Theoretical and Applied Economics
Volume XXXI (2024), No. 1(638), Spring, pp. 161-168

ESG risk rating disagreement: implications on portfolio performance

Diana-Mihaela SANDU
Bucharest University of Economic Studies, Romania
sandu1mihaela16@stud.ase.ro

Abstract. *This paper examines the ESG risk rating disagreement across two-well established rating providers and its implication on portfolio performance. By deriving a proxy for rating disagreement using the average standard deviation of pairwise percentile ranking across Refinitiv and Sustainalytics, this study examined the risk-adjusted performance of high and low disagreement portfolios. For each portfolio, four risk-adjusted measures (Sharpe ratio, Treynor ratio, Modigliani-Squared and Jensen's alpha) were calculated. In general, the study found that the best performer was the low-disagreement portfolio, but the results were not favourable for any portfolio.*

Keywords: ESG disagreement, portfolio performance, risk-adjusted measures, industry, Europe.

JEL Classification: G30, M14.

Introduction

The importance of environmental, social and governance factors has increased among investors (Aouadi and Marsat, 2018; Shakil, 2021). Investors are increasingly considering ESG performance and risks in investment decisions. One specific challenge for investors is the lack of a common framework for ESG raters. Therefore, divergences and disagreements between rating providers have real consequences for users of ESG ratings, especially for investors (Chatterji et al., 2016).

Given the reason for the differences in ESG ratings, recent research has attracted a lot of attention, in particular, to how ESG ratings are constructed and the disagreement between rating providers (Chatterji et al., 2016; Gibson et al., 2021; Avramov et al., 2022; Berg et al., 2022; Tan and Pan, 2023). However, a review of the literature has shown limited attention to the role of ESG rating disagreement in portfolio performance.

This study contributes to the literature by analysing the ESG risk rating disagreement across two-well established rating providers and its implication on portfolio performance. By deriving a proxy for rating disagreement based on ESG risk ratings from Refinitiv and Sustainalytics, this study tests the risk-adjusted performance of high and low disagreement portfolios. Four risk-adjusted measures (Sharpe ratio, Treynor ratio, Modigliani-Squared and Jensen's alpha) were calculated for each portfolio.

The sample consists of European companies, and the period analysed is January 2022 to December 2022. The ESG data sample is based on two risk rating providers, Refinitiv and Sustainalytics. The ESG Refinitiv Controversies Score is calculated based on 23 ESG Controversies Topics. During the year, if a scandal in the media occurs, the company involved is penalised, and this affects grading. Sustainalytics ESG risk rating measures a company's exposure to industry-specific ESG risks and how well it manages these risks. Sustainalytics' ESG risk ratings are categorised into five risk levels: negligible (0-10), low (10-20), medium (20-30), high (30-40) and severe (40+), while Refinitiv applies a scale of 0 to 100. Consequently, the absence of a common methodology determined the creation of a rating disagreement between these two rating providers. ESG rating disagreement was measured as the average standard deviation of pairwise percentile ranking between Refinitiv and Sustainalytics. In addition, an industry-adjusted ESG rating disagreement has been calculated, since correlations between ESG rating providers revealed significant industry differences. In general, the study found that the best performer was the low-disagreement portfolio, but the results were not favourable for any portfolio. These findings demonstrate the importance of standard reporting frameworks among ESG rating providers.

The remainder of the paper is structured as follows. Section 2 presents the theoretical background. The methodology and data are described in Section 3. Section 4 presents and discusses the results. Finally, the last section concludes.

Literature review

The existing literature examined how ESG ratings are constructed and the disagreement between rating providers (Chatterji et al., 2016; Gibson et al., 2021; Avramov et al., 2022; Berg et al., 2022; Tan and Pan, 2023). Chatterji et al. (2016) studied six ESG rating

providers (KLD Research & Analytics, Refinitiv, Calvert, FTSE4Good, DJSI and Innovest) and found a fair lack of correlation between these providers. Gibson et al. (2021) studied the relationship between stock returns and disagreement among ESG raters (Refinitiv, Sustainalytics, Inrate, Bloomberg, FTSE, KLD Research & Analytics and MSCI IVA). Their findings suggested that stock returns are positively related to ESG rating disagreement. Similarly, Avramov et al. (2022) showed that rating disagreement among six rating providers (Refinitiv, MSCI IVA, Bloomberg, Sustainalytics and S&P Global) leads to higher perceived market risk, higher market premium, and lower investor demand. Furthermore, Tan and Pan (2023) found that ESG rating disagreement among six agencies (Hexun, Huazheng, SynTao Green Finance, Wind, RKS and Bloomberg) has a significantly inverse influence on both stock returns and volatility.

However, a review of the literature has shown limited attention to the role of ESG rating disagreement in portfolio performance. Some studies have examined the relationship between ESG rating and portfolio performance, comparing ESG and non-ESG portfolio performance. Most studies have shown that ESG portfolios outperform non-ESG portfolios (Statman, 2005; Goyal and Aggarwal, 2014; Kumar et al., 2016; Diaz et al., 2021). Statman (2006) compared four ESG indexes (Domini 400 Social Index, Calvert Social Index, Citizens Index, Dow Jones Sustainability Index) to the conventional companies of the S&P 500 Index. The results show that the Domini 400 Social Index (DS 400) returns and Sharpe ratios are higher than those of the S&P 500 Index. Goyal and Aggarwal (2014) showed that the ESG stocks portfolios had outperformed the blue chip and market portfolios in the Indian context. Furthermore, Kumar et al. (2016) demonstrated that ESG companies generate higher returns and lower volatility in the majority industries. Consistent with the studies mentioned above are the findings of Diaz et al. (2021). The study evaluated the performance of the Top 25 ESG and Bottom 25 ESG portfolios of US companies compared to the S&P 500 Index from January to April 2020. The authors found that Top25 ESG outperforms the S&P 500 Index, while Bottom25 ESG underperforms the S&P 500 Index. On the contrary, some studies have shown that ESG portfolios underperformed non-ESG portfolios (Hong and Kacperczyk, 2009; Bolton and Kacperczyk, 2021; Luo, 2022).

On the basis of the prior literature, the following conclusions can be drawn. First, when comparing ESG portfolios with other portfolios, some studies showed that ESG portfolios outperform other portfolios, while others showed the opposite results. Second, ESG ratings differ from one rating provider to another. Finnaly, disagreement on the ESG risk rating and its implication on portfolio performance is not sufficiently analysed in the literature.

Data and Methodology

The sample consists of ESG ratings and financial data from listed European companies. The original sample selected for analysis consists of all listed European companies that are included in the database of Refinitiv and Sustainalytics. After data filtration, the sample contains 1004 companies with available ESG ratings. The daily closing prices for the period from January 3, 2022 to December 29, 2022 were obtained from the Refinitiv database. Table 1 shows the geographical distribution of the companies, with the sample being mainly from Germany, France, Sweden and Switzerland.

Table 1. *Geographical distribution of the sample*

Country	Freq.	Percent
Austria	30	2.99
Belgium	43	4.28
Cyprus	1	0.10
Czech Republic	3	0.30
Denmark	39	3.88
Finland	33	3.29
France	123	12.25
Germany	138	13.75
Greece	22	2.19
Hungary	5	0.50
Ireland	43	4.28
Italy	74	7.37
Luxembourg	20	1.99
Malta	3	0.30
Netherlands	63	6.27
Norway	37	3.69
Poland	32	3.19
Portugal	11	1.10
Romania	1	0.10
Spain	60	5.98
Sweden	114	11.35
Switzerland	109	10.86
Total	1004	100.00

Source: Authors' own research.

Portfolio classification is based on a proxy of disagreements measured by the standard deviation of the rating available to a given company at a given time. According to Avramov et al. (2022) and Tan and Pan (2022), disagreement over the ESG ratings is calculated using the standard deviation of the percentile rank between the rating pairs. First, all companies were sorted according to the initial rating covered by both suppliers and the percentile rank was calculated. Afterwards, for each company, the standard deviation of the percentile rank was calculated as $\frac{|ri,A-ri,B|}{\sqrt{2}}$, where $r_{i,A}$ denote the ESG rank for company i from rater Refinitiv and $r_{i,B}$ denote the ESG rank for company i from Sustainalytics.

Table 2. *Industry classification of the sample*

Industry	Freq.	Percent
Basic Materials	92	9.16
Consumer Cyclicals	142	14.14
Consumer Non-Cyclicals	72	7.17
Energy	41	4.08
Financials	147	14.64
Healthcare	91	9.06
Industrials	202	20.12
Real Estate	58	5.78
Technology	118	11.75
Utilities	41	4.08
Total	1004	100.00

Source: Authors' own research.

In addition, Gibson et al. (2021) discovered that correlations between ESG rating providers revealed significant industry differences. Consequently, an industry-adjusted ESG rating disagreement has been calculated. In this calculation, the Thomson Reuters Business Classification (TRBC) was applied to classify companies into 10 industries (basic

materials, consumer cyclicals, non-consumer cyclicals, energy, financials, healthcare, industrials, real estate, technology and utilities). As shown in Table 2, the three main industries represented are industrials, financials and consumer cyclicals.

For each of the ten industries, the industry-specific disagreement rating was calculated as $\sqrt{\frac{(ri,A-\overline{rA})^2\times(ri,B-\overline{rB})^2}{2}}$, where $r_{i,A}$ and $r_{i,B}$ denote the ESG rank for company i from Refinitiv and Sustainalytics, respectively, and $\overline{rA}$ and $\overline{rB}$ denote the industry average of the Refintiv and Sustainalytics ranks. In order to obtain the industry-adjusted ESG rating disagreement, the industry-specific disagreement rating was subtracted from the company-specific disagreement rating initially calculated.

$$(1)\,\text{Company} - \text{specific ESG disagreement} = \frac{|ri,A-ri,B|}{\sqrt{2}},$$

where $r_{i,A}$ denote the ESG rank for company i from rater Refinitiv and $r_{i,B}$ denote the ESG rank for company i from Sustainalytics.

$$(2)\,\text{Industry} - \text{specific ESG disagreement} = \sqrt{\frac{(ri,A-\overline{rA})^2\times(ri,B-\overline{rB})^2}{2}},$$

where $r_{i,A}$ and $r_{i,B}$ denote the ESG rank for company i from Refinitiv and Sustainalytics, respectively, and $\overline{rA}$ and $\overline{rB}$ denote the industry average of the Refintiv and Sustainalytics ranks.

$$(3)\,Industry - adjusted\ ESG\ disagreement = Company - specific\ ESG\ disagreement - Industry - specific\ ESG\ disagreement$$

The descriptive statistics of the rating providers are also relevant (Table 3). Given the minimum and maximum values of scores, Refintiv clearly uses a scale of 0 to 100 compared to Sustainalytics, which stops at 50. The standard deviation of the results is the highest for Refinitiv, which means a greater spread in these particular ratings.

Table 3. *Descriptive statistics on rating providers*

Variable	Obs	Mean	Std. Dev.	Min	Max
Refinitiv	1004	89.537	23.786	.439	100
Sustainalytics	1004	20.386	7.241	4.8	49.7
Company-specific ESG disagreement	1004	.251	.176	0	.670
Industry-specific ESG disagreement	1004	.005	.007	0	.054
Industry-adjusted ESG disagreement	1004	.246	.174	-.020	.645

Source: Authors' own research, using Stata

To build portfolios, the companies are sorted from high to low according to their respective industry-adjusted ESG rating disagreement. Similar to Diaz et al. (2021), the sample was divided into quartiles, and the companies in the top 25% are included in the Top25 ESG disagreement portfolio (high-disagreement portfolio) and the companies with ranks in the bottom quartile are included in the Bottom25 ESG disagreement portfolio (low-disagreement portfolio). Based on these classifications, the equally weighted portfolio return was calculated as $\frac{1}{n}\sum_{s=1}^{n} R_s$, and stock return ($R_S$) was calculated as $ln\frac{P_t}{P_{t-1}}$, where P=closing prices, t=day, n=number of shares in the portfolio.

The descriptive statistics considering the portfolio returns are displayed in order to provide an overview and understanding of the behaviour of the portfolio. Table 4 shows the mean, variance, minimum, maximum, skewness and kurtosis for each portfolio. Skewness and kurtosis represent a normal distribution if their values are 0 and 3, respectively. The skewness of the Top25 portfolio is negative, indicating a left-skewed distribution of returns, while the skewness of the Bottom25 portfolio is positive, implying a right-skewed distribution of returns. Furthermore, kurtosis lower than 3 indicates a platykurtic distribution for both portolios.

Table 4. *Descriptive statistics*

	Mean	Variance	Min.	Max.	Skew.	Kurt.
Top25 disagreement portfolio	-0.00004	0.00018	-0.04348	0.04625	-0.17693	1.48093
Bottom25 disagreement portfolio	-0.00001	0.00019	-0.04128	0.04715	0.01608	1.23559

Source: Authors' own research, using Stata.

Following Sood et al. (2022), the risk-adjusted performance measures used to evaluate the ESG rating disagreement portfolios are as follows:

1. Sharpe ratio is the fraction between the excess portfolio return above the risk-free rate and the portfolio's standard deviation. It is interpreted as the excess return per unit of risk (Sharpe, 1966).

$$(4)\ \text{Sharpe ratio} = \frac{R_P - R_F}{\sigma_P},$$

where R_P =portfolio return, R_F =risk-free rate, σ_P =standard deviation of the portfolio return.

To calculate risk-free rate, ESTER (Euro Short-Term Rate) was the interest rate for the Euro zone recommended by the European Working group on risk-free rates.

2. Treynor ratio is the excess return per unit of portfolio systematic risk (Treynor, 1962). It seems similar to the Sharpe ratio but different by referring to the systematic risk defined by portfolio betas.

$$(5)\ \text{Treynor ratio} = \frac{R_P - R_F}{\beta_P},$$

where R_P =portfolio return, R_F =risk-free rate, $\beta_P = \frac{covariance(R_P, R_M)}{variance(R_M)}$, R_M=proxy for market return (MSCI Europe Index).

The MSCI Europe Index represents large and mid-cap companies in 15 developed markets countries in Europe. With 425 components, the index covers approximately 85% of the free-float adjusted market capitalisation across the European developed markets.

3. Modigliani-squared multiplies Sharpe's ratio by the standard deviation of the benchmark index and then adds a risk-free rate (Modigliani and Modigliani, 1997).

$$(6)\ \text{Modigliani} - \text{squared} = \frac{R_P - R_F}{\sigma_P} \times \sigma_M + R_F,$$

where R_P=portfolio return, R_F=risk-free rate, σ_P=standard deviation of the portfolio return σ_M=standard deviation of the benchmark (MSCI Europe Index).

4. Jensen's alpha measures the excess return that the portfolio generates over the expected return and a positive alpha implies greater performance (Jensen, 1968).

$$(7)\,\text{Jensen}'s\ \text{alpha} = \mathrm{R_P} - (R_F + \beta_P(R_M - R_F)),$$

where R_P=portfolio return, R_F=risk-free rate, β_P=portfolio beta, $\mathrm{R_M}$=proxy for market return (MSCI Europe Index).

Results

Table 5 presents the risk-adjusted measures used to evaluate the ESG rating disagreement portfolios. The Sharpe ratio measures the risk premium per unit of risk. The Sharpe ratio is better for low-disagreement portfolio, but it has negative values for both portfolios. This results occurs because the portfolio has earned a negative return and the investment return is lower than the risk-free rate, which means a investment in these portfolios should be avoided. The Treynor ratio, is similar to the Sharpe ratio, a measure of the risk premium per unit of risk, but the Treynor ratio applies systematic risk in the form of beta. However, the Treynor ratio provides different results, with the high-disagreement portfolio acting as the best performer with a ratio of 0.0011 compared to 0.0002 for the low-disagreement portfolio. The Modigliani-squared provides results of the same magnitude as the Sharpe ratio, with the low-disagreement portfolio acting as the best performer with a ratio of 0.00075 compared to 0.00072 for the high-disagreement portfolio. The Jensen's alpha measures the excess return that the portfolio generates over the expected return. The negative alpha indicates that both portfolios performed worse than expected. Overall, the best performer is the low-disagreement portfolio, but the results are not favourable to any portfolio.

Table 5. *Risk-adjusted measures*

	Top25 disagreement portfolio	Bottom25 disagreement portfolio
Sharpe ratio	-0.00284	-0.00106
Treynor ratio	0.00106	0.00024
Modigliani-squared	0.00072	0.00075
Jensen's alpha	-0.00007	-0.00006

Source: Authors' own research.

Conclusions

This study analysed the ESG risk rating disagreement across two-well established rating providers and its implication on portfolio performance. By deriving a proxy for rating disagreement based on ESG risk ratings from Refinitiv and Sustainalytics, this study examined the risk-adjusted performance of high and low disagreement portfolios. The ESG rating disagreement was measured as the average standard deviation of pairwise percentile ranking across Refinitiv and Sustainalytics. In addition, an industry-adjusted ESG rating disagreement has been calculated, since correlations between ESG rating providers revealed significant industry differences. For each portfolio, four risk-adjusted measures, Sharpe ratio, Treynor ratio, Modigliani-Squared and Jensen's alpha, were calculated. In general, the study found that the best performer was the low-disagreement portfolio, but the results were not favourable for any portfolio. These findings demonstrate the importance of standard reporting frameworks among ESG rating providers.

Overall, the study contributes to the existing literature by examining the rating disagreement across two ESG risk ratings providers and analysing its implication on portfolio risk-adjusted performance. Furthermore, this study contributed to the recognition of differences in ESG risk scores, calculated by some raters compared to the ESG performance scores that attracted the attention of the agency ratings.

References

Aouadi, A. and Marsat, S., 2018. Do ESG controversies matter for firm value? Evidence from international data. *Journal of Business Ethics*, 151, pp. 1027-1047.

Avramov, D., Cheng, S., Lioui, A. and Tarelli, A., 2022. Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, *145*(2), pp. 642-664.

Berg, F., Koelbel, J.F. and Rigobon, R., 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), pp. 1315-1344.

Bolton, P. and Kacperczyk, M., 2021. Do investors care about carbon risk? *Journal of financial economics*, 142(2), pp. 517-549.

Chatterji, A.K., Durand, R., Levine, D. and Touboul, S., 2016. Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers. *Strategic Management Journal*, pp. 1597-1614.

Gibson, Brandon, R., Krueger, P. and Schmidt, P.S., 2021. ESG rating disagreement and stock returns. *Financial Analysts Journal*, 77(4), pp. 104-127.

Goyal, M.M. and Aggarwal, K., 2014. ESG index is Good for Socially Responsible Investor in India. *Asian Journal of Multidisciplinary Studies*, 2(11), pp. 92-96.

Hong, H. and Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. *Journal of financial economics*, 93(1), pp. 15-36.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, 23(2), pp. 389-416.

Kumar, N.C.A., Smith, C., Badis, L., Wang, N., Ambrosy, P. and Tavares, R., 2016. ESG factors and risk-adjusted performance: a new quantitative model. *Journal of sustainable finance & investment*, 6(4), pp. 292-300.

Luo, D., 2022. ESG, liquidity, and stock returns. *Journal of International Financial Markets, Institutions and Money,* 78.

Modigliani, F. and Modigliani, L., 1997. Risk-adjusted performance. *Journal of portfolio management*, 23(2), pp. 45-54.

Shakil, M.H., 2021. Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity, *Resources Policy*, 72.

Sharpe, W.F., 1966. Mutual fund performance. *The Journal of business*, 39(1), pp. 119-138.

Sood, K., Arijit, K., Pathak, P. and Purohit, H.C., 2022. Did ESG portfolio augment investors' wealth during Covid19? Evidence from Indian stock market. *Sustainability Accounting, Management and Policy Journal* (ahead-of-print).

Statman, M., 2006. Socially responsible indexes. *The Journal of Portfolio Management*, 32(3), pp. 100-109.

Tan, R. and Pan, L., 2023. ESG rating disagreement, external attention and stock return: Evidence from China. *Economics Letters*, 231.

Treynor, J.L., 1962. Toward a Theory of Market Value of Risky Assets. Available at SSRN 628187.