



Perspektif Akuntansi
Volume 7 Nomor 3 (Oktober 2024), hal. 298-314
ISSN: 2623-0194 (Print), 2623-0186 (Online)
Copyright© The Authors(s). All Rights Reserved
Center for Accounting Development and Research (CARD)
Fakultas Ekonomika dan Bisnis,
Universitas Kristen Satya Wacana
DOI: <https://doi.org/10.24246/persi.v7i3.p298-314>
<http://ejournal.uksw.edu/persi>

Understanding the ESG and Financial Performance Relationship: Different Metrics Matters

Luluk Widyawati ¹

Faculty of Economics and Business, Universitas Indonesia

Received
12/06/2023

Revised
18/08/2023

Accepted
20/12/2023

Abstract. *Studies on the relationship between environmental, social, and governance (ESG) and a firm's financial performance remain largely inconclusive. This methodological study attempts to provide empirical evidence that the disagreement among ESG ratings lead to the different results of firm-level studies on the relationship between ESG and financial performance. Tests of multiple models using different ESG rating uncover three types of variations: statistical significance, directionality, and magnitude. This study also indicates that the effect is greater in studies on accounting-based financial performance measures and studies applying composite ESG score. Social dimension consistently presents the highest number of variations compared to environmental and governance dimensions.*

Keywords: *Environmental Social and Governance (ESG), CSR, Financial Performance, Rating, Return*

Abstrak. *Studi mengenai hubungan antara aspek lingkungan, sosial, dan tata kelola (ESG) dan kinerja keuangan perusahaan belum konklusif. Studi metodologis ini berupaya memberikan bukti empiris bahwa ketidaksepakatan antar pemeringkat ESG menyebabkan perbedaan hasil penelitian pada penelitian mengenai hubungan antara ESG dan kinerja keuangan. Pengujian terhadap beberapa model menggunakan peringkat ESG yang berbeda mengungkap tiga jenis variasi: signifikansi, arah, dan besarnya koefisien asosiasi. Penelitian ini juga menunjukkan bahwa pengaruh yang lebih besar terlihat pada model penelitian yang menggunakan ukuran kinerja keuangan berbasis akuntansi dan penelitian yang menerapkan skor agregat ESG. Dimensi sosial secara konsisten menyajikan jumlah variasi tertinggi dibandingkan dengan dimensi lingkungan dan tata kelola.*

Kata kunci: *Environmental Social And Governance (ESG), CSR, Kinerja Keuangan, Rating, Return*

¹ Corresponding author: lwidya@ui.ac.id

Introduction

Studies examining the impact of environmental, social and governance (ESG)² factors on firms' financial performance (FP) provide mixed results. The use of different ESG ratings might contribute to these differences in results. This methodological study aims to investigate the effect of using different ESG ratings on the relationship between ESG and FP (ESG–FP).

Meta-analyses of the abundant literature on ESG–FP relationship suggest methodological issues associated with the use of ESG ratings (Atz *et al.*, 2023; Gillan, Koch, and Starks 2021). The ratings are typically provided by third-party providers (Christensen *et al.*, 2022; Dremptic *et al.*, 2020). Kong *et al.* (2020) and conduct an extensive on the literature about corporate social responsibility (CSR) and firm performance. They conclude that the relationship between CSR and FP remains largely inconclusive despite the advancement in the methodology applied. They note that different studies applied different methods to measure CSR. This issue is corroborated by López-Arceiz *et al.* (2018), who point to the issue of the diversity of ESG and FP measurements within the ESG–FP literature. It is likely that the measurement diversity contributes to the differences in the results of ESG–FP studies (Gillan *et al.*, 2021).

Previous studies indicate that there is a low convergent validity of ESG ratings (Berg *et al.*, 2022; Chatterji *et al.*, 2016; Dorfleitner *et al.*, 2015). Empirical evidence also shows that different ESG rating applied different ESG measurement framework, which leads to a low agreement among the ratings (Widyawati 2021). Not surprisingly, a study by Halbritter and Dorfleitner (2015) demonstrates that the significance and magnitude of ESG's impact on stock returns are different depending on the ESG ratings applied. Considering the widespread use of ESG ratings within the ESG–FP literature, further investigation into whether the low agreement affects ESG–FP relationships on multitude of model is essential.

Little is known about the impact of applying different ESG ratings on other measures of FP. A meta-analysis by López-Arceiz *et al.* (2018) suggests that accounting-based FP measures are more likely to be correlated with ESG compared to market-based FP measures. However, the study did not examine the significance and magnitude differences. This paper contributes to the literature by identifying the type of variation caused by the application of different ESG ratings and the sensitivity level of different ESG-FP models. Therefore future researcher could be more informed when selecting ESG ratings as proxies and able to prepare appropriate treatment if necessary.

Specifically, this study is a methodological study that examines the impact of four different ESG ratings on two types of FP measures. Data for the ratings are collected from four rating agencies, namely Sustainalytics, Bloomberg, Thomson Reuters, and MSCI KLD. The agencies are among the most reputable ESG data providers and used

² This study used the term ESG to also represent the concepts of corporate social responsibility (CSR) and corporate social performance (CSP).

by the majority of investors (Wong *et al.*, 2019). These ratings have been shown to have a low agreement (Berg *et al.*, 2022; Billio *et al.*, 2021). The empirical framework in this study adopts models and variables commonly used in the related literature on ESG–FP (Khan *et al.*, 2016; Li *et al.*, 2018; Nollet *et al.*, 2016). Following López-Arceiz *et al.* (2018), this framework includes both accounting-based and market-based FP measures as dependent variables. Analysis of the empirical models is performed repetitively using four different ESG ratings to examine similarities and differences in the coefficients of ESG variables.

Results of the analysis show three variations of the ESG variable: statistical significance, directionality, and magnitude. The variations suggest that the application of different ESG ratings influence the conclusions of ESG–FP studies. Studies applying composite ESG scores and/or accounting-based FP measures are more sensitive to the selection of ESG ratings. Models applying ESG scores generally exhibit fewer variations compared to models applying dimensional ESG. Among the dimensions of ESG, the social dimension consistently presents the greatest variations. The extreme variations are likely due to the social dimension having the broadest diversity of topics and substantial differences in measurement approaches.

Literature Review

The majority of studies on ESG–FP relationship apply a firm-level analysis, including cross-regression analysis of ESG and FP variables (Kong *et al.*, 2020). The firm-level analysis enables an insightful examination of the explanatory power of ESG variables (Halbritter and Dorfleitner 2015). In addition, the effect of the selection of an ESG rating as a proxy for the ESG variable can be more easily observed. This section provides an overview of the heterogeneity of ESG–FP studies that apply a firm-level analysis approach.

There are two measures of FP generally used in firm-level ESG–FP studies: accounting-based measures and market-based measures (Gillan *et al.*, 2021; López-Arceiz *et al.*, 2018). Stock returns, which represent an *ex-post* FP measure, are among the most common of market-based measures examined. Examining the impact of ESG on the stock returns, Cao *et al.* (2023) used MSCI KLD ratings to represent ESG performance and found that different aspects of MSCI KLD ratings had different effects on stock return. Using a different data source, Bird *et al.* (2012) examine the relationship between SAM ESG ratings and the stock return. However, their study indicates that CSR activities do not affect the stock return of US companies.

ESG–FP studies are also interested in ESG's effect on the cost of equity, as an *ex-ante* market-based FP measure. Such studies generally apply MSCI KLD ratings as a proxy for ESG performance. Examples of such studies include Cajias *et al.* (2014), Chava (2014), and Ng and Rezaee (2015). Cajias *et al.* (2014), and Ng and Rezaee (2015) investigate the effect of different components of MSCI KLD ratings on the cost of equity of US companies. Regression analyses of the two studies present a similar result: different components of MSCI KLD ESG ratings have different effects on the cost of equity. A more specific analysis is presented by Chava (2014) who examines

the relationship between environmental performance and cost of equity of US companies. He indicates that investors are likely to expect a higher return for companies with lower environmental performance and vice versa.

Studies of ESG–FP relations also focus on measures of FP provided by a third party. For example, Singal (2014) examines the link between a firm's credit rating, as a proxy of their FP, and their MSCI KLD rating, as a proxy for a firm's environmental performance. The study finds that superior environmental performance is linked to a higher future credit rating. Ferrero-Ferrero *et al.* (2016) apply a different strategy by using financial and ESG performance measures produced by one agency, ASSET4. They find a positive association between ASSET4's ESG scores and economic scores (as a proxy of FP).

Many researchers have applied multiple measures of FP by combining accounting- and market-based financial measures. For example, Makni *et al.* (2009) examine the causal relationship between ESG performance, proxied by ratings produced by Michael Jantzi Research Associates (MJRA)³, and FPs, proxied by return on assets (ROA), return on equity (ROE), and stock market return. The study finds no significant relationship between the aggregate ESG score and ROA or ROE. However, the ESG scores were found to affect next year's return negatively. Additional analysis in the study revealed that the environmental score has a significant negative effect on both ROA and market return. A similar combination of accounting-based and market-based FP measures is applied by Nollet *et al.* (2016) in their examination of the linear and non-linear relationship between Bloomberg ESG ratings and firms' ROA, return on capital (ROC), and excess return. Their study finds non-significant positive relationships in the linear model. However, the quadratic model of the study reveals the negative effect of ESG ratings on ROA and ROC. The study also suggests a negative but non-significant relationship between ESG and excess returns. A different result is presented by Khan *et al.* (2016) in their examination of MSCI KLD ESG ratings and FP. Their study found that the overall ESG scores of MSCI KLD had no significant effect on return on sales (ROS), which is another proxy for accounting-based FP measures. Further analysis using a continuous form of ESG rating found that overall and material⁴ ESG ratings have a positive but non-significant effect on total return.

In summary, past studies reveal diversity in ESG measures as well as in the findings regarding the ESG–FP relationship. MSCI KLD ratings appear to be a popular proxy for the ESG variable in ESG–FP studies, particularly those involving market-based FP variables (Cao *et al.*, 2023; Chava 2014; Khan *et al.*, 2016). Most of these studies arrive at a similar conclusion, namely that different aspects of MSCI KLD data affect FP differently. However, the studies disagree on the significance of specific aspects; the nature of MSCI KLD data likely causes this disagreement. The data is presented as raw binary data of indicators, and researchers might implement different methods to produce composite ESG scores for their analysis (Mattingly 2017). The literature also

³ MJRA developed to become Sustainalytics.

⁴ Materiality is defined and mapped according to the Sustainability Accounting Standards Board (SASB) guideline.

shows that, in a similar context, studies using ESG ratings from other rating agencies such as SAM (Bird *et al.*, 2012), MJRA (Makni *et al.*, 2009), and Bloomberg (Nollet *et al.*, 2016) tend to have different conclusions to the studies using MSCI KLD ratings. This difference indicates that the use of different ESG ratings could be associated with the diversity of findings in ESG–FP studies.

Research Methodology

ESG Ratings

This study investigates whether the selection of one ESG rating over another affects the results of ESG–FP studies. The ESG ratings examined include the composite ESG score and scores for each dimension: environmental (Env), social (Soc), and governance (Gov). Analyses at these different levels are essential, given that each dimension potentially has a different effect on FP (Cajias *et al.*, 2014; Cao *et al.*, 2023; Dorfleitner *et al.*, 2015).

This study employs ESG score as the continuous form of ESG rating that is generally used in ESG related studies. The four ESG ratings included in the study are from Sustainalytics, Bloomberg, Thomson Reuters, and MSCI KLD. The analysis was performed for the period from 2009 to 2013. Data availability was the primary consideration for the selection of the analysis period. Each rating agency provides data for a different period, and this study only includes the period where data from all four rating agencies are available to the researcher.

Sustainalytics provides Env, Soc, Gov, and ESG scores, which range from 1 to 100⁵. Sustainalytics updates the ESG scores several times in a year to include significant changes in companies' ESG performances. Sustainalytics's scores analyzed in this study are the Env, Soc, Gov, and ESG scores as of December of each year.

Bloomberg aggregates indicators score to create annual ESG disclosure scores and scores for each dimension (Env, Soc, and Gov). The scores are normalized and range from zero (for companies that do not provide any disclosure) to 100 (for companies that disclose every indicator applicable to the company). Bloomberg claims that the data might not fully represent ESG performance. However, the data have been used as a proxy for ESG performance in many studies (Halbritter and Dorfleitner, 2015). Consequently, the ESG disclosure scores are considered equivalent to other ESG ratings for this study.

Thomson Reuters⁶ publishes annual Thomson Reuters ESG scores as a replacement of ASSET4. The ESG scores consist of an ESG composite score and scores for ten categories (resource use, emissions, innovation, workforce, human rights, community, product responsibility, management, shareholders, and CSR strategy),

⁵ Sustainalytics launched ESG risk ratings in September 2018 as the new generation of its rating (<https://www.sustainalytics.com/esg-investing-news/sustainalytics-launches-esg-risk-ratings/>). This study applied the previous version of Sustainalytics ESG scores.

⁶ Thomson Reuters' Financial & Risk unit operate as Refinitiv since July 2018.

which range from zero to 100. As of April 2017, Thomson Reuters does not provide scores of each dimension⁷. The Env, Soc, and Gov scores for this study are calculated using a weighted-average method based on the proportion of data points in each category compared to the total data points for the dimension. Thomson Reuters applied a similar method in computing the ESG composite score.

MSCI KLD ESG ratings consist of two variables—"strength" and "concerns"—and seven categories of indicators (environmental, community, human rights, employee relations, diversity, products, and governance). MSCI KLD differs from other rating agencies in this study because it does not present composite data. Instead, MSCI KLD publishes binary scores (one or zero) for all indicators. This study applies one of the most common methods to produce the composite ESG, Env, Soc, and Gov scores of MSCI KLD by deducting the sum of concerns variable from the sum of strength variable (Chatterji, Levine, and Toffel 2009). Subsequently, this study follows Halbritter and Dorfleitner (2015) by calculating a standardized score that ranges from zero to 100 based on the lowest and highest sum.

Empirical Framework

The main interest of the empirical investigation in this study is the difference in the specification of the ESG–FP association due to different proxies of ESG variable. ESG ratings from different rating agencies are significantly different (Berg *et al.*, 2022) as they have different distribution and risk profiles (Dorfleitner *et al.* 2015), are based on different assessment methods (Hedesström *et al.*, 2011), show low correlation (Chatterji *et al.*, 2016) and present low level of agreement (Widyawati 2021). These significant differences among ESG ratings could affect the ESG–FP relationship.

To answer the question of whether the use of different ESG ratings leads to a different degree of significance of ESG variables, this study follows Halbritter and Dorfleitner (2015) by running a lag regression model repetitively using different ESG ratings as a proxy of the ESG variable. For each year of t , the FP is estimated by the following regression model:

$$FP_{i,t} = b_0 + b_1 ESG_{i,t-1} + b_3 X_{i,t-1} + IndustryFixedEffect + YearFixedEffect + \varepsilon_{i,t}$$

where $FP_{i,t}$ is the FP measures at year t , $ESG_{i,t-1}$ represents the ESG rating at year $t-1$, and $X_{i,t-1}$ represents a vector of control variables at year $t-1$.

This study employs both accounting- and market-based measures of FPs as the dependent variables. A review by López-Arceiz *et al.* (2018) suggests that accounting-based measures, including ROA, ROE, ROS, and net profit, are the most frequently used measures for ESG–FP studies. Consistent with previous studies (Khan *et al.*,

⁷ Thomson Reuters (Refinitiv) provides ESG dimensional score starting from 2019 (based on most recently published publication regarding the score's methodology which is accessed on 11 January 2020 at https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/esg-scores-methodology.pdf).

2016; Li *et al.*, 2018; Nollet *et al.*, 2016), this study uses ROA and ROS to represent the book-value-based FP.

For the market-based measures, the analysis focuses on returns. Returns are commonly used to represent market-based FP (López-Arceiz *et al.*, 2018). This study applies annualized total returns and excess returns. Annual total returns are used by Makni *et al.* (2009), while the use of annual excess returns is based on studies by Halbritter and Dorfleitner (2015) and Nollet *et al.* (2016). The risk-free rate applied to compute the annual excess returns is the yield of US Treasury Bonds.

Several factors that are confirmed to affect a firm's FP are used in this study as control variables. Consistent with previous studies (Khan *et al.*, 2016; Li *et al.*, 2018; Ng and Rezaee 2015; Nollet *et al.*, 2016), the analysis includes size, leverage, capital expenditure, research and development, liquidity, book-to-market ratio, the firm's beta, and the firm's previous returns as control variables. As the data are cross-sectional over the different industries, the analysis also includes the year and industry fixed effects.

This study analyzes an overlapping sample to minimize sample-specific effects. The overlapping sample includes companies that are rated by the four rating agencies and was manually identified. The number of companies included in the overlapping sample is different for each year, as shown in Figure 1.

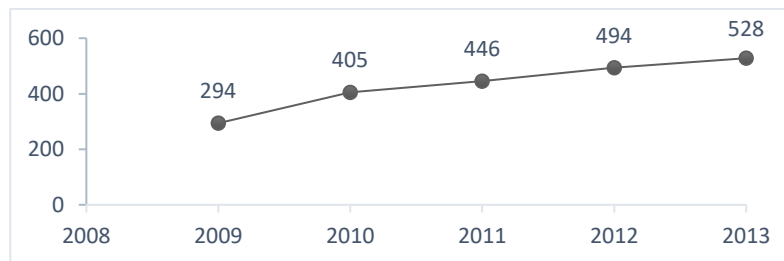


Figure 1. The number of overlapping samples

Source: Data Processed (2020)

Results And Discussion

Descriptive Statistics

Table 1 presents the univariate descriptive statistics for all variables. The financial variables are winsorized at the 1% and 99% levels to control the effect of outliers. The outliers are identified using boxplot charts. The dispersion between the minimum and maximum scores as well as the standard deviation of all variables suggests that there is sufficient variation in the data to derive meaningful analysis.

Table 1. Descriptive Statistics of All Variables

Variable	M	Minimum	Maximum	SD
ROA	.0593	-.12	.24	.0586
ROS	.0992	-.20	.40	.0939
Excess Return	.0010	-.18	.16	.0558
Return	.0222	-.17	.22	.0609
Size	9.3428	7.03	12.30	1.1676
Leverage	.2523	.00	.67	.1590
Capex	.1100	.00	1.18	.2061
R&D	.0294	.00	.38	.0650
Liquidity	2.7941	.56	12.51	1.8882
BTM	.5179	-.03	1.78	.3472
Beta	1.1549	.41	2.30	.4166
Past Return	.0317	-.14	.23	.0619
ESG				
Bloomberg	30.6975	12.00	76.35	13.5481
Thomson Reuters	60.1010	7.93	98.02	16.4442
MSCI KLD	40.5580	.00	100.00	16.3689
Sustainalytics	57.0500	33.00	86.00	8.6216
Env				
Bloomberg	21.7152	.78	82.17	17.6880
Thomson Reuters	60.3188	5.78	97.16	20.9523
MSCI KLD	51.9936	.00	100.00	15.6225
Sustainalytics	53.5720	23.00	96.00	12.7419
Soc				
Bloomberg	24.8352	3.00	86.67	16.2784
Thomson Reuters	60.5421	5.02	98.43	18.6608
MSCI KLD	41.8664	.00	100.00	16.8019
Sustainalytics	55.9810	21.00	90.00	10.0513
Gov				
Bloomberg	57.0282	37.50	85.71	6.5251
Thomson Reuters	58.3132	2.54	98.79	20.8117
MSCI KLD	59.9214	.00	100.00	13.9774
Sustainalytics	64.3600	34.00	95.00	9.3964

Note. $n = 2167$. ROA = Net income divided by average total assets; ROS = Net income over annual total sales; Return = Annualized total monthly stock returns; Excess return = Annualized monthly excess returns; Size = Natural log of market capitalization at the end of the year; Leverage = Total debt divided by total assets; Capex = Capital expenditure divided by total sales; R&D = Research and development expenditure divided by total sales; Liquidity = Common shares traded divided by total shares outstanding; BTM = Shareholders' equity divided by market capitalization; Beta = Calculated using the market model based on daily returns for a minimum of 180 days and a maximum of 730 days; Past Return = Annualized total returns over the previous 12 months.

Source: Data Processed (2020)

The descriptive statistics also show a different dispersion among the four ESG ratings. MSCI KLD ratings exhibit the highest variation of ESG ratings, as they consistently show the broadest range between minimum and maximum scores. This high variation is potentially a result of the binary nature of MSCI KLD data. Conversely, Thomson

Reuters ESG scores consistently exhibit the highest variation in the spread of the data as demonstrated by the highest standard deviation for all four ESG scores. This spread is likely related to the fact that Thomson Reuters has the highest number of indicators applicable to each rated company among the four rating agencies (Widyawati 2021), which allows for widespread variability.

The mean of the ESG scores also displays different trends among the ESG ratings. Bloomberg ESG scores consistently have the lowest mean for all levels of measurement. The low mean could be an indication of the more conservative approach taken by the rating agency in its assessment process compared to other agencies. However, the mean of Bloomberg's Gov score is more than two times the mean of its Env and Soc scores. This indicates the prominence of the Gov dimension in the Bloomberg database. In contrast, a more balanced approach is demonstrated by Thomson Reuters, which has similar mean values for all scores. The different variations and means of ESG ratings potentially reflect the different nature and measurement frameworks of the ratings.

Analysis of the correlation of the variables in both accounting-based and market-based FP models (unreported) demonstrates that each ESG rating has a different correlation with the FP measures, which is an indication that each rating affects FP differently. Examination of variance inflation factor (VIF) is also performed for all variables in all models to detect possible multicollinearity issues. Results of the VIF examination (unreported) show that no VIF exceeds three for any of the variables, which indicates that multicollinearity is not likely to be an issue.

Accounting-Based FP Measures

Panels A and B of Table 2 present results of the effect of the ESG variable on ROA and ROS, as proxies for accounting-based FP measures. All four ESG ratings indicate that composite ESG score negatively affects ROA and ROS. However, only Bloomberg and Thomson Reuters ESG ratings demonstrate statistical significance for both ROA and ROS models. The results of Bloomberg's ESG rating in the ROA model and MSCI KLD's ESG rating in the ROS model are aligned with the results of previous studies that have applied the two ratings (Khan *et al.*, 2016; Nollet *et al.*, 2016). The negative effect of overall ESG score points to the costs of implementing ESG processes and maintaining ESG performance.

Results of the analysis of the association between scores of each ESG dimension and accounting-based FP measures are also shown in Panels A and B of Table 2. The results suggest that each dimension of ESG has a different effect on ROA and ROS. Env scores consistently exhibit a significant negative relationship with FP measures, except for MSCI KLD's ratings in the ROA model. This result is aligned with the previous study by Makni *et al.* (2009), who argues that environmental initiatives, which often involve long-term projects, are often too costly and lead to poor FP in the short term, even though it is expected to improve long-term FP.

Nevertheless, the empirical analysis shows that the extent of the negative impact varies depending on the ESG rating applied. This variation occurs in a pattern that can be observed in both the ROA and ROS models. In both models, Bloomberg's and

Thomson Reuters' ratings tend to achieve a similar statistical significance and coefficient value for the Env score. Extreme values in both models are consistently observed in MSCI KLD's and Sustainalytics' ratings: MSCI KLD constantly presents the weakest effect while Sustainalytics always presents the strongest effect. The ROA and ROS models show a comparable disparity between extreme values with 57 points difference for the ROA model and 61 points difference for the ROS model.

The impact of Soc and Gov scores on both ROA and ROS is considerably more heterogeneous compared to the impact of the Env score. An examination of the impact of the Soc score on ROA and ROS shows different statistical significances and directionality as both significant positive and negative impacts are presented. For example, in the ROS model, Thomson Reuters' ratings indicate that the Soc score has a significant negative impact. In contrast, Sustainalytics's ratings suggest that the Soc score has a significant positive impact. A less extreme disparity in statistical significance is observed in the Gov score. Only the Gov score of MSCI KLD consistently presents a significant positive effect for both ROA and ROS.

A similar variation is observed in regard to the magnitude of the effect. The highest discrepancy between weakest and strongest effect is observed in the Soc scores of the ROS model. Two ratings consistently show the strongest effect in the ROA and ROS models. Sustainalytics's ratings consistently have the strongest effect in the Soc dimension, while MSCI KLD's ratings have the strongest effect in the Gov dimension.

Panels A and B of Table 2 also present the association between ESG rankings and proxies of accounting-based FP measures. In general, ESG ranking generates a weaker effect compared to the ESG score, as indicated by the lower value of the coefficient. However, the use of ESG ranking improves the convergence of the association between ESG and accounting-based FP. The use of composite ESG ranking instead of score reduces variability in both statistical significance and magnitude of the effect. This reduction is most notable in the ROA model: use of ESG ranking in the ROA model leads to a full agreement among ESG ratings, as all ratings show a significant negative regression coefficient. Correspondingly, the magnitude variation is smaller. MSCI KLD's and Sustainalytics' ratings also exhibit notable changes. The two ratings show a significant effect of ESG ranking compared to the non-significant effect of ESG score.

Table 2. Coefficient of the ESG variables in Different ESG–FP Models

	Bloomberg		Thomson Reuters		MSCI KLD		Sustainalytics	
	(I)	(II)	(I)	(II)	(I)	(II)	(I)	(II)
PANEL A: Accounting-Based FP Measures—ROA								
ESG	-.00030*** (-3.29)		-.00045*** (-5.73)		-.00003 (-.36)		-.00021 (-1.39)	
Env		-.00032*** (-3.13)		-.00039*** (-5.39)		-.00009 (-1.09)		-.00066*** (-5.57)
Soc		.00010 (-3.13)		.00002 (.34)		-.00009 (-1.20)		.00054*** (3.82)
Gov		-.00003 (.98)		-.00006 (-1.06)		.00024*** (2.99)		.00007 (.42)
PANEL B: Accounting-Based FP Measures—ROS								
ESG	-.00031** (-1.98)		-.00083*** (-6.33)		-.00006 (-.45)		-.00040 (-1.60)	
Env		-.00035** (-2.05)		-.00040*** (-3.29)		-.00025* (-1.82)		-.00086*** (-4.30)
Soc		.00022 (1.22)		-.00024* (-1.71)		-.00008 (-.652)		.00082*** (3.461)
Gov		-.00022 (-.52)		-.00015 (-1.53)		.00038*** (2.79)		-.00029 (-1.10)
PANEL C: Market-Based FP Measures—Total Return								
ESG	-.00003 (-.34)		-.00005 (-.57)		.00011 (1.39)		.00017 (1.06)	
Env		-.00006 (-.52)		-.00012 (-1.47)		-.00001 (-.16)		-.00003 (-.24)
Soc		-.00007 (-.62)		.00017* (1.81)		.00017** (2.04)		.00013 (.85)
Gov		.00036 (1.32)		-.00007 (-1.06)		-.00009 (-.992)		.00011 (.66)
PANEL D: Market-Based FP Measures—Excess Return								
ESG	-.00004 (-.35)		-.00005 (-.63)		.00006 (.73)		.00009 (.58)	
Env		-.00005 (-.47)		-.00016** (-1.96)		-.00006 (-.61)		-.00013 (-1.03)
Soc		-.00005 (-.39)		.00018* (1.92)		.00011 (1.35)		.00016 (1.05)
Gov		.00025 (.95)		-.00004 (-.64)		-.00003 (-.38)		.00014 (.80)

Note. ESG, Env, Soc, and Gov are absolute scores. Each panel presents a different set of models, based on the dependent variable applied. There are no significant differences regarding the coefficients of intercept and control variables within the same model despite using different ESG proxy. Model (I) considers overall ESG score while model (II) examines the effect of the particular ESG dimension. Analysis of all models is performed using a linear mixed model to account for autocorrelation and includes year and industry fixed effects with $n = 2167$. ***, **, and * indicate significance levels of 1%, 5%, and 10% respectively.

Source: Data Processed (2020)

Market-Based FP Measures

The estimates of ESG variables in the market-based FP models are displayed in Panels C and Panel D of Table 2. The empirical models involving composite ESG scores reveal a consensus among the ratings that ESG does not significantly affect either total or excess returns. Results of MSCI KLD's ESG rating in both return models, particularly the direction and the non-significant effect, are consistent with Halbritter and Dorfleitner (2015) and Khan *et al.* (2016). Similarly, the results of Bloomberg's ESG scores in the excess return model are consistent with Nollet *et al.* (2016). The non-significant result supports the argument that ESG is likely to affect FP in a non-linear manner.

Differences exist in the directionality and magnitude of the effect. Bloomberg's and Thomson Reuters' ratings exhibit negative coefficients while MSCI KLD's and Sustainalytics's present positive coefficients. There is a pattern of coefficient rank with Sustainalytics consistently presenting the highest coefficient, followed by MSCI KLD, Thomson Reuters, and Bloomberg (which consistently presents the lowest coefficient). However, the differences do not affect the interpretation of the model due to the non-significant result.

Panels C and D of Table 2 also present an estimation of the effect of ESG dimensions on total returns and excess returns, respectively. The panels show that two ratings, Bloomberg's and Sustainalytics's, have a non-significant impact for all dimensions.

Regarding the Env dimension, all four ESG ratings again agree on the negative relationship between Env score and market-based FP measures. However, only Thomson Reuters's Env score has a significant negative effect on excess returns. A similar significant negative effect of the Env dimension is observed by Makni *et al.* (2009). The negative effect is potentially due to the perceived cost of environmental initiatives. There are also slight variations in magnitudes.

More variations are observed in the models for the Soc and Gov dimensions. All ratings agree on the non-significance of the Gov score on total and excess returns. In contrast, there is a different conclusion regarding the statistical significance of the Soc score. In the total returns model, Thomson Reuters and MSCI KLD agree that the Soc dimension significantly improves returns. In the excess returns model, only Thomson Reuters presents a significant positive coefficient of the Soc score.

Similar to the composite ESG score, the directionality of the effect varies as the two return models present both positive and negative coefficients of Soc and Gov scores. In both models, the Gov dimension exhibits more variation, as only two ratings agree on a direction, compared to the Soc dimension, where three ratings agree on the positive direction.

Regarding the extent of the effect of the Soc and Gov dimensions, the ratings show a relatively low discrepancy. In both return models, Thomson Reuters is consistent in presenting the strongest effect of the Soc score while Bloomberg consistently presents the highest coefficient of the Gov score.

Conclusion

The proliferation of studies about the ESG–FP relationship within the last decades is yet to be followed by a general conclusion despite significant improvements to some methodological issues. One of the main issues is related to the proxies used to represent ESG performance (Gillan *et al.*, 2021; Kong *et al.*, 2020; López-Arceiz *et al.*, 2018). Many studies use ESG ratings produced by rating agencies as a proxy (Huang 2021). However, studies about the ESG ratings demonstrate that there is a lack of standardization and transparency regarding the construction of the ratings, which results in low convergence (Chatterji *et al.*, 2016). Widyawati (2021) expanded the understanding of the divergence by presenting evidence of significant differences in the construct of the ESG measurement framework and low agreement among ESG ratings.

The low convergence of ESG ratings produced by different rating agencies indicates that different ratings may have potentially different influences on the conclusions drawn regarding the ESG–FP relationship. However, since most ESG–FP studies only use rating from one agency, there is a lack of understanding about this influence. Expanding on a previous study by Halbritter and Dorfleitner (2015), this study aimed to provide evidence of this influence by testing whether the use of different ESG rating leads to a different conclusion regarding the ESG–FP relationship.

The analysis employed four different ESG ratings on the same empirical framework. Multiple models were used to examine the ESG–FP relationship to ensure robustness. The empirical framework consisted of eight ESG–FP models considering the application of four FP measures (ROA, ROS, total returns, and excess returns) and two levels of the ESG variable (composite ESG and dimensional ESG). Linear mixed-model analysis of all models was performed repetitively using the four different ESG ratings. The base empirical framework involved an analysis of the use of ESG scores as a proxy for the ESG variable. Analysis of an alternate empirical framework was performed to test whether the application of an ordinal form of ESG rating improves agreement among ESG ratings.

Analysis of the empirical framework revealed three variations of the ESG variable: statistical significance, directionality, and magnitude. Variation of statistical significance appears as ESG ratings disagreed on whether the ESG variable has a significant effect on FP measures. Directionality variation involved differences in the direction of the effect of ESG on FP. Variation in magnitude was represented by differences in the value of the coefficient of the ESG variable.

Among the three variations, the variation of statistical significance can be considered as the most important variation. Statistical significance is the first criteria generally used by the researcher in drawing conclusions regarding a linear model. This study shows that, in certain models, the use of different ESG ratings could lead to different statistical significance, which could entail different conclusions about the ESG–FP relationship. In cases where there is an agreement in the significance of the effect, variation in directionality and magnitude then needs to be considered. For example, this paper showed that, in the ROS model, the Soc scores of Thomson Reuters and

Sustainalytics both had significant effects but in different directions. This difference in directionality could lead to a different conclusion about the role of the Soc dimension on FP. The presence of these variations suggests that ESG ratings have considerable influence on the conclusion of ESG–FP studies, particularly in studies that use the ESG score (rather than ranking).

Analysis of empirical framework of this study also uncovered several patterns regarding variations. The accounting-based models showed more variations, namely the statistical significance and magnitude variation. In contrast, there was mainly a directionality variation in market-based FP models, especially in the case of overall ESG performance. It can be concluded that accounting-based FP measures are more sensitive to differences in the ESG ratings used. This sensitivity is a concern considering that a meta-analysis by López-Arceiz *et al.* (2018) found that studies using accounting-based FP measures are more likely to conclude significant ESG–FP relationships compared to studies using market-based FP measures.

In general, the results of the analysis using composite ESG score exhibited fewer variations compared to results using dimensional ESG. For example, the composite ESG score of the four ESG ratings in market-based FP models exhibited the same non-significant result and relatively comparable values even though they presented a different directionality. This trend is aligned with Widyawati (2021), who found that the agreement of composite ESG ratings is higher than the agreement of individual ESG dimensions. These results point to the importance of information about the aggregation method applied to produce overall ESG so that users can assess whether the method is justifiable.

Among the dimensions of ESG, the Soc dimension presented the most variation, showing a substantial disparity in all three types of variation. The ratings from the four rating agencies never reached an agreement on statistical significance in all base models. At most, only two ratings presented the same significant coefficient. Likewise, at most, only two ratings agreed on the direction of the effect. The Soc dimension also presented the highest magnitude variation, as observed in the Soc score in the ROS model. The highest variations presented by the Soc dimension were consistent with the findings from Widyawati (2021) that the Soc dimension showed the greatest diversity of topics and substantial differences in measurement approaches.

The Gov dimension was second in terms of variation as it always showed directionality variation in all base models. However, all four ratings agreed on the non-significant impact of the Gov dimension on market-based FP measures. The Env dimension was the most consistent as it always showed negative and sometimes significant association with FP. The magnitude variation of the Gov and Env dimensions were also considerably less than those of the Soc dimension.

Despite the apparent variations, a pattern is observed regarding the ESG ratings. For the composite ESG scores, there are two sets of ESG ratings. The members of each set present similar directionality and statistical significance. The first set consists of ratings from Bloomberg and Thomson Reuters while the second set consists of those from MSCI KLD and Sustainalytics. The two sets frequently present opposite results.

However, the similarity between members of each set is not observed at the dimensional level, as all four ratings show different patterns of results. This suggests that the similarity of effect is not derived from the similarity of measurement frameworks. Therefore, the composite ESG scores of the set's members are not entirely interchangeable.

This methodological paper presents evidence that the significant differences in the measurement framework and measurement concerns to different specifications of the ESG–FP relationship. Therefore, researchers need to consider the differences when using ESG ratings as proxies for ESG performance. This study provides evidence on the importance to use multiple ratings as different rating might lead to a different conclusion. Therefore, future researchers should consider the differences and performed appropriate strategy to treat potential issues it may caused.

This study attempts to extend the knowledge about ESG ratings in ESG–FP studies. However, some limitations need to be considered regarding the analysis. The analysis only involved a limited number of sample companies for each rating agency. The analysis was also performed only for the period until 2013. Therefore, recent changes in the measurement framework of each rating agency or recent developments in the ESG rating industry are not yet included. It is also important to note that this study does not aim to replicate previous ESG–FP studies, as there are numerous control variables, interaction effects, and fixed effects that cannot be applied in this study due to data availability and resource limitation. Despite these limitations, this paper provides additional insight into the influence of ESG rating choice on the association between ESG and FP.

References

- Atz, Ulrich, Tracy Van Holt, Zongyuan Zoe Liu, and Christopher C. Bruno. 2023. "Does Sustainability Generate Better Financial Performance? Review, Meta-Analysis, and Propositions." *Journal of Sustainable Finance & Investment* 13(1):802–25. <https://doi.org/10.1080/20430795.2022.2106934>
- Berg, F., J. F. Kölbel, and R. Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Ratings*." *Review of Finance* 26(6):1315–44. <https://doi.org/10.1093/rof/rfac033>
- Billio, M., M. Costola, I. Hristova, C. Latino, and L. Pelizzon. 2021. "Inside the ESG Ratings: (Dis)Agreement and Performance." *Corporate Social Responsibility and Environmental Management* 28(5):1426–45. <https://doi.org/10.1002/csr.2177>
- Bird, R., F. Momenté, and F. Reggiani. 2012. "The Market Acceptance of Corporate Social Responsibility: A Comparison across Six Countries/Regions." *Australian Journal of Management* 37(2):153–68. <https://doi.org/10.1177/0312896211416136>
- Cajias, M., F. Fuerst, and S. Bienert. 2014. "Can Investing in Corporate Social Responsibility Lower a Company's Cost of Capital?" *Studies in Economics and Finance* 31(2):202–22. <https://doi.org/10.1108/SEF-05-2013-0067>
- Cao, J., S. Titman, X. Zhan, and W. Zhang. 2023. "ESG Preference, Institutional Trading, and Stock Return Patterns." *Journal of Financial and Quantitative Analysis* 58(5):1843–77. <https://doi.org/10.1017/S0022109022000916>

- Chatterji, A. K., R. Durand, D. I. Levine, and S. Touboul. 2016. "Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers." *Strategic Management Journal* 37(8):1597–1614. <https://doi.org/10.1002/smj.2407>
- Chatterji, A. K., D. I. Levine, and M. W. Toffel. 2009. "How Well Do Social Ratings Actually Measure Corporate Social Responsibility?" *Journal of Economics and Management Strategy* 18(1):125–69. doi: <https://doi.org/10.1111/j.1530-9134.2009.00210.x>
- Chava, S. 2014. "Environmental Externalities and Cost of Capital." *Management Science* 60(9):2223–47. <https://doi.org/10.1287/mnsc.2013.1863>
- Christensen, D. M., G. Serafeim, and A. Sikochi. 2022. "Why Is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings." *Accounting Review* 97(1):147–75. <https://doi.org/10.2308/TAR-2019-0506>
- Dorfleitner, G., G. Halbritter, and M. Nguyen. 2015. "Measuring the Level and Risk of Corporate Responsibility - An Empirical Comparison of Different ESG Rating Approaches." *Journal of Asset Management* 16(7):450–66. <https://doi.org/10.1057/jam.2015.31>
- Drempetic, S., C. Klein, and B. Zwergel. 2020. "The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review." *Journal of Business Ethics* 167(2):333–60. <https://doi.org/10.1007/s10551-019-04164-1>
- Ferrero-Ferrero, I., M. Fernández-Izquierdo, and M. J. Muñoz-Torres. 2016. "The Effect of Environmental, Social and Governance Consistency on Economic Results." *Sustainability (Switzerland)* 8(10). <https://doi.org/10.3390/su8101005>
- Gillan, S. L., A. Koch, and L. T. Starks. 2021. "Firms and Social Responsibility: A Review of ESG and CSR Research in Corporate Finance." *Journal of Corporate Finance* 66. <https://doi.org/10.1016/j.jcorpfin.2021.101889>
- Halbritter, G., and G. Dorfleitner. 2015. "The Wages of Social Responsibility - Where Are They? A Critical Review of ESG Investing." *Review of Financial Economics* 26:25–35. <https://doi.org/10.1016/j.rfe.2015.03.004>
- Hedesström, M., U. Lundqvist, and A. Biel. 2011. "Investigating Consistency of Judgement across Sustainability Analyst Organizations." *Sustainable Development* 19(2):119–34. <https://doi.org/10.1002/sd.511>
- Huang, D. Z. X. 2021. "Environmental, Social and Governance (ESG) Activity and Firm Performance: A Review and Consolidation." *Accounting and Finance* 61(1):335–60. <https://doi.org/10.1111/acfi.12569>
- Khan, M., G. Serafeim, and A. Yoon. 2016. "Corporate Sustainability: First Evidence on Materiality." *Accounting Review* 91(6):1697–1724. <https://doi.org/10.2308/accr-51383>
- Kong, Y., A. Antwi-Adjei, and J. Bawuah. 2020. "A Systematic Review of the Business Case for Corporate Social Responsibility and Firm Performance." *Corporate Social Responsibility and Environmental Management* 27(2):444–54. <https://doi.org/10.1002/csr.1838>
- Li, Y., M. Gong, X. Y. Zhang, and L. Koh. 2018. "The Impact of Environmental, Social, and Governance Disclosure on Firm Value: The Role of CEO Power." *British Accounting Review* 50(1):60–75. <https://doi.org/10.1016/j.bar.2017.09.007>
- López-Arceiz, F. J., A. J. Bellostas, and P. Rivera. 2018. "Twenty Years of Research on the Relationship Between Economic and Social Performance: A Meta-Analysis Approach." *Social Indicators Research* 140(2):453–84. <https://doi.org/10.1007/s11205-017-1791-1>

- Makni, R., C. Francoeur, and F. Bellavance. 2009. "Causality between Corporate Social Performance and Financial Performance: Evidence from Canadian Firms." *Journal of Business Ethics* 89(3):409–22. <https://doi.org/10.1007/s10551-008-0007-7>
- Mattingly, J. E. 2017. "Corporate Social Performance: A Review of Empirical Research Examining the Corporation–Society Relationship Using Kinder, Lydenberg, Domini Social Ratings Data." *Business and Society* 56(6):796–839. <https://doi.org/10.1177/0007650315585761>
- Ng, A. C., and Z. Rezaee. 2015. "Business Sustainability Performance and Cost of Equity Capital." *Journal of Corporate Finance* 34:128–49. <https://doi.org/10.1016/j.jcorpfin.2015.08.003>
- Nollet, J., G. Filis, and E. Mitrokostas. 2016. "Corporate Social Responsibility and Financial Performance: A Non-Linear and Disaggregated Approach." *Economic Modelling* 52:400–407. <https://doi.org/10.1016/j.econmod.2015.09.019>
- Singal, M. 2014. "The Link between Firm Financial Performance and Investment in Sustainability Initiatives." *Cornell Hospitality Quarterly* 55(1):19–30. <https://doi.org/10.1177/1938965513505700>
- Widyawati, L. 2021. "Measurement Concerns and Agreement of Environmental Social Governance Ratings." *Accounting and Finance* 61(S1):1589–1623. <https://doi.org/10.1111/acfi.12638>
- Wong, C., A. Brackley, and E. Petroy. 2019. "Rate the Raters 2019: Expert Views on ESG Ratings." <https://Sustainability.Com/Wp-Content/Uploads/2019/02/SA-RateTheRaters-2019-1.Pdf>.