



Managing ESG Ratings Disagreement in Sustainable Portfolio Selection

Manuel Luis Martino, Roma Tre University

About the speaker



- **Manuel Luis Martino** – *PhD Candidate, Roma Tre University*
 - *Research focus on Portfolio Optimization, Financial Modelling, Sustainable Finance, and Risk Management*
 - *Teaching Assistant, MSc in Finance – ESADE Business School*
 - *Visiting Researcher, Banco Mediolanum (Barcelona, Spain)*
 - *Awarded 1st place in the “Orazio Ruggeri ASSIOM FOREX 2023” contest.*
-



- Roma Tre University – Department of Business Economics
- PhD Candidate in Quantitative Finance
- Research Group on Mathematical Finance



Motivations



“Environmental, social, and governance (ESG) rating providers have become influential institutions” [Berg et al. 2022].

- A total of 5,372 investors, representing more than USD 121.3 trillion in AUM have committed to integrating ESG information into their investment decisions [PRI, 2023].
- The main strand of the literature studies has fueled over years the debate on the relationship between ESG ratings and financial performance [Friede et al. 2015, Cesarone et al. 2022].
- However, ESG ratings from different data providers differ substantially [Billio et al. 2021, Gibson et al. 2021, Chatterji et al. 2016] and often produce disagreement on the evaluation of the same company.
- Indeed, disagreement over ESG ratings could generate conflicting information for investors, potentially underestimating or overestimating market opportunities for sustainable investments.

Aims of the work



1. to introduce a new general method to manage the ESG ratings disagreement in portfolio selection.
2. to include in the Mean-Variance framework a new objective that takes into account the ESG ratings provided by multiple data providers.
3. to formulate the resulting multi-objective nonlinear portfolio optimization problem as a convex quadratic programming, by exploiting the k -sum optimization strategy [Puerto et al. 2017, Ponce et al. 2018].
4. to provide out-of-sample evaluation of the proposed approach on some real-world data sets.

On the disagreement of ESG scores



- ESG rating agencies produce scores based on proprietary methodologies and metrics, which generally result in different information and interpretations [Gibson et al. 2021]:

Rating Provider	Abbreviation	Score Range	Metric
Refinitiv	RFT	0 (Brownest) - 100 (Greenest)	ESG Score
Bloomberg	BMG	0 (Brownest) - 100 (Greenest)	ESG Disclosure Score
Morningstar Sustainalytics	MNG	0 (Greenest) - 100 (Brownest)	ESG Risk Score
S&P Global	S&P	0 (Brownest) - 100 (Greenest)	ESG Global Rank

- This implies that the scores assigned to the same asset by two different agencies may be different.
- To ensure that ESG scores from different data providers receive the same weight and give the same information, we perform the *feature scaling* of ESG.

On the disagreement of ESG scores

Consider an investment universe of n assets and m different rating agencies. Let e_{ij} be ESG score that agency i , $i = 1, \dots, m$ assigns to the asset j , $j = 1, \dots, n$.

- We first map the ESG scores onto a new scale according to the min-max normalization:

$$\bar{e}_{ij} = \frac{e_{ij} - e_j^{\min}}{e_j^{\max} - e_j^{\min}}, \quad \text{where } e_i^{\min} = \min_{1 \leq j \leq n} e_{ij}, \quad e_i^{\max} = \max_{1 \leq j \leq n} e_{ij}, \quad \text{and with } \bar{e}_{ij} \in [0,1].$$

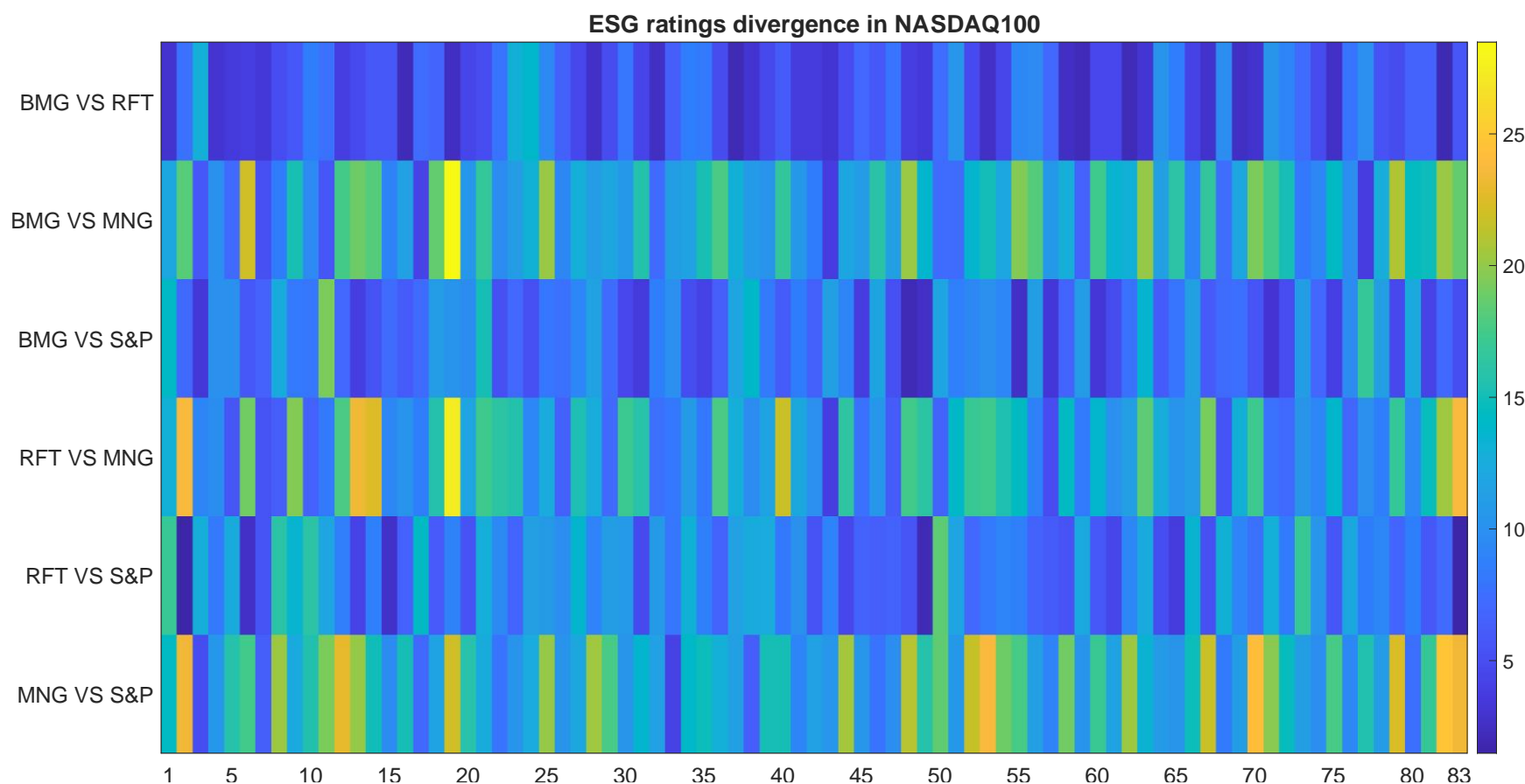
- We then define the **Non-ESG score** as the complement of the ESG score, namely $s_{ij} = 1 - \bar{e}_{ij}$

where $s_{ij} \in [0,1]$.

- Clearly, the lower the s_{ij} , the greener the asset j , according to the rating of agency i .

On the disagreement of ESG scores

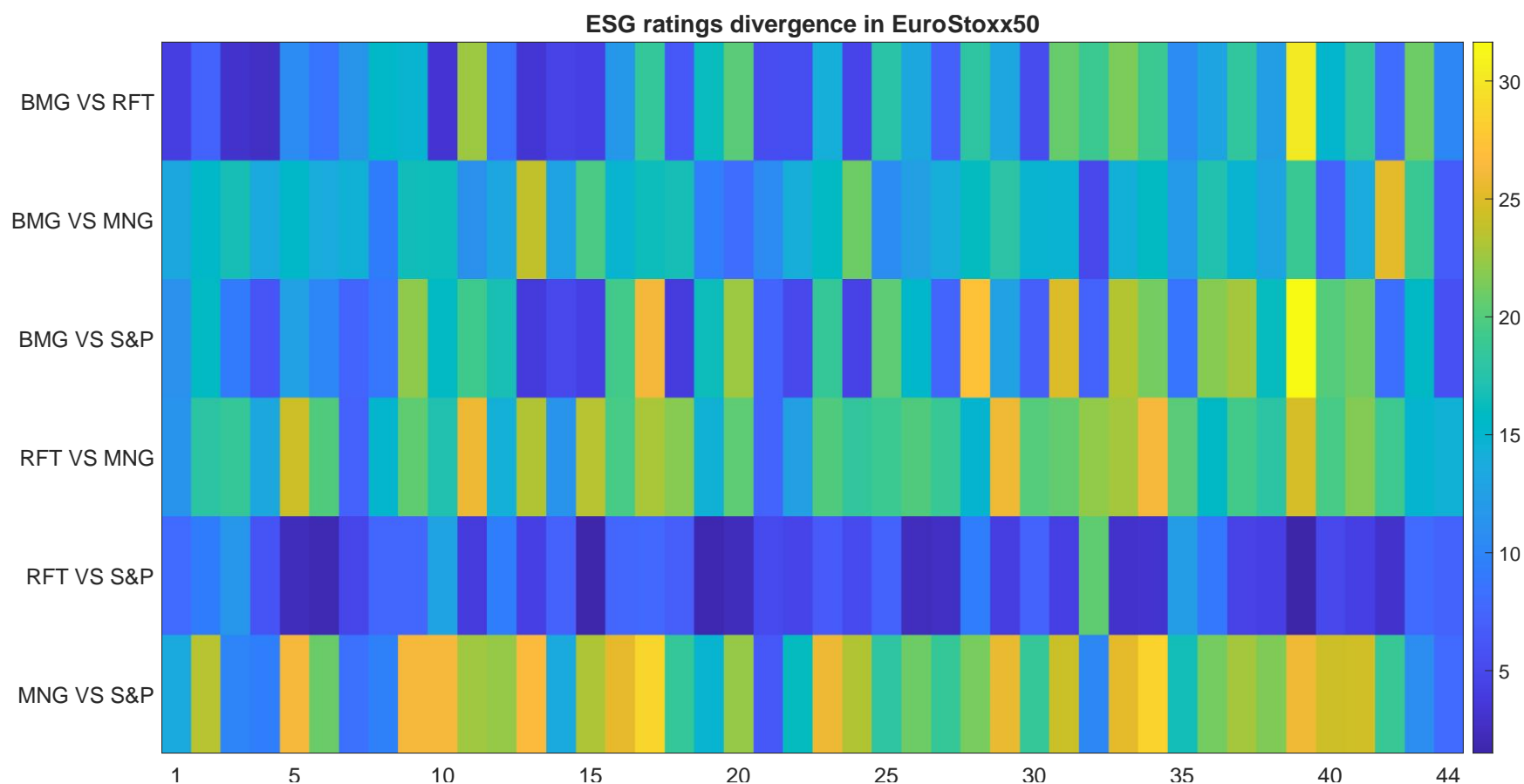
- For each asset we compute the pairwise Euclidean distances between ESG scores assigned by the four data providers, on NASDAQ100 ($n = 83$ asset constituents):



On the disagreement of ESG scores



- For each asset we compute the pairwise Euclidean distances between ESG scores assigned by the four data providers, on EUROSTOXX50 ($n = 44$ asset constituents):



On the disagreement of ESG scores

- To understand whether there are valuation differences in NASDAQ100 and EUROSTOXX50 stock markets, we also compute the average distances of the ESG scores over all assets and data providers, using several distance metrics:

Market Index	Euclidean	Chebychev	Cosine	Correlation
NASDAQ100	10.22	0.40	0.05	0.35
EuroStoxx 50	16.00	0.56	0.07	0.75

- The rating providers selected seem to agree more among themselves when assigning ESG ratings in the US market than in the European market.
- We expect that our k -sum optimization approach should be more effective in the case of markets characterized by more intense disagreement among rating providers.

Theoretical Framework (1)

- We denote by s^i the column vector of the Non-ESG scores assigned to the n assets by agency $i, i = 1, \dots, m$.
- The global Non-ESG score of a portfolio according to agency i is defined as:

$$s^i(x) = (s^i)^T x$$

Where x is the vector of portfolio weights with $\sum_{j=1}^n x_j = 1$ and $x_j \geq 0, j = 1, \dots, n$.

- For a given portfolio we sort in a non-increasing order the m global Non-ESG scores assigned by the m agencies:

$$s^{\bar{\pi}_1(x)}(x) \geq s^{\bar{\pi}_2(x)}(x) \geq \dots \geq s^{\bar{\pi}_m(x)}(x) \quad \text{where} \quad \bar{\pi}(x) = (\bar{\pi}_1(x), \bar{\pi}_2(x), \bar{\pi}_i(x), \dots, \bar{\pi}_m(x)).$$

Theoretical Framework (2)

- We define the q -Worst Non-ESG score of a portfolio as the sum of the q worst largest scores in the above ranking $1 \leq q \leq m$:

$$s_q^{\bar{\pi}(x)}(x) = \sum_{i=1}^q s^{\bar{\pi}_i(x)}(x) = \max_{\pi_1(x), \dots, \pi_m(x)} \sum_{i=1}^q s^{\pi_i(x)}(x)$$

- For a given q , our investment sustainability objective is to find a portfolio that minimizes $s_q^{\bar{\pi}(x)}(x)$:
 - If $q = m$, we minimize the sum of all scores (i.e., the permutation does not apply).
 - If $q = 1$, we minimize the worst portfolio score among those assigned by the m agencies.
 - If $1 < q < m$, we have an intermediate worst-case approach.

Theoretical Framework (3)

- We are interested in solving the following three-objective portfolio optimization problem

$$\begin{aligned}
 & \min_x \quad \sigma_P^2(x) = x^T \Sigma x \\
 & \max_x \quad \mu_P(x) = \mu^T x \\
 & \min_x \quad s_q^{\bar{\pi}(x)}(x) = \max_{\pi_1(x), \dots, \pi_m(x)} \sum_{i=1}^q s^{\pi_i(x)}(x) \\
 & \text{s.t.} \\
 & \quad x \in \Omega
 \end{aligned} \tag{1}$$

Where Σ and μ represent the covariance matrix of the asset returns and the vector of expected returns.

- The feasible domain is defined as:

$$\Omega = \left\{ x \in \mathbb{R}^n : \sum_{j=1}^n x_j = 1, x_j \geq 0, j = 1, \dots, n \right\}$$

Theoretical Framework (4)



- Since $\sigma_P^2, \mu_P, s_q^{\bar{\pi}(x)}(x)$ are convex functions, the three-objective optimization problem (1) can be solved by applying a weighted sum scalarization. Thus, we obtain:

$$\min_{x \in \Omega} \left\{ \lambda_1 (x^T \Sigma x) - \lambda_2 (\mu^T x) + \lambda_3 \left(\max_{\pi_1(x), \dots, \pi_m(x)} \sum_{i=1}^q s^{\pi_i(x)}(x) \right) \right\} \quad (2)$$

Where $\lambda_i \geq 0, i = 1, 2, 3$, are scalars.

Theoretical Framework (5)

- Following the approach in [Ponce et al. 2018, Puerto et al. 2017], the inner maximization problem in (1) can be re-written, for any given x , by introducing m new binary variables $z_i, i = 1, \dots, m$

$$\max_{\pi_1(x), \dots, \pi_m(x)} \sum_{i=1}^q s^{\pi_i(x)}(x) = \max_z \left\{ \sum_{i=1}^m s^i(x) z_i : \sum_{i=1}^m z_i = q, \quad z_i \in \{0,1\}, \quad i = 1, \dots, m \right\}$$

For a given x , the above problem finds the sum of the **q worst (largest) Non-ESG scores**.

- The constraint matrix is TU, hence the integrality constraints on the variables $z_i, i = 1, \dots, m$ can be relaxed to $0 \leq z_i \leq 1, i = 1, \dots, m$.
- Therefore, we obtain a linear equivalent program and exploit its exact dual formulation.

Theoretical Framework (6)



- Our original three-objective portfolio selection problem (1) can now be formulated as a Convex Quadratic Program

$$\begin{aligned} \min_{(x,v,u)} \quad & \lambda_1 x^T \Sigma x - \lambda_2 \mu^T x + \lambda_3 \left[qu + \sum_{i=1}^m v_i \right] \\ \text{s.t.} \quad & v_i + u \geq (s^i)^T x & i = 1, \dots, m \\ & v_i \geq 0 & i = 1, \dots, m \\ & u \geq 0 \\ & \sum_{k=1}^n x_k = 1 \\ & x_k \geq 0 & k = 1, \dots, n \end{aligned} \tag{3}$$

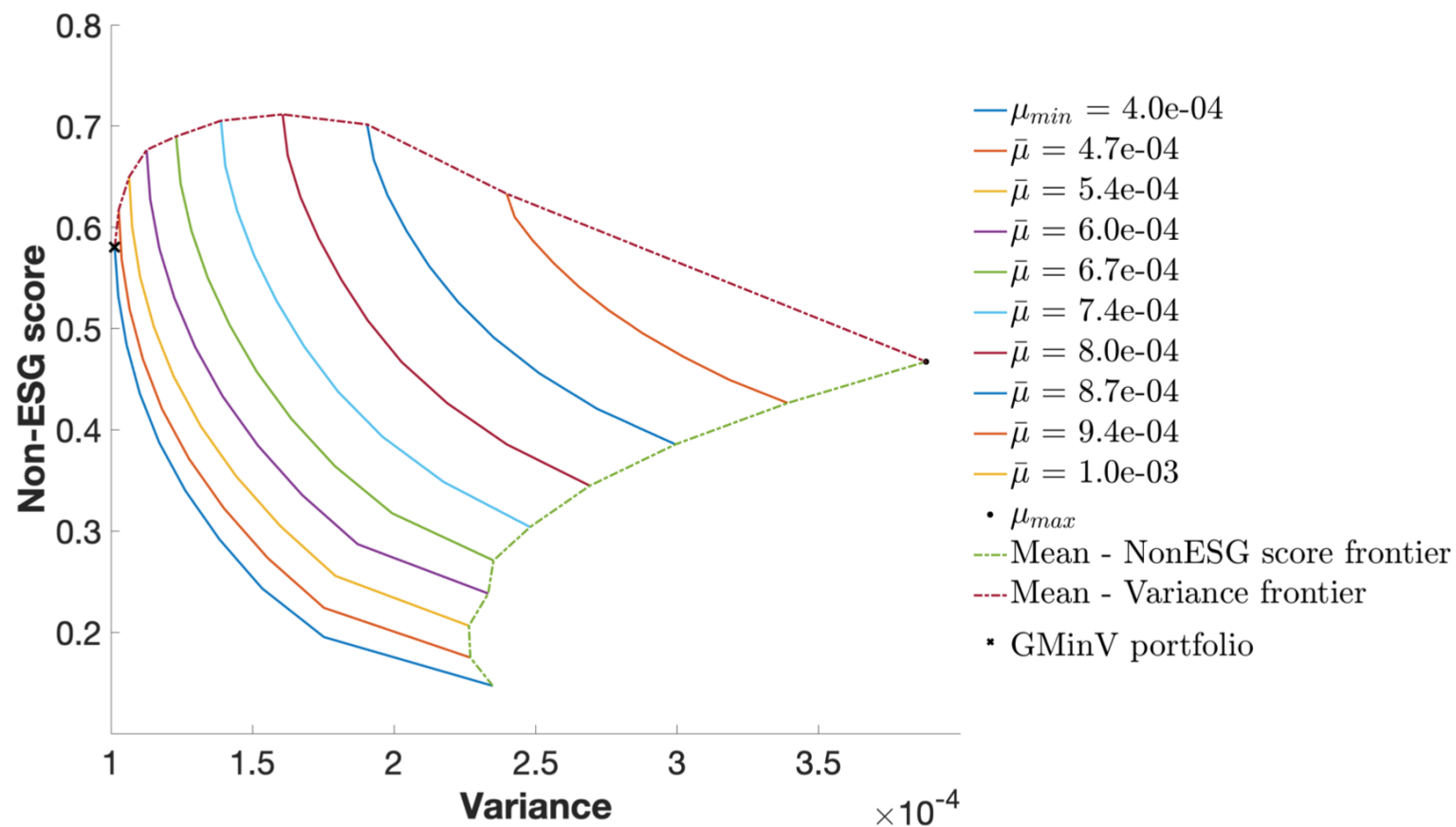
Theoretical Framework (7)

- To obtain a single-objective formulation of our Problem, we apply the **ϵ -constraint** method (see *Theorem 4.6* in [Ehrgott 2005]):

$$\begin{aligned}
 & \min_{(x,v,u)} && x^T \Sigma x \\
 & \text{s.t.} && \\
 & && \mu^T x \geq \bar{\mu} \\
 & && qu + \sum_{i=1}^m v_i \leq \bar{\gamma} \\
 & && v_i + u \geq (s^i)^T x \quad i = 1, \dots, m \\
 & && \sum_{k=1}^n x_k = 1 \\
 & && u \geq 0 \\
 & && v_i \geq 0 \quad i = 1, \dots, m \\
 & && x_k \geq 0 \quad k = 1, \dots, n
 \end{aligned} \tag{4}$$

Finding the three-objective efficient surface (2)

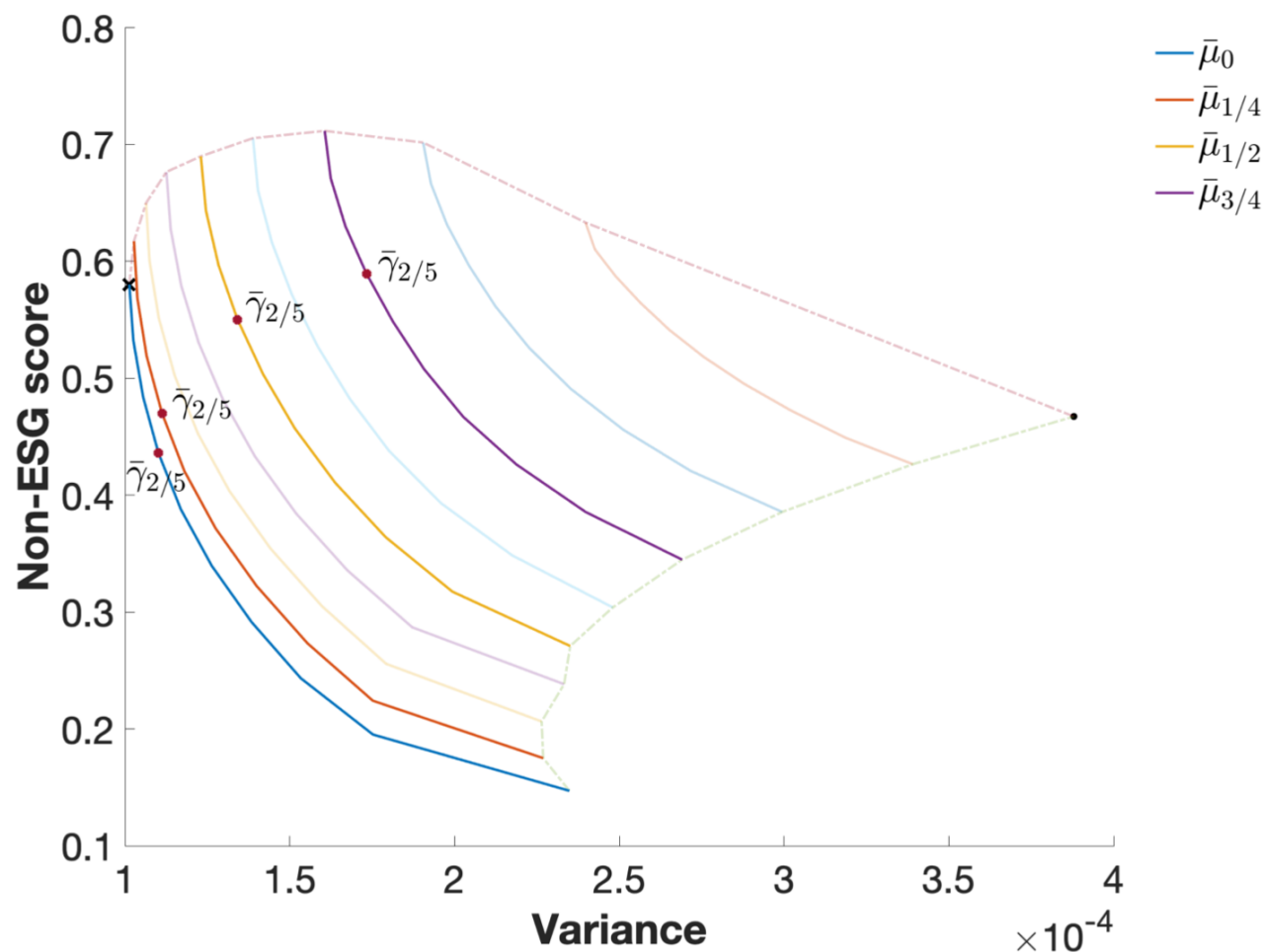
- The Pareto-optimal solutions of problem (4) are obtained by appropriately varying the target level of the portfolio expected return $\bar{\mu}$ and the target level of the portfolio Non-ESG score $\bar{\gamma}$



Experimental setup



- Our empirical analysis is based on **4 Pareto-optimal portfolios** obtained from Problem (4):



Mean-Variance-ESG (Cesarone et al. 2022)

- For comparison purposes, we consider the three-objective Mean-Variance-ESG optimization model in [Cesarone et al. 2022], where they include in the Mean-Variance framework, the maximization of the portfolio expected ESG

$$\min_x (\sigma_P^2(x), -\mu_P(x), -ESG_P(x))$$

s.t.

$$x \in \Delta = \left\{ x \in \mathbb{R}^n : \sum_{k=1}^n x_k = 1, x_k \geq 0, k = 1, \dots, n \right\}$$

With, where $ESG_P(x) = \sum_{k=1}^n ESG_k x_k$, where ESG_k denotes the expected ESG score assigned to asset k by a **single** ratings provider.

Mean-Variance-ESG (Cesarone et al. 2022)



Similarly, we select several Mean-Variance-ESG Pareto-optimal portfolios by appropriately varying the target level of the portfolio expected return η and the target level of the portfolio expected ESG λ :

$$\begin{aligned} \min_x \quad & \sum_{k=1}^n \sum_{j=1}^n x_k x_j \sigma_{kj} \\ \text{s.t.} \quad & \sum_{k=1}^n \mu_k x_k \geq \eta \\ & \sum_{k=1}^n ESG_k x_k \geq \lambda \\ & \sum_{k=1}^n x_k = 1 \\ & x_k \geq 0 \quad k = 1, \dots, n \end{aligned}$$

Datasets and ratings providers

We investigate the behaviour of the optimal portfolios by considering:

- four rating providers:

Rating provider	Time	Frequency
Refinitiv (Datastream)	Jan 2016–Dec 2021	Monthly
Bloomberg	Jan 2016–Dec 2021	Yearly
RobecoSAM	Jan 2016–Dec 2021	Yearly
Sustainalytics (Morningstar)	Jan 2016–Dec 2021	Monthly

- two real-world datasets with daily frequency

Market Index	#Assets	Country	Time Interval
EuroStoxx50	44	EU	Jan 2016–Dec 2021
NASDAQ100	83	USA	Jan 2016–Dec 2021

Summary of portfolio strategies analyzed



Table: List of portfolio strategies

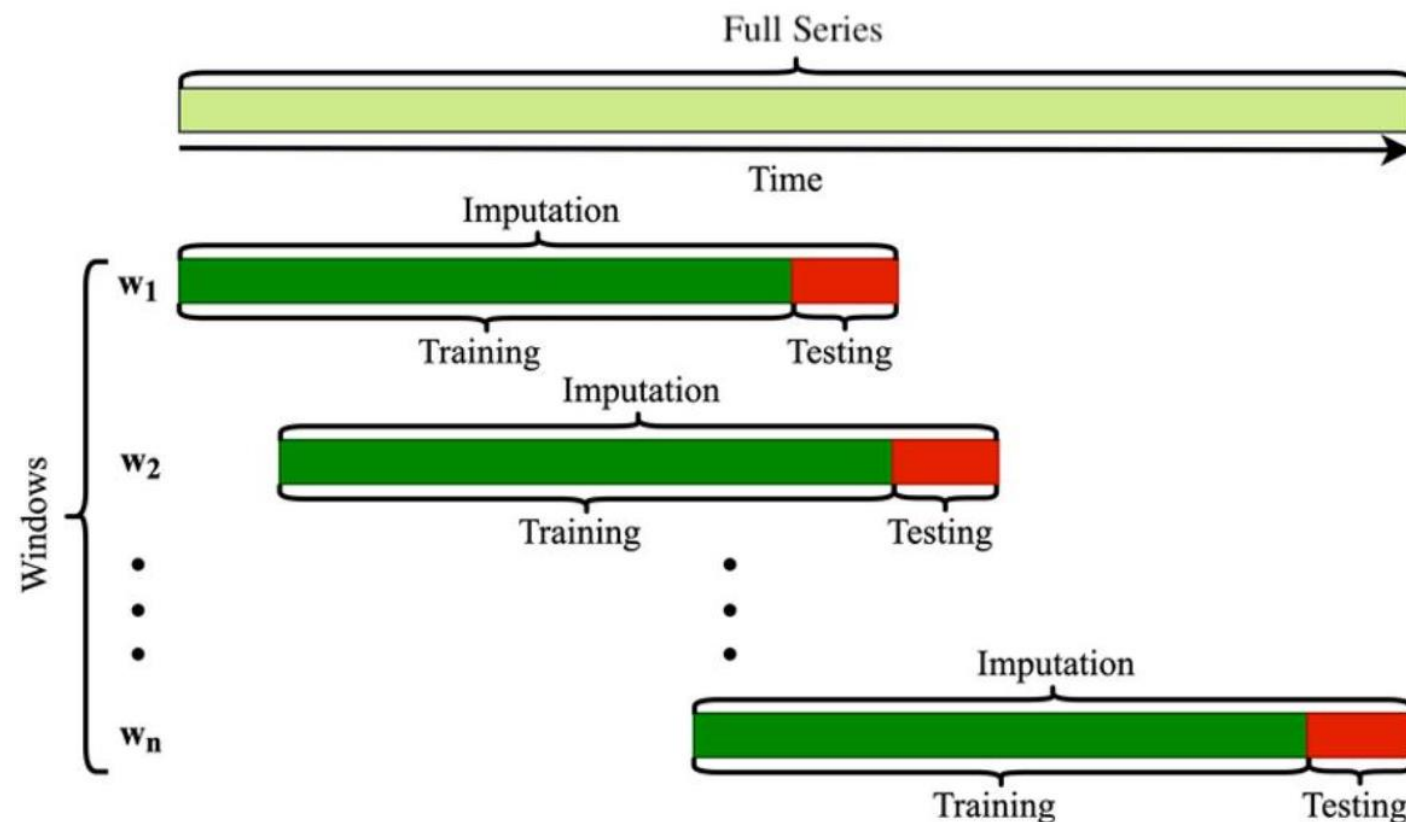
Approach	Abbreviation
Global Minimum-Variance portfolio	GMinV
Equally Weighted	EW
Risk Parity	RP
Most Diversified Portfolio	MDP
<i>Mean-Variance-ESG optimization strategy (Cesarone et al. 2022)</i>	
MV-ESG: Low Gain (η_0)	Sust_1
MV-ESG: Intermediate Low Gain ($\eta_{1/4}$)	Sust_2
MV-ESG: Intermediate High Gain ($\eta_{1/2}$)	Sust_3
MV-ESG: High Gain ($\eta_{3/4}$)	Sust_4
<i>k – sum optimization strategy</i>	
MV q –Worst NonESG: Low Gain ($\bar{\mu}_0$)	Sust_1_qWorst
MV q –Worst NonESG: Intermediate Low Gain ($\bar{\mu}_{1/4}$)	Sust_2_qWorst
MV q –Worst NonESG: Intermediate High Gain ($\bar{\mu}_{1/2}$)	Sust_3_qWorst
MV q –Worst NonESG: High Gain ($\bar{\mu}_{3/4}$)	Sust_4_qWorst

Rolling time windows evaluation scheme



The empirical analysis is based on a **rolling time windows approach**:

- We consider **in-sample** time windows of **2 year**.
- We assess the portfolio performance in the following **1 month**, called **out-of-sample** window.
- The out-of-sample performance of each portfolio strategy is evaluated using the Market Index (**Index**) as a benchmark.



Performance measures

To evaluate the **out-of-sample** of the optimal portfolios, we consider the following performance measures:

- Mean (**ExpRet**), volatility (**Vol**), Sharpe ratio (**Sharpe**) [Sharpe 1966, 1994];
- Maximum drawdown (**MDD**) [Chekhlov et al. 2005], ulcer index (**Ulcer**) [MacCann, 1989], Rachev ratio (with a confidence level equal to 0.10, named **Rachev10**) [Rachev et al. 2004], turnover (**Turn**) [DeMiguel et al. 2009];
- Jensen's Alpha (**AlphaJ**) [Jensen, 1968], Information ratio (**InfoR**) [Treyner and Black, 1973], Value-at-Risk (with a confidence level equal to 0.05, named **VaR5**), Omega ratio (**Omega**) [Harlow and Rao, 1989], the average number of selected assets (**ave#**)
- Return On Investment (**ROI**).

Out-of-sample performance results (EuroStoxx 50)



Table: EuroStoxx 50 (with $q = 1$)

Approach	ExpRet	Vol	Sharpe	MDD	Ulcer	Rachev10	Turn	AlphaJ	InfoRatio	VaR5	Omega	ave #
MV0	0.042%	0.939%	4.45%	-0.321	8.53%	0.920	0.13	0.025%	2.46%	1.22%	1.145	12
EW	0.040%	1.297%	3.11%	-0.407	9.44%	0.919	-	0.014%	7.29%	1.78%	1.105	44
RP	0.039%	1.201%	3.26%	-0.387	8.79%	0.905	0.02	0.015%	7.77%	1.60%	1.110	44
MDP	0.050%	1.040%	4.82%	-0.331	6.08%	0.920	0.13	0.031%	4.26%	1.47%	1.155	17
Sust_1	0.029%	0.976%	2.93%	-0.324	10%	0.899	0.17	0.011%	0.50%	1.36%	1.094	16
Sust_2	0.035%	0.979%	3.60%	-0.320	9.36%	0.883	0.22	0.018%	1.53%	1.38%	1.115	15
Sust_3	0.054%	1.108%	4.90%	-0.301	7.70%	0.896	0.36	0.036%	4.03%	1.76%	1.154	11
Sust_4	0.070%	1.318%	5.28%	-0.304	8.41%	0.922	0.43	0.048%	5.21%	2.13%	1.162	7
Sust_1_qWorst	0.063%	0.972%	6.52%	-0.302	5.98%	0.956	0.27	0.047%	5.38%	1.26%	1.217	12
Sust_2_qWorst	0.069%	0.992%	6.97%	-0.298	5.93%	0.945	0.29	0.052%	6.17%	1.33%	1.233	12
Sust_3_qWorst	0.087%	1.157%	7.51%	-0.296	6.66%	0.930	0.37	0.068%	7.93%	1.84%	1.243	8
Sust_4_qWorst	0.092%	1.373%	6.70%	-0.304	8.14%	0.925	0.45	0.070%	7.46%	2.16%	1.209	6

Out-of-sample ROI (EuroStoxx 50)



Table: ROI based on a 3-years time horizon (with $q = 1$)

Approach	ExpRet	Vol	5%-perc	25%-perc	50%-perc	75%-perc	95%-perc
MV0	21%	11%	9%	13%	17%	30%	41%
EW	30%	16%	1%	19%	31%	43%	56%
RP	28%	14%	4%	19%	28%	39%	51%
MDP	38%	12%	25%	28%	33%	48%	59%
Sust_1	13%	5%	4%	9%	12%	16%	23%
Sust_2	17%	7%	7%	12%	16%	21%	29%
Sust_3	38%	15%	21%	25%	33%	50%	65%
Sust_4	51%	20%	29%	33%	43%	68%	88%
Sust_1_qWorst	49%	13%	29%	40%	47%	60%	71%
Sust_2_qWorst	55%	15%	34%	43%	52%	66%	82%
Sust_3_qWorst	76%	23%	49%	57%	71%	95%	118%
Sust_4_qWorst	78%	28%	46%	55%	70%	102%	129%

Out-of-sample performance results (NASDAQ100)



Table: NASDAQ100 (with $q = 1$)

Approach	ExpRet	Vol	Sharpe	MDD	Ulcer	Rachev10	Turn	AlphaJ	InfoRatio	VaR5	Omega	ave #
MV0	0.045%	1.124%	3.99%	-0.319	6.83%	0.901	0.15	-0.013%	-5.80%	1.43%	1.145	19
EW	0.108%	1.435%	7.53%	-0.312	5.61%	0.930	-	0.016%	1.76%	2.05%	1.265	83
RP	0.094%	1.319%	7.10%	-0.306	5.02%	0.931	0.02	0.010%	-1.50%	1.81%	1.256	83
MDP	0.102%	1.269%	8.01%	-0.316	4.69%	1.017	0.15	0.031%	0.10%	1.63%	1.291	20
Sust_1	0.044%	1.127%	3.90%	-0.299	6.94%	0.896	0.16	-0.014%	-5.96%	1.48%	1.139	18
Sust_2	0.066%	1.197%	5.52%	-0.289	5.58%	0.941	0.32	0.0001%	-4.14%	1.62%	1.194	18
Sust_3	0.170%	1.983%	8.58%	-0.364	9.51%	0.981	0.59	0.064%	5.95%	3.19%	1.287	11
Sust_4	0.281%	2.738%	10.25%	-0.440	13.13%	1.043	0.48	0.152%	9.10%	4.47%	1.348	7
Sust_1_qWorst	0.059%	1.266%	4.62%	-0.312	7.52%	0.908	0.36	-0.010%	-4.87%	1.76%	1.158	17
Sust_2_qWorst	0.078%	1.385%	5.64%	-0.314	8.07%	0.922	0.46	0.001%	-2.74%	2.01%	1.188	15
Sust_3_qWorst	0.182%	2.144%	8.48%	-0.375	11.22%	0.982	0.57	0.071%	5.97%	3.32%	1.277	9
Sust_4_qWorst	0.285%	2.858%	9.98%	-0.434	13.79%	1.045	0.47	0.156%	8.62%	4.33%	1.334	6

Out-of-sample ROI (NASDAQ100)



Table: ROI based on a 3-years time horizon (with $q = 1$)

Approach	ExpRet	Vol	5%-perc	25%-perc	50%-perc	75%-perc	95%-perc
MV0	34%	9%	22%	26%	33%	40%	52%
EW	121%	18%	100%	109%	116%	132%	159%
RP	100%	14%	82%	92%	97%	108%	128%
MDP	117%	14%	94%	108%	117%	124%	144%
Sust_1	35%	10%	22%	27%	35%	40%	54%
Sust_2	64%	12%	52%	57%	60%	69%	94%
Sust_3	287%	55%	227%	246%	270%	315%	414%
Sust_4	825%	215%	603%	668%	747%	929%	1344%
Sust_1_qWorst	44%	14%	30%	34%	39%	55%	72%
Sust_2_qWorst	72%	19%	50%	55%	69%	83%	113%
Sust_3_qWorst	321%	81%	229%	260%	298%	365%	511%
Sust_4_qWorst	862%	243%	599%	690%	777%	992%	1453%

On the ESG ratings divergence

- For each market index, we calculate the average divergence of ESG ratings assigned by the data providers to all assets, using several distance metrics:

Market Index	Euclidean	Chebychev	Cosine	Correlation
NASDAQ100	10.22	0.40	0.05	0.35
EuroStoxx 50	16.00	0.56	0.07	0.75

- The different data providers seem to agree more among themselves in assigning ESG ratings in the **US market** (i.e. the NASDAQ100) than in the **European market** (i.e. the EuroStoxx 50).
- This might justify the greater contribution of our k -sum optimization approach in the case of markets characterised by more intense disagreement between rating providers.

Conclusion (1)



- We have introduced a new general method to manage the ESG ratings disagreement between providers for portfolio selection purposes.
- We have included in the Mean-Variance framework a third objective represented by the q -Worst Non-ESG score. Such a goal has been treated by using the k -sum operator.
- The resulting Mean-Variance-(Robust)ESG has been reformulated as a convex quadratic programming problem.

Conclusion (2)

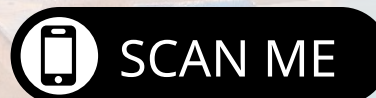


- The empirical analysis based on EuroStoxx 50 and NASDAQ100 datasets confirms that requiring higher sustainability to the selected portfolios improves their performance.
- Comparing the MVESG and MVqW-ESG portfolios, the latter show better results. This means that the portfolio sustainability seems to be better represented by the k -sum operator which considers all providers' information, with respect to the ESG evaluated by a single provider.
- Furthermore, it seems that the greater the disagreement between the rating providers, the more effective the use of the proposed q -Worst Non-ESG score.

Thank you!!

Manuel Luis Martino – manuelluis.martino@uniroma3.it

→ Check out the paper!



Bibliography (1)



- Berg, F., Koelbel, J.F., Rigobon, R. **(2022)**. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26: 1315–1344.
- PRI **(2023)**: Quarterly Signatory Update.
- Friede, G., Busch, T., Bassen, A. **(2015)**. ESG and Financial Performance: Aggregated Evidence from More Than 2000 Empirical Studies. *Journal of Sustainable Finance & Investment*, 5: 210–233.
- Cesarone, F., Martino, M.L., Carleo, A. **(2022)**. Does ESG Impact Really Enhance Portfolio Profitability? *Sustainability*, 14(4) 2050.
- Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L. **(2021)**. Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28: 1426–1445.

Bibliography (2)



- Chatterji, A. K., Durand, R., Levine, D. I., Touboul S. **(2016)**. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8:) 1597–1614.
- Gibson Brandon, R., Krueger, P., Schmidt, P.S. **(2021)**. ESG rating disagreement and stock returns. *Financial Analysts Journal*, 77: 104–127.
- Ponce, D., Puerto, J., Ricca, F., Scozzari, A. **(2018)**. Mathematical programming formulations for the efficient solution of the k-sum approval voting problem. *Computers & Operations Research*, 98: 127–136.
- Puerto, J., Rodríguez-Chá, A. M., Tamir, A. **(2017)**. Revisiting k-sum optimization. *Mathematical Programming*, 165: 579–604.