating agencies

Table 3 Out-of-sample average (standard deviation) of the ESG scores for the EW portfolio and optimal strategies with $\lambda = 0$ and two different values of the parameter ϵ and α (rows), evaluated with respect to each rating agency (columns)

	BL	MN	RV	SP	TV
EW	41.8 (5.2)	78.9 (1.2)	72.1 (2.9)	72.5 (8.8)	58.6 (2.9)
α = 0.05					
ε = 0.00	40.8 (5.4)	80.1 (1.4)	71.4 (4.5)	71.9 (11.3)	59.4 (2.6)
ε = 1.00	41.1 (5.4)	79.8 (1.1)	71.2 (4.4)	71.8 (10.8)	59.3 (2.5)
α = 0.2					
ε = 0.00	40.7 (5.5)	80.1 (1.5)	71.3 (4.6)	71.6 (11.6)	59.5 (2.6)
ε = 1.00	41.0 (5.0)	79.7 (1.1)	71.1 (4.1)	71.6 (10.7)	59.3 (2.1)

The values for the parameter ϵ are expressed in thousands

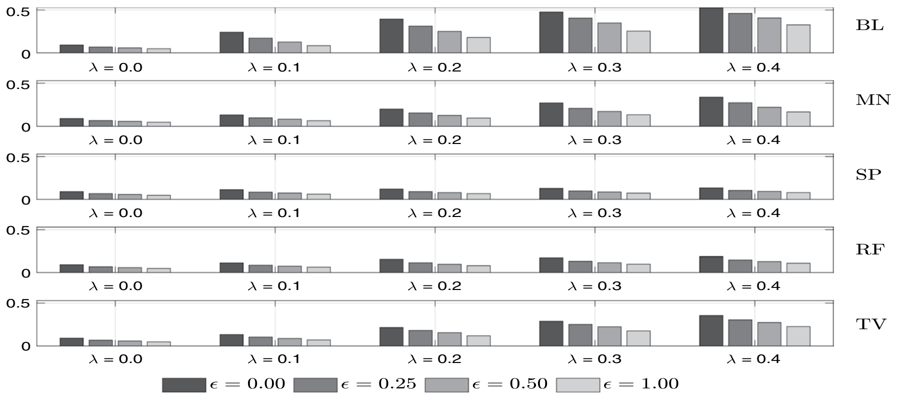

Fig. 3 The Herfindahl-Hirschman Index (HHI) for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the HHI of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.2$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

portfolios. Table 3 in the online appendix reports the BB test results for pairs of optimal strategies with identical parameter values but differing in the choice of rating agency. Overall, the test confirms that the strategies differ in terms of CVaR and, in many cases, also in MDD, whereas the three ratios (Sortino, STAR, and Rachev) are statistically indistinguishable with the exception of Truvalue, which always differs.

3.3 Case $\epsilon > 0$: mean-CVaR-ESG model with individual providers under distributional uncertainty

We now investigate the impact on optimal strategies as we progressively assign strictly positive values to ϵ. In particular, we analyze the effect of different choices of the DRO parameter ϵ on the following statistics: i) the average concentration index, ii) the average turnover of the strategy, and iii) the STAR ratio. To illustrate this, we present the values of each strategy for these statistics in Figs. 3, 4, and 5, respectively. Each vignette show the values of one particular statistics of the optimal strategy for four different choices of the parameter ϵ, where each row is relative to a particular ESG rating agency used to compute the optimal strategy for increasing values of ESG

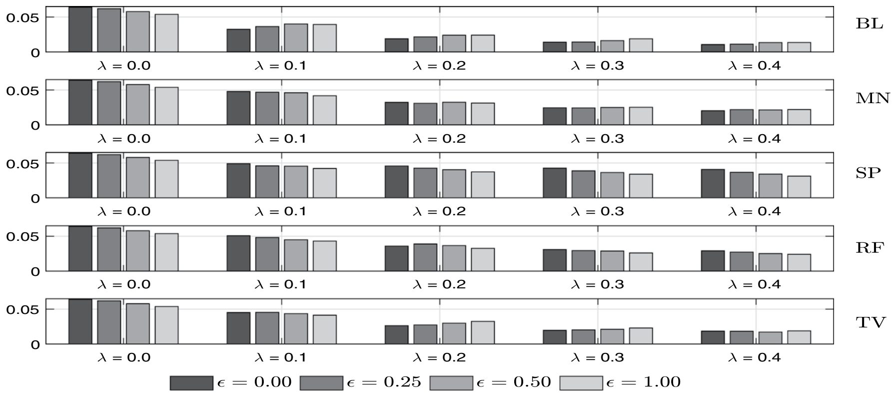

Fig. 4 The Average Turnover for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the Average Turnover ratio of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.2$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

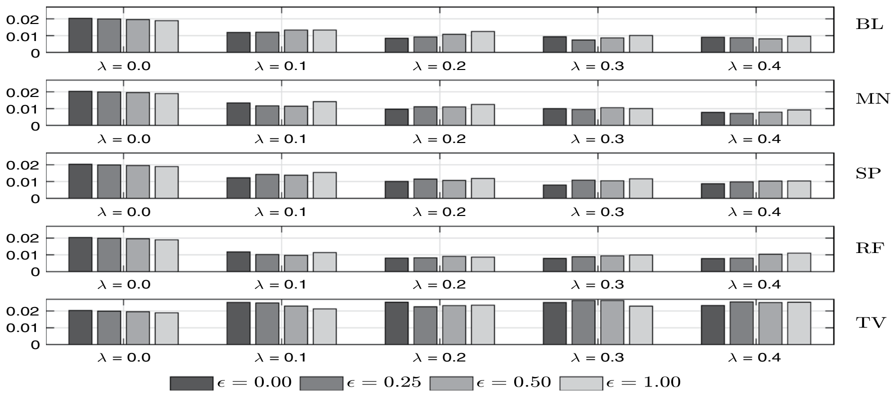

Fig. 5 The STAR ratio for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the STAR ratio of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.2$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

appetite. As before, the analogous results for $\alpha = 0.05$ are stored in Figs. 8, 9 and 10 in Appendix B.

The first implication of considering a DRO approach is that the concentration of the portfolio decreases. This is in line with the theory, which ensure a convergence toward the equally weighted portfolio for ϵ approaching infinity. This is particularly important when investors have a medium-high attitude toward the ESG score, i.e., strictly positive λ, as the classical Mean-CVaR-ESG optimization presented in the previous section produces highly concentrated optimal portfolios.

The effect of increasing ϵ on the average turnover varies depending on the values assigned to α and λ. When $\alpha = 0.2$, increasing λ significantly reduces the average turnover, although the effects are mixed for different choices of ϵ. However, when $\alpha = 0.05$, the average turnover decreases as ϵ increases, for all values of λ and for all rating agencies. Indeed, as α increases, the impact of a positive λ on the solution of the optimization problem is higher, and the concentration of the portfolio increases, as the optimizer assigns more weights on the ESG component of the objective function. This will in turn also decrease the transaction cost of the strategy.

The impact of increasing ϵ on the STAR ratio is more complex. As already discussed, increasing the parameter λ while keeping all other parameters fixed leads to a decrease in the performance measure. The only exception occurs with Truvalue's scores, which generally produce higher STAR ratios as the ESG parameter increases. This aligns with the broader disagreement in the literature regarding the effect of incorporating ESG preferences on the financial performance of the optimal portfolio. Fixing all parameters and varying ϵ does not produce consistent results: in some cases, we observe an increase in the performance measure, while in others, we see a reduction. However, in all cases, the effect is relatively small, allowing us to conclude that increasing the radius of the uncertainty ball around the empirical distribution function primarily affects the concentration index, rather than the STAR ratio. This result suggests that, in general, we can achieve similar financial reward/risk profiles with portfolios that exhibit significantly different concentration measures. We will come back to this point later in Sect. 4.

3.4 ESG discrepancy and impact on portfolio weights

It is relevant to examine the average ESG score of each optimal portfolio across all five rating providers. Let $\boldsymbol{w}_{\boldsymbol{t}}^{(i),*}(\lambda, \epsilon, \alpha)$ be the optimal portfolio weights obtained as the solution to problem (7)-(11) for given values of the parameters λ, ϵ, and α, at date t of the out-of-sample analysis, using the ESG scores of the i-th rating agency. We can now compute the average ESG score of the optimal portfolios over the out-of-sample horizon $t = 1, \dots, T$ with respect to the ESG scores of any other rating agency $j \in \mathcal{J}$, namely:

$$\bar{z}_{i,j}(\lambda, \epsilon, \alpha) := \sum_{t=1}^{T} \boldsymbol{z}_{\boldsymbol{t}}^{(j)} \boldsymbol{w}_{\boldsymbol{t}}^{(i),*}(\lambda, \epsilon, \alpha), \quad \text{for some } i, j \in \mathcal{J}. \tag{15}$$

Of course, when $\lambda = 0$, the ESG scores do not enter into the optimization problem, and the average ESG score $\bar{z}_{i,j}(0, \epsilon, \alpha)$ for some $j \in \mathcal{J}$, is the same for every $i \in \mathcal{J}$. We can assess the improvement in the average ESG score of the optimal portfolios obtained using the scores of the i-th agency when setting a higher value of λ by considering its relative variation with respect to the case of $\lambda = 0$:

$$\delta_{i,j}(\lambda) := \frac{\bar{z}_{i,j}(\lambda, \epsilon, \alpha)}{\bar{z}_{i,j}(0, \epsilon, \alpha)} - 1. \tag{16}$$

Table 3 shows the average ESG score $\bar{z}_{i,j}(0, \epsilon, \alpha)$ and its standard deviation (in parentheses) for the extreme values of the parameter ϵ (rows) considered in the experiment. We emphasize that these figures represent the out-of-sample average ESG scores of the optimal Mean-CVaR-ESG portfolio selected with no attention to ESG (i.e., $\lambda = 0$) and $\alpha = 0.2$, providing a baseline for future comparisons. The average ESG score, according to a given rating agency, remains generally stable for all values of ϵ considered, with the only observable effect being a marginal reduction in its standard deviation. This is interesting because, in general, increasing ϵ leads to optimal strategies with different concentration indices, yet the average ESG score remains largely unaffected. It is also noteworthy that the average ESG score of the two optimal portfolios reported in Table 3 is consistently close to that of the equally weighted (EW) portfolio, regardless of the rating agency chosen.

The values in Table 3 will then be used to compute the relative variations in percentage, $\delta_{i,j}(\lambda)$, for $i, j \in \mathcal{J}$, and $\lambda \in \{0.1, 0.2, 0.3, 0.4\}$, as reported in Table 4. The left panel of Table 4 compares the relative variation of the average ESG score of each optimal portfolio for different models (rows) and rating agencies (columns) for

Table 4 Table The left panel shows the average ESG relative variation $\delta_{i,j}(\lambda)$, as defined in eq. (16), of the optimal portfolios with respect to the case $\lambda = 0$

	ESG relative variation (%)					Average turnover (%)				
	BL	MN	RF	SP	TV	BL	MN	RF	SP	TV
	λ = 0.1					λ = 0.1				
BL	60.6	0.4	12.6	19.8	2.4	0.0	91.5	84.0	85.1	91.3
MN	3.0	8.8	2.6	7.9	− 1.6	91.5	0.0	80.9	79.9	83.4
RF	13.0	− 1.4	21.2	23.1	− 1.6	84.0	80.9	0.0	59.0	85.9
SP	10.8	− 0.3	12.9	33.1	− 1.0	85.1	79.9	59.0	0.0	83.8
TV	4.9	− 0.1	− 1.7	− 0.6	19.5	91.3	83.4	85.9	83.8	0.0
	λ = 0.2					λ = 0.2				
BL	72.6	2.2	13.1	26.9	4.1	0.0	98.8	88.6	90.2	97.1
MN	3.2	12.7	4.0	8.3	− 3.8	98.8	0.0	93.9	87.9	95.9
RF	19.3	− 2.8	26.4	29.0	− 2.3	88.6	93.9	0.0	65.2	95.3
SP	12.7	− 0.6	15.9	35.8	− 1.1	90.2	87.9	65.2	0.0	94.6
TV	3.6	− 0.01	− 2.3	− 0.6	26.1	97.1	95.9	95.3	94.6	0.0
	λ = 0.3					λ = 0.3				
BL	76.5	3.0	12.7	28.2	4.8	0.0	99.6	91.5	92.0	97.1
MN	5.4	14.6	5.1	6.4	− 4.4	99.6	0.0	97.5	93.5	98.9
RF	18.9	− 3.3	27.8	29.7	− 3.0	91.5	97.5	0.0	68.7	96.3
SP	14.0	− 0.9	17.2	37.0	− 0.9	92.0	93.5	68.7	0.0	97.8
TV	3.7	0.1	− 2.4	0.6	28.9	97.1	98.9	96.3	97.8	0.0
	λ = 0.4					λ = 0.4				
BL	78.5	3.4	12.8	28.9	5.0	0.0	100.0	92.5	93.7	97.9
MN	7.2	15.5	6.0	4.9	− 4.5	100.0	0.0	98.9	96.0	99.1
RF	18.4	− 3.6	28.5	29.6	− 3.2	92.5	98.9	0.0	71.9	96.2
SP	14.5	− 1.1	17.9	37.6	− 0.8	93.7	96.0	71.9	0.0	98.3
TV	2.4	0.1	− 1.9	1.1	30.4	97.9	99.1	96.2	98.3	0.0

The right panel shows the average turnover between Mean-CVaR-ESG optimal portfolio strategies when switching the optimization from one rating agency to another, for different values of λ. In all the cases the other parameters are $N = 252$, $\beta = 0.90$, $\epsilon = 0$, $\alpha = 0.2$. The rating agencies considered are Bloomberg (BL), Refinitiv Eikon (RF), Morningstar Sustainalytics (MN), S&P Global (SP), and Truvalue Labs (TV) the four different values of λ and with $\epsilon = 0$. Each row represents the optimal strategy obtained by optimizing based on a specific provider's ESG scores, while the columns show the out-of-sample relative variation of average ESG scores with respect to the case $\lambda = 0$. On the diagonals we find the relative variation of the average ESG score with respect to the optimal strategy obtained within the same rating agency. For instance, the first row in the left panel shows the relative variation of the average ESG score with respect to all rating agencies (columns) of the optimal strategy obtained with the ESG scores from Bloomberg: while the average Bloomberg ESG score of the strategy increases of 60.6% when λ rises from 0 to 0.1, the improvement in the Morningstar Sustainalytics average score is irrelevant (0.4%).

We observe a certain agreement between BL, RF, and SP: when the portfolio is optimized using one of these providers, the ESG score of the portfolio tends to increase for the others as well. However, in some cases, optimizing with one rating agency may lead to a decrease in the portfolio's ESG score according to another provider. The only exception is Bloomberg: when the portfolio is optimized using its ESG scores, the average ESG score improves compared to the case of no ESG awareness, regardless of the provider chosen.

However, in general, optimizing the portfolio based on the ratings of one agency does not necessarily result in an increase in the ESG value of the portfolio according to the ratings of another agency. Similar results for $\alpha = 0.05$ are presented in Table 11 in Appendix B.

Another important aspect of our analysis is assessing the extent to which the weights of the optimal strategy vary depending on the choice of the rating agency used to optimize the portfolio. At each date t, for a given set of parameters $(\lambda, \epsilon, \alpha)$, we obtain J different optimal weight vectors:

$$\boldsymbol{w}_t^{(j),*}(\lambda, \epsilon, \alpha), \quad \text{for} \quad j = 1, \dots, J. \tag{17}$$

From these, we can calculate the average potential turnover required to switch between each pair of optimal strategies as follows:

$$\Delta_{j,k}(\lambda, \epsilon, \alpha) := \sum_{t=1}^{T} \frac{\boldsymbol{e}^\top |\boldsymbol{w}_t^{(j),*}(\lambda, \epsilon, \alpha) - \boldsymbol{w}_t^{(k),*}(\lambda, \epsilon, \alpha)|}{2}, \text{ for } j = 1, \dots, J \text{ and } k < j. \tag{18}$$

These values are reported in the right panel of Table 4 for the cases $\alpha = 0.2$, $\epsilon = 0$ and $\lambda \in \{0.1, 0.2, 0.3, 0.4\}$. The case with $\alpha = 0.05$ is in the Appendix in Table 11. It is clear that the average portfolio composition changes significantly when switching to another rating agency, even when the values of the parameters α and λ are fixed. For instance, when $\alpha = 0.2$, the average turnover between the strategies exceeds 80%, even for the lowest value of the ESG parameter, $\lambda = 0.1$. Even with low values of α (e.g., 0.05), the average relative variation ranges from 33% for $\lambda = 0.1$ to 87.2% for $\lambda = 0.4$.

To further evaluate the differences between optimal strategies, we compute, for each company, the average portfolio weight over the entire out-of-sample period. This is obtained by summing the weights across all dates and normalizing by the number of periods T, yielding the vector

$$\bar{\boldsymbol{w}}^{(j),*}(\lambda,\epsilon,\alpha)=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{w}_t^{(j),*}(\lambda,\epsilon,\alpha).$$

Figure 6 displays, in each panel, the 10 companies with the largest average portfolio weights under a specific parameter configuration: rating agency, $\epsilon=0$, $\alpha=0.2$, and $\lambda=0.4$. Each company is assigned a distinct color according to the FactSet Revere Business Industry Classification System (RBICS). The results highlight that the composition of the portfolio varies substantially across strategies.

Increasing the uncertainty radius ϵ does not significantly reduce the variability between optimal strategies computed under the same parameter settings but with different ESG scores (results are omitted for brevity). This finding highlights the challenge of implementing an optimal ESG portfolio strategy when there is uncertainty regarding which rating agency to select or trust.

We can summarize the result of this section as follows: i) the choice of the ESG rating agency leads to significantly different optimal strategies, even for moderate values of the λ; ii) improving the average ESG value of the portfolio with respect to the scores of one rating agency does not imply that an improvement occurs for the scores of any other ESG rating agency, as in some cases the latter evaluation decreases; iii) the DRO approach does not help reduce this discrepancy, but it effectively reduces portfolio concentration while maintaining similar levels of financial performance, as shown by the STAR ratio analysis in Fig. 5. These conclusions remain valid for any other parameter combination considered, including the choice of the time window.

These results highlight the need for statistics based on the ESG evaluations from all providers, rather than relying on the scores from a single rating agency, $\boldsymbol{z}_t^{(j)}$, in the optimization problem (7)-(11). The next section is dedicated to discussing the benefits of using a statistic computed from the ESG scores of all J providers.

4 DRO portfolios on ESG indeces

In each optimization problem considered so far, a single ESG rating agency has been selected, and the optimization has been run using that agency's scores. Ambiguity among ESG ratings was addressed by constructing an uncertainty ball around an observed data matrix $\hat{\Xi}^{(j)}$, for some j in $\mathcal{J}$, separately for each provider. We observed that the distributionally robust approach produces less concentrated portfolios while maintaining similar financial performance and ESG profiles. However, the resulting optimal strategies are substantially affected by the choice of rating agency, as reflected in the portfolio weights.

Aggregating data from multiple sources provides a practical means of addressing the complexities inherent in ESG assessment, reducing the impact of methodological differences and temporal variability. Consequently, this approach offers stakeholders a more consistent and comparable basis for decision-making. An alternative is to use a statistic computed from the ESG scores of all J providers, thereby supplying the optimizer with aggregated ESG values that incorporate information from multiple

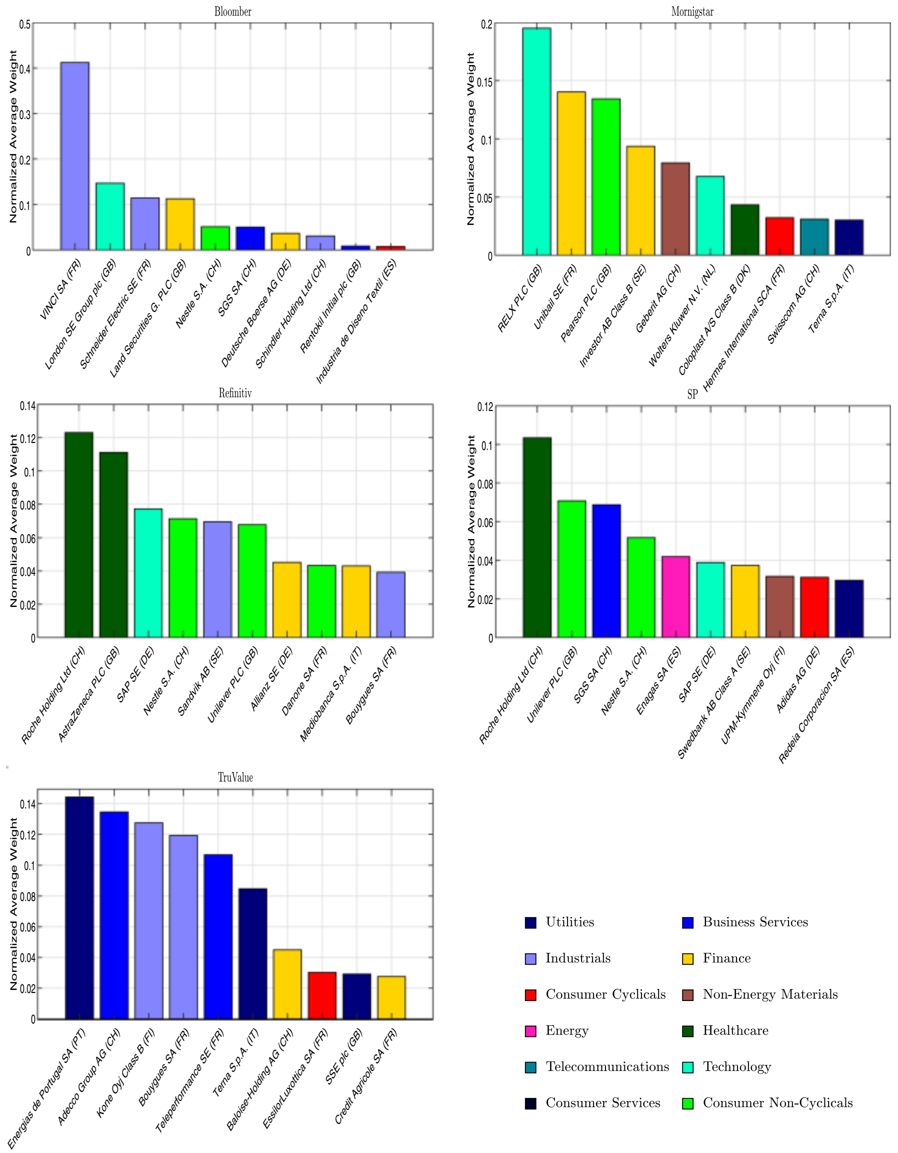

Fig. 6 Each panel shows the relative contribution of ten assets with highest weights along the entire out-of-sample horizon, obtained by compute optimal portfolios with ESG scores provided by a particular rating agency. The color of the bar reflects the industry accordingly to the FactSet's Revere Business Industry Classification System (RBICS)

sources. The aim is to obtain an optimal portfolio whose ESG score exceeds that of any individual provider. Various aggregation approaches have been proposed in the literature to address ESG divergence among rating agencies, although their focus is not on portfolio optimization. For a recent review, see Agosto and Tanda (2025). A common approach is the equal-weighted average, which assigns the same importance to each provider when computing the composite score. A refinement is the voting-average method, which adjusts weights according to the degree of consensus among agencies Bissoondoyal-Bheenick et al. (2024). More recent studies propose an optimized-weighting method, where weights are determined via linear optimization to maximize the correlation between the aggregated ESG score and subsequent one-year excess stock returns. Data-driven techniques can also be used to derive aggregated measures from multiple inputs. For instance, Principal Component Analysis (PCA) extracts uncorrelated components that capture the linear relationships among observed variables. In the ESG context, the first principal component is often adopted as a synthetic ESG score, summarizing information from multiple providers while accounting for inter-provider correlations Bissoondoyal-Bheenick et al. (2024), Gucciardi et al. (2025). Gai et al. (2023) propose the Multidimensional Synthesis of Indicators (MSI) method for aggregating ESG indicators that are not fully substitutable. MSI combines multiple ESG metrics into a single score while limiting compensations between dimensions, thereby providing a more nuanced assessment of corporate ESG performance and helping to detect potential greenwashing. Finally, Agosto et al. (2023) adopt a Bayesian approach to obtain an aggregated ESG indicator by integrating ratings from different providers. Their weighting procedure is data-driven, relying on the empirical relationship between ESG performance and creditworthiness, as measured by credit ratings issued by recognized agencies.

In what follow we test two simple aggregation approaches to maintain tractability, as this is the first study to address both financial and ESG uncertainty within a distributionally robust optimization (DRO) framework. The first approach uses the average aggregation of ESG scores across different providers, while the second considers a linear combination of the average and standard deviation, with the aim of penalizing companies whose ESG scores vary substantially across providers.

4.1 ESG average aggregation

The first approach assigns to each asset i the mean of its ESG scores at a given time n, defined as

$$\bar{\mu}_{i,n} := \frac{1}{J}\sum_{j=1}^{J} z_{i,n}^{(j)}, \tag{19}$$

and collects these values for all assets into the vector $\bar{\boldsymbol{\mu}}_n$. Table 5 presents the same quantities reported for the out-of-sample results in the previous section, but with the ESG scores of the J providers replaced by the mean index $\bar{\boldsymbol{\mu}}_n$ for the case $\alpha = 0.2$. Results for the case $\alpha = 0.05$ can be found in Table 12 in the Appendix. We show results for all four choices of the distributionally robust parameter ϵ in the same table, expressed in thousands. We observe that when $\alpha = 0.2$, the annualized return exceeds the benchmark only for $\lambda \in [0, 0.1]$ (moderate ESG appetite). When $\alpha = 0.05$ the annualized return of optimal portfolios is always superior to the benchmark (the EW portfolio) for all the choice of λ.

Table 5 Summary statistics and risk/reward ratios for out-of-sample daily optimal portfolios from 1-June-2017 to 28-June-2024 using the ESG average index $\bar{\mu}_n$ among the ESG scores of each company and for different values of ϵ

	AnnRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	StdHHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
ε	λ = 0.0						λ = 0.0				
0.00	8.86	6.84	6.39	9.48	9.12	2.91	6.67	2.03	82.21	1.77	24.01
0.25	8.68	6.67	6.20	8.96	6.75	2.37	6.62	1.99	81.97	1.77	23.89
0.50	8.48	6.16	5.79	8.48	5.78	1.77	6.50	1.96	81.51	1.76	24.41
1.00	8.11	5.64	5.38	7.68	4.80	1.39	6.27	1.89	81.61	1.75	24.57
	λ = 0.1						λ = 0.1				
0.00	8.79	5.21	5.13	8.23	11.31	2.78	6.29	1.96	85.95	1.84	26.84
0.25	7.46	5.09	5.22	8.00	8.56	2.23	5.46	1.69	84.53	1.84	27.52
0.50	6.52	4.74	4.99	7.73	7.38	2.03	4.91	1.51	84.54	1.83	27.74
1.00	6.53	4.47	4.65	7.11	5.91	1.55	5.00	1.55	84.11	1.77	27.15
	λ = 0.2						λ = 0.2				
0.00	7.54	3.74	3.89	6.84	16.52	5.51	5.26	1.62	82.92	1.97	29.37
0.25	7.52	3.72	3.87	6.90	11.95	2.97	5.33	1.64	83.84	1.93	30.20
0.50	6.19	3.71	4.00	6.67	9.78	2.42	4.52	1.40	82.93	1.91	30.57
1.00	5.35	3.65	4.06	6.68	7.68	2.01	3.99	1.23	82.64	1.92	32.96
	λ = 0.3						λ = 0.3				
0.00	6.07	2.77	3.17	5.90	21.41	6.75	4.16	1.28	80.46	2.10	32.49
0.25	5.58	2.78	3.17	5.81	15.61	3.88	3.96	1.22	82.58	2.03	31.64
0.50	4.85	2.79	3.22	5.90	12.89	3.12	3.51	1.08	81.25	2.05	32.97
1.00	4.64	2.95	3.41	6.06	9.92	2.58	3.41	1.05	82.66	2.03	33.77
	λ = 0.4						λ = 0.4				
0.00	5.68	1.95	2.36	4.93	25.97	9.17	3.84	1.19	83.91	2.16	34.37
0.25	5.65	2.16	2.50	5.02	19.89	6.20	3.86	1.21	85.39	2.12	34.98
0.50	6.01	2.17	2.45	5.02	16.68	5.17	4.10	1.28	85.69	2.09	35.10
1.00	4.92	2.27	2.67	5.23	12.23	2.64	3.46	1.07	82.82	2.12	35.64

All the optimization in the table have the significance level of the CVaR $\beta = 0.90$, $\alpha = 0.2$ and $N = 252$

In general, with only a few exceptions, increasing ϵ while keeping the other parameter values fixed leads to a decrease in annualized returns. However, we observe that the high annualized return of the EW portfolio is associated with a substantially higher downturn risk, as measured by both CVaR_{95} and the maximum drawdown. The optimized portfolio generally outperforms the EW portfolio in terms of SS and STAR ratios, except under strong ESG preferences (i.e., $\alpha = 0.2$ and $\lambda > 0.2$), although the BB test detects a statistically significant difference only when $\lambda = 0$. In line with expectations, CVaR is consistently lower for the optimized strategies than for the EW portfolio, while MDD is also lower provided that $\lambda < 0.3$. The turnover, and consequently the cost relative to the initial investment, also decreases with ϵ, except in high ESG attitude scenarios where both α and λ take relatively high values. At the same time, both the average and standard deviation of the concentration index decrease considerably. In some cases, concentration is more than halved when comparing the extreme values of ϵ considered. We can note that increasing the parameter

ϵ does not substantially affect the values of all the statistics in Table 5, meaning that we can achieve similar performance and risk measures but reducing the concentration of the portfolio. This finding is confirmed by the BB test when applied to pairs of strategies that share identical parameters except for ϵ, as reported in Table 6 of the online appendix. In most cases, the test does not reject the null hypothesis of equality between the two statistics; however, there are occasional exceptions, particularly for CVaR, which tends to be statistically lower when ϵ is larger.

The average ESG values of optimal portfolios across the five providers and for different levels of λ and ϵ are shown in Table 6. The relative variation compared to the case without ESG preferences is presented in Table 7 (Table 13 for $\alpha = 0.05$ is provided in the Appendix). We observe that using the average index $\bar{\mu}_n$, instead of a single ESG rating score, produces optimal portfolios whose average ESG score increases across all rating agencies as λ increases. Of course, these results come at a cost, which can be observed by comparing the left-panels of Tables 4 with 6. For a fixed choice of λ, the average ESG score of the optimal portfolio obtained using the ESG index is lower than that obtained by directly optimizing with the score of a specific rating agency $j \in \mathcal{J}$.

This difference becomes more pronounced as the ESG attitude increases and also depends on the distribution of scores for each agency. To illustrate this, we compare the values in Tables 7 and 4 for the corresponding values of λ. For instance, when $\alpha = 0.2$ and $\lambda = 0.4$, optimizing with Bloomberg ESG scores produces an optimal portfolio with a Bloomberg ESG score of 63.3 and a standard deviation of 5 points.

Table 6 Out-of-sample average ESG portfolio scores (standard deviation) for different models (rows) with respect to each rating agencies (columns) obtained with the ESG average index $\bar{\boldsymbol{\mu}}_n$ for $\alpha = 0.2$ and different choices of the ESG weights λ and the uncertainty radius ϵ expressed in thousands (rows)

	BL	MN	RV	SP	TV	BL	MN	RV	SP	TV
ε	λ = 0.1					λ = 0.3				
0.00	55.3	83.0	82.6	92.4	63.6	63.9	85.4	83.8	96.0	66.4
	(4.3)	(1.8)	(3.1)	(3.1)	(2.2)	(4.8)	(2.9)	(2.4)	(2.5)	(2.3)
0.25	53.9	82.8	82.3	91.9	63.5	63.5	85.1	83.7	95.9	66.2
	(4.0)	(1.3)	(2.5)	(3.1)	(2.1)	(4.0)	(2.8)	(1.9)	(2.3)	(1.9)
0.50	53.2	82.6	82.2	91.5	63.3	62.9	85.0	83.7	95.9	66.1
	(3.6)	(1.2)	(2.2)	(3.2)	(2.0)	(3.5)	(2.7)	(1.7)	(2.0)	(1.9)
1.00	52.3	82.6	81.8	91.0	63.2	61.8	84.8	83.7	95.9	66.1
	(3.2)	(1.1)	(1.8)	(3.2)	(1.8)	(2.9)	(2.3)	(1.6)	(1.9)	(2.1)
	λ = 0.2					λ = 0.4				
0.00	61.8	84.3	84.2	95.7	65.5	65.3	86.2	83.3	96.2	67.0
	(4.3)	(2.8)	(2.2)	(2.1)	(1.8)	(5.1)	(2.7)	(2.5)	(2.6)	(2.8)
0.25	60.8	84.2	84.0	95.7	65.4	64.9	85.8	83.3	96.0	66.8
	(3.7)	(2.3)	(1.8)	(1.8)	(1.8)	(4.4)	(2.8)	(2.2)	(2.4)	(2.8)
0.50	59.6	84.1	83.8	95.6	65.3	64.4	85.6	83.4	95.8	66.5
	(3.7)	(2.0)	(1.6)	(1.6)	(2.1)	(3.9)	(2.8)	(1.9)	(2.3)	(2.3)
1.00	57.8	84.0	83.6	95.4	65.1	63.4	85.3	83.5	95.8	66.5
	(3.6)	(1.6)	(1.4)	(1.7)	(2.2)	(3.3)	(2.5)	(1.7)	(2.0)	(2.3)
	α = 0.05					α = 0.2				

All the optimizations in the table have the significance level of the CVaR $\beta = 0.90$ and $N = 252$

Table 7 Out-of-sample relative variation of the average ESG portfolio scores obtained with the ESG average index $\bar{\mu}_n$ with respect to the case $\lambda = 0$, $\alpha = 0.2$, for different values of ϵ expressed in thousands (rows)

	ESG Relative Variation (%)						ESG Relative Variation (%)				
	BL	MN	RV	SP	TV		BL	MN	RV	SP	TV
ε	λ = 0.1						λ = 0.3				
0.00	36.0	3.6	16.0	29.3	6.8	0.00	57.2	6.6	17.6	34.3	11.6
0.25	32.7	3.4	15.5	28.6	6.6	0.25	56.1	6.3	17.5	34.2	11.1
0.50	30.9	3.2	15.3	28.1	6.4	0.50	54.7	6.1	17.4	34.2	11.1
1.00	28.7	3.2	14.9	27.4	6.1	1.00	52.0	5.9	17.4	34.2	11.1
	λ = 0.2						λ = 0.4				
0.00	52.0	5.3	18.2	33.9	10.1	0.00	60.6	7.6	16.9	34.6	12.6
0.2	49.6	5.1	17.9	33.9	9.9	0.2	59.7	7.2	16.9	34.3	12.2
0.50	46.6	5.2	17.6	33.7	9.7	0.50	58.5	6.9	17.0	34.1	11.7
1.00	42.2	5.0	17.6	33.4	9.3	1.00	55.9	6.6	17.2	34.1	11.6

When running the optimization with the ESG index at $\epsilon = 0$, the average ESG score is 55.8 with a standard deviation of 4.2. Although the latter value is lower than the former, it is important to note that optimizing with any other provider would result in an average Bloomberg ESG score ranging from 42.0 to 45.5, well below what is achieved with the index. Increasing the value of ϵ generally results in a lower average ESG score for the optimal portfolio based on the ESG index, but with a reduced standard deviation. The extent of this reduction varies across providers, as the distribution of ESG scores differs significantly among them, and on the level of λ. Furthermore, this average reduction is typically very small, on the order of 2–3 points, which allows us to conclude that increasing the uncertainty radius does not significantly affect the average ESG score, while it slightly does reduce its standard deviation.

4.2 Dispersion-adjusted ESG aggregation

The simple average $\bar{\mu}_{i,n}$ provides a straightforward cumulative ESG measure for the i-th asset across all providers; however, it does not capture the dispersion of ESG scores among rating agencies. Although the DRO approach accounts for uncertainty in the distribution of $\bar{\mu}_n$, investors may also benefit from penalizing assets that exhibit high dispersion in their scores across providers. For example, let $\bar{\sigma}_{i,n} := \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(z_{i,n}^{(j)} - \bar{\mu}_{i,n}\right)^2}$ denote the standard deviation of ESG scores for the i-th asset at time n. We then construct an alternative index

$$\bar{u}_{i,n} := \bar{\mu}_{i,n} - b\,\bar{\sigma}_{i,n}, \quad i = 1, \ldots, I, \tag{20}$$

where the parameter b controls the penalization assigned to the standard deviation relative to the mean. This index penalizes assets with high disagreement across rating agencies, reflecting a more conservative assessment in which assets with widely varying ESG scores are considered less favorable due to the ambiguity or risk arising from the lack of consensus. The aim is to incorporate a measure of confidence or reliability into the ESG evaluation.[12] This construction can be interpreted as a lower confidence bound on the "true" ESG score: the higher the disagreement among rating agencies, the more the score is reduced. If ESG scores are assumed to follow a Gaussian distribution, b can be set as a quantile of the normal distribution, so that $\bar{u}_{i,n}$ corresponds to the lowest quantile of the ESG score distribution. For example $b = 1.96$ yields the 5% lower quantile, producing a conservative estimate that emphasizes robustness to disagreement. This approach follows the spirit of the *k*-sum method in Cesarone et al. (2024), although they apply the ESG enhancement to Mean-Variance optimal portfolios. When *k* is set equal to the number of rating agencies providing ESG scores, the *k*-sum method of Cesarone et al. (2024) becomes similar to our earlier approach of using the simple average of the ESG scores across all providers. For lower values of *k*, the *k*-sum method focuses on the lowest ESG scores within the optimal portfolio, effectively emphasizing the worst contributions. This is analogous to our alternative index $\bar{u}_{i,n} = \bar{\mu}_{i,n} - b\,\bar{\sigma}_{i,n}$, which penalizes assets with high dispersion among rating agencies, thereby reducing the score of assets with potentially inconsistent ESG evaluations. A key difference is that the *k*-sum method evaluates ESG scores during the portfolio optimization phase, so that the contribution of each asset to the portfolio's ESG score is weighted by its allocation. As a result, the impact of diverging ESG ratings is scaled by the portfolio weights, and assets with small weights have limited influence, even if their ESG scores vary substantially. In contrast, our approach computes ESG divergence *ex-ante* at the asset level, assigning a penalized ESG score to each company independently of its eventual portfolio allocation.

To provide a qualitative understanding of the interaction between $\bar{u}_n, \bar{\mu}_n$ and $\bar{\sigma}_n$ Fig. 7 presents scatter plots of $\bar{\mu}_{i,n}$ versus $\bar{\sigma}_{i,n}$ for $i = 1, \ldots, I$ on four representative dates. Each point corresponds to an asset, with its color indicating the value of $\bar{u}_{i,n}$ obtained with $b = 1.96$: brighter colors correspond to higher values, while darker colors represent lower values. The figure highlights an intriguing pattern. In 2014, there is a clear proportional relationship between $\bar{\mu}_n$ and $\bar{\sigma}_n$, indicating that assets with lower average ESG scores tended to have lower dispersion, and thus greater consensus among rating agencies. Over time, this relationship evolves, passing through periods of apparent independence, and by 2024 it shifts to an inverse proportional relationship. This latter pattern reflects a dynamic shift over the years, evolving from a situation where consensus was primarily among assets with lower ESG scores to one where consensus is mainly observed for assets with higher ESG scores.

The summary statistics for the out-of-sample optimal strategies, where the ESG score is replaced by the measure $\boldsymbol{\mu}_n$ and $\alpha = 0.2$, are presented in Table 8. The case $\lambda = 0$ is omitted, as its values coincide with those reported in Table 6. The corresponding results for $\alpha = 0.05$ can be found in Table 14 in Appendix B. A comparison with Table 6, which presents results based on the ESG average index $\boldsymbol{\mu}_n$, reveals that the alternative index $\boldsymbol{\mu}_n$ leads to more concentrated portfolios and lower average turnover. Specifically, it exhibits higher standard deviations for concentration and lower standard deviations for turnover. As before, the DRO approach significantly helps

[12] Other measures that combine central tendency and dispersion can also be considered. For example, the signal-to-noise ratio $\frac{\bar{\mu}_{i,n}}{\bar{\sigma}_{i,n}}$ for the *i*-th asset is one such measure. However, this ratio can be misleading as it may yield high values when both the mean and standard deviation are small.

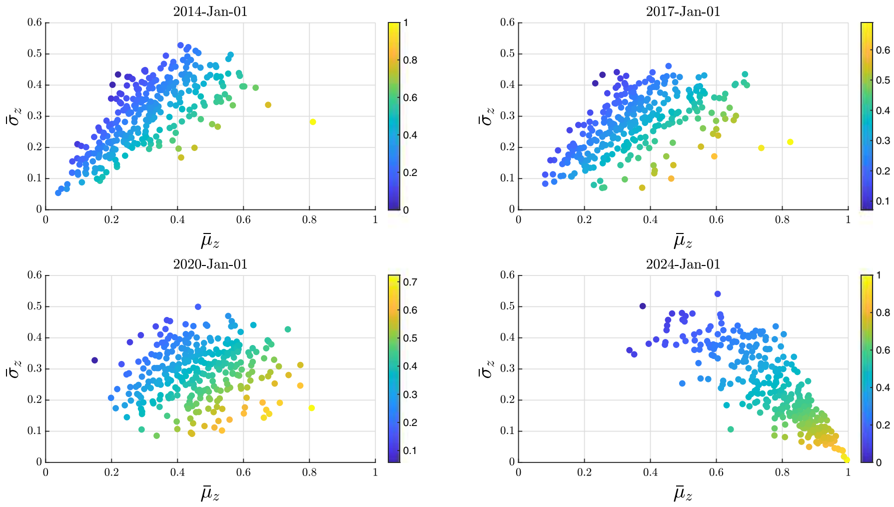

Fig. 7 Scatter plots of $\bar{\mu}_{i,n}$ versus $\bar{\sigma}_{i,n}$ for $i = 1, \ldots, I$ on four representative dates. Each point corresponds to an asset, with its color indicating the value of $\bar{u}_{i,n}$: brighter colors correspond to higher values, while darker colors represent lower values

reduce portfolio concentration as the uncertainty radius ϵ increases, holding other parameters fixed. For example, when the ESG appetite $\lambda = 0.4$, the average portfolio concentration is 46.64% without DRO constraints, which decreases to 25.84% for $\epsilon = 0.1$. The tail risk measures, CVaR and MD, increase slightly compared to the case in which the ESG average index $\boldsymbol{\mu}_n$ was used. This behavior can be attributed to the more concentrated portfolios generated when using the dispersion-penalized index $\boldsymbol{u}_n$ for high values of λ. By penalizing assets with high disagreement across ESG providers, $\boldsymbol{u}_n$ may overweight a smaller subset of assets, increasing exposure to extreme outcomes and, consequently, tail risk. The DRO approach helps mitigate this concentration, improving portfolio robustness. When the risk-aversion parameter α is reduced to 0.05, even for high λ values, the ESG preference remains relatively low compared to financial risk. As a result, extreme concentration and tail risk measures such as CVaR and MD are less pronounced. Remarkably, the DRO approach does not adversely affect financial performance in terms of annualized returns and reward-risk ratios when compared to the case without distributional uncertainty. The BB test performed between each optimal strategies and the EW portfolio in general confirms that the CVaR of the optimal strategies is still statistically lower then that of the EW portfolio for $\lambda \leq 0.3$, see Table 5 in the online appendix. Similarly, when the BB test is applied to pairs of optimal strategies sharing identical parameters except for the uncertainty radius ϵ (see Table 7 in the online appendix), the only statistically significant difference is observed for CVaR, which tends to improve as ϵ increases.

Table 9 reports the relative improvement in average ESG scores of the optimal portfolio compared with the case where ESG is not included in the objective function, as defined in (16). As expected, the average ESG score improve with respect to all the five rating agencies considered in the study. Compared with the values in Table 7, we observe that the index $\boldsymbol{u}_n$ tends to penalize assets with higher ESG scores from

Table 8 Summary statistics and risk/reward ratios for out-of-sample daily optimal portfolios from 1-June-2017 to 28-June-2024 using the ESG index u_n among the ESG scores of each company and for different values of ϵ

	AnnRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	Std-HHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
	λ = 0.1						λ = 0.1				
0.00	7.42	4.34	4.33	7.28	15.31	3.95	5.26	1.62	86.48	1.94	29.25
0.25	9.75	4.70	4.31	7.11	11.57	2.69	6.75	2.07	85.27	1.91	28.83
0.50	8.85	4.24	4.14	6.73	9.86	2.15	6.27	1.94	84.62	1.87	28.37
1.00	9.18	3.88	3.82	6.26	7.85	1.39	6.43	2.00	81.43	1.87	30.66
	λ = 0.2						λ = 0.2				
0.00	8.84	3.09	3.00	5.63	25.50	11.52	5.47	1.68	82.18	2.23	31.12
0.25	9.82	3.10	2.89	5.37	19.95	7.72	6.19	1.87	81.90	2.17	32.65
0.50	9.96	2.91	2.81	5.34	16.59	5.64	6.29	1.91	81.84	2.16	33.50
1.00	9.22	2.89	2.89	5.39	12.53	2.87	5.84	1.80	81.10	2.14	35.86
	λ = 0.3						λ = 0.3				
0.00	8.59	2.38	2.50	5.22	39.12	24.41	4.91	1.50	84.62	2.53	35.70
0.25	9.15	2.59	2.66	5.25	31.84	21.78	5.32	1.61	84.05	2.44	34.47
0.50	8.77	2.71	2.80	5.23	24.52	14.92	5.22	1.59	84.30	2.38	33.77
1.00	10.63	2.67	2.52	4.89	18.03	6.92	6.25	1.89	84.38	2.34	34.27
	λ = 0.4						λ = 0.4				
0.00	8.71	2.06	2.23	5.10	46.64	24.95	4.70	1.43	87.68	2.75	37.79
0.25	8.29	2.18	2.41	5.01	40.48	25.33	4.58	1.39	86.38	2.70	36.43
0.50	9.48	2.21	2.32	4.92	35.68	24.80	5.21	1.59	88.19	2.61	35.29
1.00	7.93	2.11	2.33	4.69	25.84	18.29	4.60	1.40	85.11	2.54	35.86

All the optimization in the table have the significance level of the CVaR $\beta = 0.90$, $\alpha = 0.2$ and $N = 252$

Table 9 Out-of-sample relative variation of the average ESG portfolio scores obtained with the ESG index u_n with respect to the case $\lambda = 0$, $\alpha = 0.2$, for different values of ϵ expressed in thousands (rows)

	ESG Relative Variation (%)						ESG Relative Variation (%)				
	BL	MN	RV	SP	TV		BL	MN	RV	SP	TV
ε	λ = 0.1						λ = 0.3				
0.00	39.5	5.0	5.0	12.5	13.5	0.00	56.9	8.2	9.5	22.9	16.6
0.25	38.6	4.7	5.1	12.6	13.2	0.25	56.1	8.0	9.5	22.7	16.7
0.50	37.2	4.5	4.7	12.3	12.8	0.50	54.9	7.8	9.3	22.1	16.8
1.00	35.7	4.1	4.7	12.1	11.6	1.00	53.1	7.5	8.9	21.1	16.8
	λ = 0.2						λ = 0.4				
0.00	52.2	7.2	7.8	19.2	16.0	0.00	58.5	8.6	9.1	23.3	17.4
0.25	50.8	6.9	7.6	18.6	16.1	0.25	57.7	8.4	9.3	23.2	17.1
0.50	49.4	6.6	7.4	18.4	16.2	0.50	57.0	8.3	9.7	23.2	17.0
1.00	46.4	6.1	6.7	17.6	16.0	1.00	55.7	8.1	9.8	22.7	17.0

Refinitiv (RF) and Standard & Poor (SP), while favoring assets with higher ESG scores from Morningstar (MN) and Truvalue (TV).

We conclude this section with some remarks on the robustness of the results under variations in the estimation time window N and the CVaR parameter β, considering the cases $N \in \{504, 756\}$ and $\beta \in \{0.95, 0.98, 0.99\}$, respectively. In general, increasing N leads to lower financial performance, both in terms of annualized returns and RRRs. It also reduces ESG values and relative variation, as shown in Tables 6–7. A possible explanation is that a larger estimation window increases uncertainty over the distributional estimates, resulting in more conservative portfolio allocations. We observe in Table 5 that, when all other parameters are fixed, increasing ϵ leads to lower financial performance. However, this is not always the case. In experiments with larger time window estimates (N), the DRO approach sometimes outperforms the analogous case with the same parameters but $\epsilon = 0$. Regarding the parameter β, we do not observe substantial differences from the findings discussed above. In summary, we can state that the real impacts of the DRO approach, as discussed in Sect. 3.4, are twofold. First, in situations of ESG discrepancy, DRO improves the ESG score across all providers compared to the case with no ESG appetite. Second, it leads to a substantial reduction in portfolio concentration. Since this reduction comes with similar ESG and financial performance, the DRO model can be of significant value to fund managers of ESG funds.

5 Conclusions

In this paper, we have applied an historical distributionally robust optimization approach to the Mean-CVaR-ESG optimal portfolio, enhanced with ESG considerations. We have followed a recent stream of academic literature by linearly incorporating the ESG score of the portfolio into the objective function. We have shown that ambiguity in ESG values leads to significantly different optimal solutions depending on the choice of the rating agency, even when the distributionally robust technique is applied to both sources of uncertainty, i.e., financial returns and ESG values. The optimal portfolio composition changes significantly depending on the sources of ESG scores used in the optimization phase. Even when two optimal portfolios obtained with data from two different ESG rating agencies exhibit similar results in terms of ESG scores, their compositions are profoundly different. Furthermore, the average ESG score of an optimal portfolio, when evaluated according to a specific provider, does not necessarily improve if the optimization is performed using scores from a different rating agency.

We have shown that optimizing a portfolio using the historical Mean-CVaR-ESG approach, based on two simple statistics of ESG scores from multiple providers, yields optimal portfolios that enhance ESG performance across all providers. However, this method tends to produce highly concentrated portfolios, especially when ESG preferences and risk tolerance are high. The DRO (Distributionally Robust Optimization) approach effectively reduces portfolio concentration, with only a minor impact on financial performance metrics. We argue that these features of the DRO model are particularly relevant to managers of ESG-focused funds, as it ensures more diversified portfolios compared to traditional historical Mean-CVaR-ESG optimization.

Appendix A: Derivation of the linear programming reformulation

Using the CVaR representation from Rockafellar and Uryasev (2002), the objective function

$$\mathbb{E}^{\mathbb{P}}\left[\ell\left(\boldsymbol{w},\boldsymbol{\xi}\right)\right] = -\alpha\left[\lambda\,\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{z}^{(j)}\right] + (1-\lambda)\,\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{r}\right]\right] + (1-\alpha)\,\mathrm{CVaR}_{\beta}^{\mathbb{P}}\left(-\boldsymbol{w}^{\top}\boldsymbol{r}\right), \tag{A1}$$

can be equivalently written as

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\left[\ell\left(\boldsymbol{w},\boldsymbol{\xi}\right)\right] &= -\alpha\lambda\,\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{z}^{(j)}\right] - \alpha(1-\lambda)\,\mathbb{E}^{\mathbb{P}}\left[\boldsymbol{w}^{\top}\boldsymbol{r}\right] + \\ &\quad + (1-\alpha)\inf_{\tau\in\mathbb{R}}\left\{\tau + \frac{1}{1-\beta}\,\mathbb{E}^{\mathbb{P}}\left[\left(-\boldsymbol{w}^{\top}\boldsymbol{r}-\tau\right)_{+}\right]\right\}.\end{aligned} \tag{A2}$$

The last equation can be further modified by taking out the expected value and the infimum problem as:

$$\inf_{\tau\in\mathbb{R}}\left\{\mathbb{E}^{\mathbb{P}}\left[-\alpha\lambda\,\boldsymbol{w}^{\top}\boldsymbol{z}^{(j)} - \alpha(1-\lambda)\,\boldsymbol{w}^{\top}\boldsymbol{r} + (1-\alpha)\tau + \frac{1}{1-\beta}\left(-\boldsymbol{w}^{\top}\boldsymbol{r}-\tau\right)_{+}\right]\right\}. \tag{A3}$$

Recall that the random vector $\boldsymbol{\xi}^{(j)}$, representing the uncertainty, can be written as $\boldsymbol{\xi}^{(j)} = \left[\boldsymbol{r}, \boldsymbol{z}^{(j)}\right]^{\top}$ Thus, the loss function formulation in Esfahani and Kuhn (2018), expressed as

$$\ell\left(\boldsymbol{w},\boldsymbol{\xi}\right) := \max_{k\in K}\left[a_{1,k}\langle\boldsymbol{w},\boldsymbol{r}_n\rangle + a_{2,k}\langle\boldsymbol{w},\boldsymbol{z}_n\rangle + b_k\right], \tag{A4}$$

can be recovered by simply setting $K = 2$ and the coefficients as follows:

$$\begin{aligned} a_{1,1} &= -\alpha(1-\lambda) - \frac{1-\alpha}{1-\beta}, & a_{2,1} &= -\alpha\lambda, & b_1 &= (1-\alpha) - \frac{1}{1-\beta},\\ a_{1,2} &= -\alpha(1-\lambda), & a_{2,2} &= -\alpha\lambda, & b_2 &= 1-\alpha. \end{aligned}$$

It is now straightforward to show that, by assuming the daily returns lie in the interval $[-1, 1]$ and the ESG scores in [0, 1], the polytope describing the support set

$$\Xi^{(j)} = \left\{\boldsymbol{\xi}^{(j)} \in \mathbb{R}^{2I} : C\boldsymbol{\xi}^{(j)} \le \boldsymbol{d}\right\}$$

of the uncertainty vector can be written as

$$C = \left[\boldsymbol{I}_I \quad -\boldsymbol{I}_I \quad \boldsymbol{I}_I \quad -\boldsymbol{I}_I\right]^{\top}, \qquad \boldsymbol{d} = \begin{bmatrix}\mathbf{1}_{3I}\\ \mathbf{0}_I\end{bmatrix}.$$

Appendix B: Additional figures and tables

Table 10 and Figs. 8, 9 and 10.

Table 10 Summary statistics for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024

	AnnRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	Std-HHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
λ = 0.0											
NO ESG	8.95	5.97	5.62	9.79	9.02	2.81	6.76	2.06	82.34	1.76	23.95
λ = 0.1											
BL	9.63	5.97	5.55	9.69	9.47	2.51	7.18	2.17	81.59	1.78	24.48
MN	7.86	5.64	5.40	9.46	9.18	2.78	6.07	1.83	82.89	1.76	23.31
RF	7.38	5.67	5.60	9.51	9.10	2.73	5.67	1.73	82.57	1.76	25.02
SP	8.15	5.42	5.53	9.47	9.54	2.92	6.21	1.89	85.71	1.76	25.98
TV	9.93	5.89	5.36	9.29	9.60	2.85	7.42	2.23	82.64	1.78	22.89
λ = 0.2											
BL	8.86	5.01	4.73	8.55	11.82	3.61	6.57	2.01	84.70	1.80	26.81
MN	7.75	4.96	4.95	8.95	9.92	3.17	5.99	1.83	84.45	1.74	23.50
RF	7.17	5.27	5.39	9.38	9.32	2.38	5.50	1.68	84.43	1.77	24.96
SP	8.13	4.74	4.89	8.52	10.13	3.53	6.12	1.87	86.18	1.79	26.70
TV	9.84	5.81	5.06	8.83	10.03	2.68	7.33	2.23	85.37	1.77	22.95
λ = 0.3											
BL	6.09	3.93	4.12	7.78	15.49	5.50	4.55	1.38	79.81	1.89	28.93
MN	6.64	4.38	4.71	8.39	11.14	3.77	5.18	1.61	86.31	1.73	23.30
RF	6.20	4.50	5.00	8.93	9.84	2.54	4.76	1.46	83.21	1.81	27.13
SP	6.36	4.27	4.73	8.24	10.79	3.50	4.87	1.48	83.82	1.83	26.59
TV	11.67	6.02	4.88	8.27	10.88	3.12	8.47	2.57	86.49	1.79	24.03
λ = 0.4											
BL	5.39	3.22	3.51	6.88	20.20	8.23	3.99	1.20	78.05	1.98	30.94
MN	5.98	4.10	4.47	7.89	12.09	3.87	4.66	1.46	87.16	1.75	24.22
RF	4.71	4.01	4.74	8.32	10.60	2.79	3.74	1.15	82.87	1.82	26.83
SP	5.93	4.02	4.48	8.01	11.19	3.20	4.56	1.39	83.81	1.83	26.20
TV	12.44	5.81	4.51	8.15	12.03	3.76	8.76	2.63	87.09	1.86	25.51

Results are presented for each choice of rating agency used for the ESG value in the optimization and for four values of the parameter λ. All optimizations in the table follow the same parameter configuration: $\alpha = 0.05$, $\beta = 0.90$, $\epsilon = 0$, and $N = 252$. When $\lambda = 0$, the optimizer does not consider the ESG value, resulting in a single identical row across all rating agencies

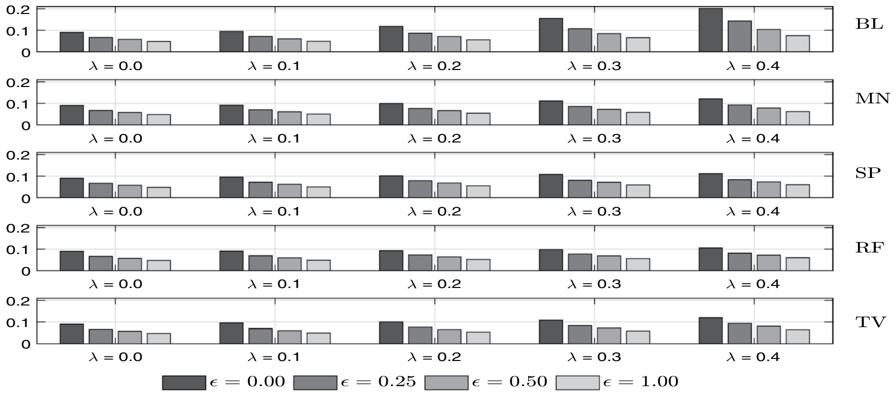

Fig. 8 The Herfindahl-Hirschman Index (HHI) for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the HHI of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.05$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

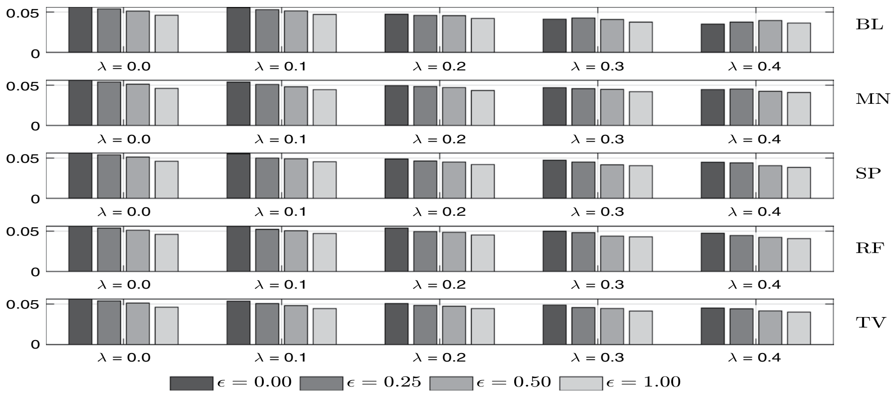

Fig. 9 The Average Turnover for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the Average Turnover ratio of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.05$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

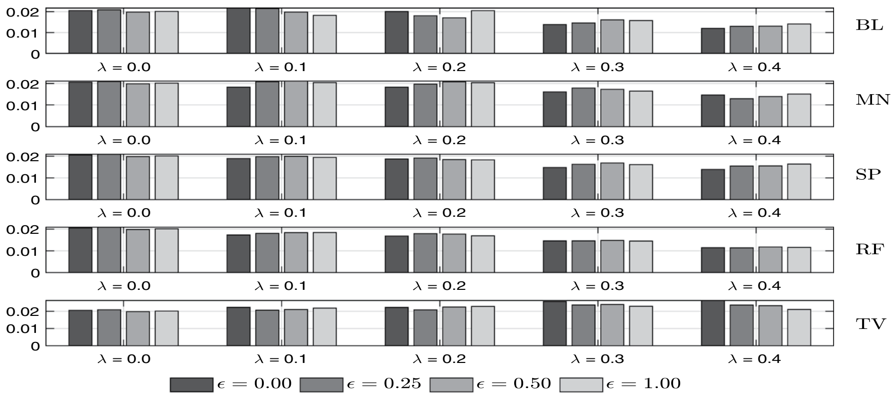

Fig. 10 The STAR ratio for out-of-sample daily optimal portfolios from June 1, 2017, to June 28, 2024, is shown for the EUX universe and for each choice of a rating agency used to determine the ESG value in the optimization. Each vignette displays the STAR ratio of the strategy for four different values of the parameter ϵ, expressed in thousands, with $\alpha = 0.05$. Each row corresponds to a particular ESG rating agency used to compute the optimal strategy

Tables 11, 12, 13 and 14.

Table 11 The left panel shows the average ESG relative variation $\delta_{i,j}(\lambda)$, as defined in eq. (16), of the optimal portfolios with respect to the case $\lambda = 0$

	ESG relative variation (%)					Average turnover (%)				
	BL	MN	RF	SP	TV	BL	MN	RF	SP	TV
	λ = 0.1					λ = 0.1				
BL	17.3	0.2	3.0	5.9	1.7	0.0	42.3	40.3	44.3	40.3
MN	0.8	2.9	0.4	2.6	0.0	42.3	0.0	37.8	42.4	38.7
RF	3.7	− 0.3	7.7	8.3	− 0.0	40.3	37.8	0.0	33.5	38.3
SP	5.9	0.5	7.0	19.3	0.0	44.3	42.4	33.5	0.0	46.5
TV	2.8	0.2	0.1	− 0.2	6.4	40.3	38.7	38.3	46.5	0.0
	λ = 0.2					λ = 0.2				
BL	34.1	0.1	7.0	11.3	2.5	0.0	64.7	62.3	64.9	62.8
MN	1.5	4.8	0.7	4.4	− 0.1	64.7	0.0	57.1	60.4	55.6
RF	7.1	− 0.5	13.4	14.8	− 0.6	62.3	57.1	0.0	46.2	59.9
SP	8.7	0.2	10.0	27.3	− 0.7	64.9	60.4	46.2	0.0	65.0
TV	4.0	0.2	− 0.3	− 0.1	10.9	62.8	55.6	59.9	65.0	0.0
	λ = 0.3					λ = 0.3				
BL	46.3	0.3	9.7	14.4	2.7	0.0	78.5	72.8	75.5	76.1
MN	2.5	6.5	1.5	6.3	− 0.6	78.5	0.0	68.1	70.3	69.0
RF	9.8	− 0.9	17.1	18.7	− 1.0	72.8	68.1	0.0	53.1	72.9
SP	9.9	− 0.2	11.4	30.5	− 0.8	75.5	70.3	53.1	0.0	74.4
TV	4.4	0.1	− 1.0	− 0.4	14.7	76.1	69.0	72.9	74.4	0.0
	λ = 0.4					λ = 0.4				
BL	55.2	0.4	11.1	17.2	2.5	0.0	87.2	80.4	81.9	85.9
MN	2.9	7.9	2.2	7.1	− 1.1	87.2	0.0	75.7	76.7	77.9
RF	11.5	− 1.6	19.6	21.2	− 1.3	80.4	75.7	0.0	57.2	81.2
SP	10.4	− 0.3	12.3	32.0	− 0.9	81.9	76.7	57.2	0.0	80.3
TV	4.6	− 0.1	− 1.8	− 0.9	17.7	85.9	77.9	81.2	80.3	0.0

The right panel shows the average turnover between Mean-CVaR-ESG optimal portfolio strategies when switching the optimization from one rating agency to another, for different values of λ. In all the cases the other parameters are $N = 252$, $\beta = 0.90$, $\epsilon = 0$, $\alpha = 0.05$. The rating agencies considered are Bloomberg (BL), Refinitiv Eikon (RF), Morningstar Sustainalytics (MN), S&P Global (SP), and Truvalue Labs (TV)

Table 12 Summary statistics for out-of-sample daily optimal portfolios from 1-June-2017 to 28-June-2024 using the ESG average index $\bar{\boldsymbol{\mu}}_n$ among the ESG scores of each company and for different values of ϵ expressed in thousands

	AnnRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	Std-HHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
ε	λ = 0.0										
0.00	8.95	5.97	5.62	9.79	9.02	2.81	6.76	2.06	82.34	1.76	23.95
0.25	9.06	5.83	5.39	9.12	6.67	2.36	6.90	2.09	82.11	1.75	23.93
0.50	8.57	5.42	5.13	8.67	5.76	1.78	6.57	1.98	81.56	1.75	24.52
1.00	8.61	4.83	4.60	7.65	4.78	1.40	6.64	2.01	82.61	1.73	24.39
λ = 0.1											
0.00	9.25	5.76	5.38	9.41	9.06	2.71	7.05	2.14	83.14	1.74	24.60
0.25	8.67	5.53	5.22	8.79	6.89	2.34	6.61	2.02	83.21	1.74	25.01
0.50	8.71	5.39	5.14	8.35	5.93	1.86	6.67	2.03	81.61	1.74	24.77
1.00	8.12	4.97	4.85	7.81	4.88	1.34	6.29	1.92	81.70	1.72	25.64
λ = 0.2											
0.00	8.93	5.60	5.39	9.55	9.48	2.31	6.77	2.07	85.06	1.75	24.88
0.25	9.63	5.27	4.86	8.43	7.21	2.15	7.36	2.25	86.27	1.71	24.55
0.50	8.62	4.90	4.78	8.18	6.29	1.89	6.60	2.03	83.76	1.72	25.85
1.00	8.46	4.74	4.66	7.87	5.03	1.34	6.48	1.99	82.78	1.73	26.08
λ = 0.3											
0.00	8.89	5.39	5.24	9.18	9.97	2.31	6.58	2.03	86.50	1.78	27.19
0.25	8.84	4.90	4.83	8.42	7.61	2.03	6.68	2.05	85.54	1.75	26.36
0.50	8.64	4.87	4.76	8.12	6.55	1.85	6.56	2.02	85.11	1.73	25.86
1.00	7.60	4.51	4.52	7.70	5.29	1.35	5.85	1.80	84.06	1.74	25.60
λ = 0.4											
0.00	9.11	5.20	4.96	8.66	10.65	2.46	6.60	2.05	86.98	1.81	27.10
0.25	7.85	4.68	4.80	8.22	8.12	2.03	5.80	1.80	85.30	1.81	27.16
0.50	6.63	4.39	4.64	7.94	6.97	1.86	5.03	1.56	83.88	1.79	27.74
1.00	6.64	4.14	4.33	7.40	5.63	1.43	5.12	1.58	83.25	1.77	26.95

All the optimization in the table have the significance level of the CVaR $\beta = 0.90$, $\alpha = 0.05$ and $N = 252$

Table 13 Out-of-sample relative variation of the average ESG portfolio scores obtained with the ESG index $\bar{\boldsymbol{\mu}}_n$ with respect to the case $\lambda = 0$, $\alpha = 0.05$, for different values of ϵ expressed in thousands (rows)

	ESG Relative Variation (%)						ESG Relative Variation (%)				
	BL	MN	RV	SP	TV		BL	MN	RV	SP	TV
ε	λ = 0.1						λ = 0.3				
0.00	12.5	1.7	7.2	14.7	3.0	0.00	30.5	3.6	13.9	26.2	6.6
0.25	11.2	1.3	7.2	14.2	2.9	0.25	27.7	2.9	13.9	25.3	6.0
0.50	11.0	1.1	7.3	13.8	2.8	0.50	26.2	2.6	13.9	25.0	5.8
1.00	10.8	0.9	6.9	13.2	2.3	1.00	24.3	2.6	13.3	24.3	5.5
	λ = 0.2						λ = 0.4				
0.00	22.4	2.6	11.1	21.8	5.2	0.00	36.8	3.6	16.0	29.2	7.2
0.25	20.7	2.1	11.4	21.3	4.8	0.25	33.3	3.6	15.4	28.6	7.1
0.50	19.9	2.0	11.3	21.1	5.0	0.50	31.2	3.3	15.3	28.1	6.8
1.00	18.6	1.9	11.0	20.3	4.1	1.00	28.9	3.2	14.9	27.2	6.5

Table 14 Summary statistics for out-of-sample daily optimal portfolios from 1-June-2017 to 28-June-2024 using the ESG index $\bar{u}_n$ among the ESG scores of each company and for different values of ϵ expressed in thousands

	An-nRet (%)	Cost (%)	AvgTO (%)	StdTO (%)	HHI (%)	Std-HHI (%)	SS (%)	STAR (%)	Rachev (%)	CVaR (%)	MDD (%)
EW	8.01	0.49	0.26	0.11	0.33	0.00	4.57	1.38	89.56	2.61	39.19
λ = 0.1											
0.00	9.04	5.79	5.36	9.22	9.67	2.59	6.78	2.07	83.71	1.76	24.15
0.25	9.46	5.54	5.07	8.59	7.23	2.19	7.20	2.18	83.28	1.74	24.23
0.50	8.67	5.22	4.96	8.24	6.10	1.55	6.64	2.01	82.37	1.75	24.94
1.00	8.19	4.85	4.71	7.76	4.95	1.19	6.33	1.93	81.69	1.73	24.85
λ = 0.2											
0.00	9.27	5.57	5.03	8.77	11.04	2.51	6.81	2.10	86.83	1.79	24.88
0.25	10.02	5.23	4.64	8.03	8.15	1.79	7.48	2.32	86.60	1.73	25.23
0.50	10.24	5.08	4.46	7.64	6.92	1.48	7.63	2.35	86.19	1.74	25.56
1.00	9.08	4.61	4.30	7.25	5.60	1.18	6.84	2.11	83.47	1.74	26.65
λ = 0.3											
0.00	9.73	4.99	4.55	7.99	12.39	2.76	6.95	2.14	86.84	1.84	27.01
0.25	9.36	4.49	4.20	7.40	9.36	1.82	6.80	2.11	85.64	1.80	27.19
0.50	9.66	4.36	4.12	7.22	8.03	1.54	7.06	2.19	86.18	1.78	27.45
1.00	8.54	4.21	4.06	6.97	6.50	1.12	6.39	1.98	85.08	1.76	27.14
λ = 0.4											
0.00	8.37	4.42	4.31	7.55	14.13	3.34	5.94	1.83	86.79	1.90	29.38
0.25	9.69	4.21	3.97	7.28	10.60	2.39	6.88	2.14	86.71	1.84	28.14
0.50	9.31	3.92	3.74	6.62	9.07	1.87	6.70	2.09	85.58	1.81	28.11
0.50	8.91	3.67	3.61	6.42	7.34	1.22	6.40	2.01	83.85	1.81	28.70

All the optimization in the table have the significance level of the CVaR $\beta = 0.90$, $\alpha = 0.05$ and $N = 252$

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10287-025-00548-z.

Acknowledgements This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 - CUP F83C22001720001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Funding Open access funding provided by Università degli Studi di Bergamo within the CRUI-CARE Agreement.

References

Agosto A, Giudici P, Tanda A (2023) How to combine esg scores? A proposal based on credit rating prediction. Corp Soc Responsib Environ Manag 30(6):3222–3230
Agosto A, Tanda A (2025) Divergence and aggregation of esg ratings: asurvey. Open Res Eur 5(28):28
Avramov D, Cheng S, Lioui A, Tarelli A (2022) Sustainable investing with esg rating uncertainty. J Financ Econ 145(2):642–664
Berg F, Koelbel JF, Rigobon R (2022) Aggregate confusion: the divergence of esg ratings. Rev Finance 26(6):1315–1344
Biglova A, Ortobelli S, Rachev S, Stoyanov S (2004) Different approaches to risk estimation in portfolio theory. J Portf Manag 31(1):103
Billio M, Costola M, Hristova I, Latino C, Pelizzon L (2021) Inside the esg ratings:(dis) agreement and performance. Corp Soc Responsib Environ Manag 28(5):1426–1445
Billio M, Fitzpatrick AC, Latino C, Pelizzon L (2024) Unpacking the esg ratings: Does one size fit all?
Bissoondoyal-Bheenick E, Bennett S, Lehnherr R, Zhong A (2024) Esg rating disagreement: implications and aggregation approaches. Int Rev Econ Finance 96:103532
Cesarone F, Martino ML, Carleo A (2022) Does esg impact really enhance portfolio profitability? Sustainability 14(4):2050
Cesarone F, Martino ML, Ricca F, Scozzari A (2024) Managing esg ratings disagreement in sustainable portfolio selection. Comput Oper Res 170:106766
Cheridito P, Kromer E (2013) Reward-risk ratios. J Invest Strateg 3(1):3–18
de Mello TH, Pagnoncelli BK (2016) Risk aversion in multistage stochastic programming: a modeling and algorithmic perspective. Eur J Oper Res 249(1):188–199
Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Math Program 171:115–166
Gai L, Bellucci M, Biggeri M, Ferrone L, Ielasi F (2023) Banks' esg disclosure: a new scoring model. Financ Res Lett 57:104199
Gucciardi G, Ossola E, Parisio L, Pelagatti M (2025) Common factors behind companies' environmental ratings. Int Rev Financ Anal 100:103961
Ledoit O, Wolf M (2008) Robust performance hypothesis testing with the sharpe ratio. J Empir Finance 15(5):850–859
Lotfi S, Pagliardi G, Paparoditis E, Zenios SA (2025) Hedging political risk in international portfolios. Eur J Oper Res 322(2):629–646
Markowitz H (1952) Modern portfolio theory. J Finance 7(11):77–91
Martin RD, Rachev ST, Siboulet F (2003) Phi-alpha optimal portfolios and extreme risk management. Wilmott 2003:70–83
Pedersen LH, Fitzgibbons S, Pomorski L (2021) Responsible investing: the ESG-efficient frontier. J Financ Econ 142(2):572–597
Pichler A, Shapiro A (2021) Mathematical foundations of distributionally robust multistage optimization. SIAM J Optim 31(4):3044–3067
Rachev ST (1991) Probability metrics and the stability of stochastic models. Wiley, Chichester
Rockafellar R, Uryasev S (2002) Conditional value-at-risk for general loss distributions. J Bank Finance 26(7):1443–1471
Shapiro A (2021) Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. Eur J Oper Res 288(1):1–13
Sortino FA, Satchell S (2001) Managing downside risk in financial markets. Butterworth-Heinemann, Oxford Boston
Utz S, Wimmer M, Steuer RE (2015) Tri-criterion modeling for constructing more-sustainable mutual funds. Eur J Oper Res 246(1):331–338
Varmaz A, Fieberg C, Poddig T (2024) Portfolio optimization for sustainable investments. Ann Oper Res 341(2):1151–1176

Authors and Affiliations

Davide Lauria[1,2] · **Marco Bonomelli**[1] · **Gabriele Torri**[1] · **Rosella Giacometti**[1]

✉ Davide Lauria
davide.lauria@unibg.it

Marco Bonomelli
marco.bonomelli@unibg.it

Gabriele Torri
gabriele.torri@unibg.it

Rosella Giacometti
rosella.giacometti@unibg.it

1 Department of Management, University of Bergamo, 24127 Bergamo, Italy

2 Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409-1042, USA