Journal of Risk **22**(6), 1–33
DOI: 10.21314/JOR.2020.437

Research Paper

The impact of corporate social and environmental performance on credit rating prediction: North America versus Europe

Gregor Dorfleitner,[1] Johannes Grebler[1] and Sebastian Utz[2]

[1]Department of Finance, University of Regensburg, 93040 Regensburg, Germany;
emails: gregor.dorfleitner@ur.de, j-grebler@web.de
[2]School of Finance, University of St. Gallen, 9000 St. Gallen, Switzerland;
email: sebastian.utz@unisg.ch

(Received April 25, 2019; revised November 12, 2019; accepted February 18, 2020)

ABSTRACT

We quantify the extent to which the quality of credit rating predictions improves by integrating measures of corporate social performance (CSP) in an established credit risk model. Our analysis provides comprehensive evidence of the comparative informational advantage of considering CSP in predicting credit ratings of North American and European firms. In the North American sample, both environmental and social performance have an explanatory impact. The out-of-sample prediction quality improves by more than 0.8%. By contrast, only social performance increases the explanatory power in the European sample; environmental performance does not. Overall, we show that CSP is a relevant variable for predicting credit ratings. In general, our findings support the risk mitigation view of CSP, indicating that firms with high CSP are less risky and thus have better credit ratings. However, the quality of the relationship depends on the socioeconomic and cultural environment as well, as can be seen from the differing results in North America and Europe.

Keywords: credit risk; credit rating prediction; corporate social performance; risk mitigation.

Corresponding author: G. Dorfleitner

Print ISSN 1465-1211 | Online ISSN 1755-2842

1 INTRODUCTION

This paper analyzes whether, how and the extent to which the prediction quality of a firm's credit rating improves by integrating its corporate social performance (CSP) into the forecasting model. To capture regional differences, we analyze two samples of firm-level data including North American firms and one including a European firm. We unify the work of Jiraporn *et al* (2014) for North America and Stellner *et al* (2015) for Europe by providing a framework in which results for the two regions can credibly be compared in an established credit risk model. We use CSP measures of the globally available Asset4 framework and investigate the impact of these CSP measures on both the explanation and the prediction of credit ratings. We capture region-specific differences by estimating models for North America and Europe separately in a first step and for a merged data set in a second step. In particular, we apply a two-stage approach with an estimation of credit risk models, including CSP variables in the first stage and an out-of-sample analysis using the estimate of the first stage to predict the credit ratings in a second stage. Finally, we measure the prediction quality by comparing predicted and actual credit ratings. In North America, both the environmental and social CSP show an explanatory impact on credit ratings, while only the social CSP is relevant in Europe. Further, we find the prediction power of the credit risk model improved when using CSP scores for the North American sample, while we document no improvement in the European sample.

In theory, corporate social responsibility – which we repeatedly address in this paper through the narrower yet measurable concept of CSP – can coexist with both better and worse credit ratings, provided there is evidence of an impact. According to Goss and Roberts (2011), there are two contrary views for the impact of CSP: the risk-mitigation view and the overinvestment view. According to the risk-mitigation view, a firm with high CSP faces lower risks than a firm with a low CSP if all other aspects of these firms are comparable. High CSP protects firms from legal, reputational and regulatory risks (Bauer and Hann 2010), allows firms to hire better qualified employees (Turban and Greening 1997), and lowers agency risks (Oikonomou *et al* 2014). The opposite (overinvestment) view regards investments in CSP as a waste of scarce resources. An increase in fixed costs related to sustainable investments in CSP increases the volatility of earnings and thus the default risk (Frooman *et al* 2008). Except for the environmental score in Europe, credit ratings are significantly positively correlated with the CSP scores, and thus our findings are consistent with the risk-mitigation view.

Whether CSP adds informational power to the explanation of credit ratings has been the subject of two studies, namely Jiraporn *et al* (2014), analyzing a North American sample of firms, and Stellner *et al* (2015), analyzing a European one. The evidence these studies provide is inconsistent, which could be due to either the different methodological designs or the differing regional focus of their samples. More precisely, Stellner *et al* (2015) show that CSP has no impact on the credit ratings of European firms, while Jiraporn *et al* (2014) conclude that CSP does have an impact on the credit ratings of North American firms. Nevertheless, these results are inappropriate for concluding that regional differences exist, since both studies used different model specifications and concepts for measuring CSP (data providers Asset4 and Kinder Lydenberg Domini (KLD), respectively). In particular, Asset4 and KLD CSP data shows major differences even after adjustment for different CSP definitions (Chatterji *et al* 2016; Dorfleitner *et al* 2015). Asset4 provides a comprehensive calculation of the scores based on more than 750 indicators, which are ordinal or metric, unlike KLD, which only uses a binary rating system to reflect CSP strengths and concerns for US firms (Humphrey *et al* 2012). We use the CSP measures of Asset4 due to their global coverage, which allows for a consistent estimation of a well-established credit risk model for North American and European firms. The major limitation of the existing studies in the CSP–credit rating context concerns the retrospective contemplation by measuring the correlation of credit ratings and lagged CSP, which lacks out-of-sample predictions. In our study, we predict the next period's credit rating based on all available information at a certain point in time. Finally, the prediction quality is determined by comparing actual and predicted credit ratings.

Our data set includes Standard & Poor's (S&P) counterparty ratings matched with Asset4 CSP scores and a set of control variables. We estimate several versions of an established ordered probit credit risk model, which can handle rating migrations. To be specific, we estimate one baseline model without CSP factors and three CSP models for each sample (ie, the North American sample, the European sample and the merged sample comprising both), resulting in a total of twelve model specifications. The three CSP model specifications comprise one aggregated CSP measure model, a model specification with a score for the environmental dimension of CSP and a model specification with a score for the social dimension of CSP. The in-sample period for determining the models' coefficients ranges from 2003 to 2013 for credit ratings. Subsequently, we predict credit ratings on the data set covering the years 2014–17. This two-step process ensures that only the available level of information is used to predict the following periods' credit ratings.

We find that the integration of all of our measures for CSP in the credit risk model increases the explanatory power in the North American sample. The quality of out-of-sample credit rating predictions is improved by 0.8%. For European firms, only the social dimension of CSP shows a significant (positive) correlation with credit ratings. The prediction quality experiences no improvement. Distinct findings for North America and Europe result from the geographical, social and political environments of the two regions, which are reflected in the Asset4 scores. The average level of CSP scores of North American firms is lower, and their variance is higher, than that of European firms. This pattern is one possible reason for CSP scores having a higher explanatory power in predicting credit ratings in North America, since the explanatory variables show a certain degree of variance there. In a nutshell, CSP has an impact on credit ratings in both regions, although to a different extent. For North America, our findings are consistent with those of Jiraporn *et al* (2014), while our results suggest contrasting implications to those of Stellner *et al* (2015).

Our findings reveal valuable insights for researchers, debt holders and debt issuers. Based on our approach of incorporating CSP into credit risk models, we find that debt holders experience greater accuracy in their credit rating predictions if they include CSP factors as explanatory variables in their credit risk models. With this higher prediction power, they profit twofold by preventing misjudgments: in the case of overestimation of risk, they could lose business due to excessive price-setting, while in the case of underestimation of risk, the applied pricing might not cover the anticipated risk. The latter leads to immediate losses in the risk-adjusted performance measurement and material losses when risks become imminent. Finally, debt issuers can improve their credit rating prediction, and hence their cost of debt, by increasing their CSP. For instance, an increase in the aggregated CSP score by one standard deviation for a BBB-rated North American firm results in average savings of 14.5 basis points (bps).

The remainder of this paper is organized as follows. We review the related literature and discuss theoretical concepts in Section 2. Section 3 describes the data set, and Section 4 introduces the methodology employed. Section 5 presents the empirical results, Section 6 discusses the findings and Section 7 gives an overview of the robustness tests. Finally, Section 8 states our conclusions.

2 THEORETICAL CONSIDERATIONS

Our study considers two streams of the literature: the impact of CSP on credit ratings and the regional differences regarding the attitude of firms toward CSP.

2.1 The impact of CSP on credit ratings

From a theoretical perspective, there is an indirect link between CSP and credit ratings in the context of financing cost and corporate financial performance (CFP).

First, previous studies (Dhaliwal *et al* 2011; El Ghoul *et al* 2011; Goss and Roberts 2011) suggest that firms with high CSP have lower financing costs in terms of both cost of equity and cost of debt. As creditworthiness is negatively related to interest rates payable on debt (Kisgen 2006), we expect to observe a positive relationship between CSP and credit ratings based on this consideration. From this perspective, CSP can be seen as an underlying factor, having an impact on both financing cost and credit ratings. Concerning single CSP pillars, there exists a negative relationship between CSP and financing cost for environmental performance (Schneider 2011; Sharfman and Fernando 2008) and for social performance (Chen *et al* 2011).

Second, a similar argument considers CFP as opposed to financing costs. CSP is positively related to CFP (Dorfleitner *et al* 2018; Kang *et al* 2016; von Arx and Ziegler 2014) in the sense of sustainable future cashflows. Further, CFP is positively related to creditworthiness (Standard & Poor's 2013). Finally, firms with a high CSP tend to have a lower idiosyncratic risk due to the risk-mitigation effect of CSP, which corresponds to both lower financing costs and a higher CFP (Orlitzky 2008). Therefore, we expect a positive relationship between CSP and creditworthiness.

By reexamining the different pillars of CSP, we can expose the underlying mechanisms. Firms with a low level of environmental performance face legal, reputational and regulatory risks (Bauer and Hann 2010). Moreover, a good social performance allows firms to hire better qualified employees, who are a key factor in future success (Turban and Greening 1997). It should be noted that a contrasting view (the overinvestment view) exists, according to which CSP lowers CFP when costs exceed additional positive returns (Aupperle *et al* 1985; Brammer and Millington 2008; Cornell and Shapiro 1987). However, there is less supporting evidence for this view.

From an empirical perspective, a few studies examine the impact of CSP on credit ratings by approaching an ordered-response credit risk model and show that CSP is positively related to (good) credit ratings. Stellner *et al* (2015) find no significant relationship between CSP and credit ratings in the eurozone based on the Asset4 equal-weighted rating score. However, high (low) CSP results in better credit ratings if the country's sustainability performance is high (low). Jiraporn *et al* (2014) use the KLD composite score and find that the CSP policies of US firms are affected by those of other firms in the same three-digit zip code area. Firms with high CSP have better credit ratings. A deeper look at single dimensions of CSP by utilizing KLD data shows that US firms with high environmental and social performance have better credit ratings (Attig *et al* 2013; Bauer and Hann 2010; Bauer *et al* 2009; Oikonomou *et al* 2014).

Although there exists empirical evidence on the general CSP–credit rating link, it is still not clear whether CSP has an impact on prediction quality. In addition, there is no consistent evidence on the question of whether the impact of CSP on credit ratings depends on regional differences, a matter we will treat in the next subsection.

2.2 CSP in North America and Europe

In general, CSP is higher in Europe than North America. This is true at least for the United States, as the dominant country in our North American sample. Various explanations for the differences between both regions include the legal origin (common law in North America versus civil law in Europe), the divergent institutional and political setups, the level of economic development, the historic tendency toward liberal democracy and the perception of stakeholders (Cai *et al* 2016; Doh and Guay 2006; Liang and Renneboog 2017; Maignan 2001; Welford 2005).

In particular, the stakeholder perception is linked to the differing ideologies as defined by Lodge (1990). European countries are more closely tied to a communitarian ideology, which means that they tend to pursue the goal of common, long-term goods. Conversely, the United States tends to adopt an individualistic ideology, implying that individual, short-term improvements are pursued instead. The motivation of the companies to act in a socially responsible way differs between the two regions, depending on firm size and financial performance (Sotorrío and Sánchez 2008).

CSP in the United States is more ingrained in society, while CSP in Europe appears to be more state-oriented. Historically, CSP in the United States has been driven by concrete corporate policies and programs that contribute to social concerns, while in Europe the contribution to social concerns predominantly occurs in the context of values, norms and rules. According to Matten and Moon (2008), the rise of CSP in Europe in recent decades has been the result of incentives for corporate engagement provided by the European Union.

Empirical evidence shows that North America's CSP only exceeds Europe's with respect to rare aspects such as business communication (Maignan and Ralston 2002); in terms of most aspects and measurement concepts, CSP is higher in Europe. We expect the regional differences in the CSP level to have an impact on credit ratings and their predictions in this study.

3 DATA

We match the S&P credit ratings of North American and European counterparties from Compustat with the ratings universe of the sustainability rating agency Asset4, provided by Thomson Reuters Datastream. Moreover, we use firm-year financial and accounting data from Datastream and Worldscope to control for well-documented influencing factors on credit risk. Financial counterparties are excluded based on the economic sector level of Thomson Reuters Business Classification. Our final data set comprises a panel of 724 North American firms (5393 firm-year observations) and

218 European firms (1712 firm-year observations). Both the North American panel and the European panel follow the region classification of Fama and French (2012).[1]

3.1 S&P credit ratings and Asset4 CSP scores

We use S&P long-term borrower credit ratings, reflecting the obligors' creditworthiness over a long time horizon (greater than one year) as the independent variable. The S&P issuer credit rating is defined as the current assessment of an obligor's overall financial capacity to serve its debt, ie, its creditworthiness. The rating grades comprise AAA, AA, A, BBB, BB, B, CCC, CC and D, where D is assigned to obligors that are overdue in either their interest or their capital payments. Credit ratings of BBB or better are often referred to as "investment" grade, while credit ratings below this threshold are often called "noninvestment" or "speculative" grade. Vazza and Kraemer (2017) give a detailed description of the rating methodology.

We capture a company's overall CSP (which we refer to as ES in the following), ie, the average of its environmental performance and social performance, following the methods of El Ghoul *et al* (2017), Ioannou and Serafeim (2012) and Luo *et al* (2015). We use the ES score and the score of each of the two pillars from the Asset4 database, ie, the environmental (ENV) score and the social (SOC) score.[2] The ENV and SOC scores measure corporate activities along environmental and social dimensions. The environmental score evaluates a firm's "impact on living and nonliving natural systems, including the air, land, and water, as well as complete ecosystems" (see Thomson Reuters 2011). For instance, this measure captures resource reduction, emission reduction and product innovation benefiting the environment. The social score measures the ability of a firm to "generate trust and loyalty with its workforce, customers and society" through investment in customer/product responsibility, community, human rights, diversity, employee training and development, health and safety, and employment quality. We calculate the ES score as the average of the ENV and SOC scores, respectively. The ES score represents the aggregated performance of a firm according to the environmental and social dimensions in a particular year.

Asset4 publishes scores that act as external measures for sustainable business models (Chatterji *et al* 2016; Humphrey *et al* 2012; Ioannou and Serafeim 2012). These scores are based on publicly available and traceable information, eg, websites; United States Securities and Exchange Commission (SEC) filings such as forms

---

[1] The North American panel comprises the United States and Canada, while the European panel comprises Austria, Belgium, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom, according to the FTSE Country Segment classification.

[2] Some studies also consider a third score, ie, the corporate governance score. However, we follow the definition of the studies mentioned above to determine our measures for CSP.

TABLE 1 Definitions of the subvariables.

	c_{it}^{A}	c_{it}^{B}	c_{it}^{C}	c_{it}^{D}
If $C_{it} \in [0, 5)$	C_{it}	0	0	0
If $C_{it} \in [5, 10)$	5	$C_{it} - 5$	0	0
If $C_{it} \in [10, 20)$	5	5	$C_{it} - 10$	0
If $C_{it} \in [20, 100]$	5	5	10	$C_{it} - 20$

10-K, DEF 14A and 10-Q; sustainability reports; media sources; and nongovernmental organization reports. To guarantee a high level of ratings integrity, every data point is cross-checked by at least one additional analyst and by further analyses through statistical tools. Therefore, using the Asset4 scores eliminates, as far as possible, weaknesses such as the lack of transparency in the KLD, FTSE4Good and Dow Jones rating approaches (Chatterji and Levine 2006). Accordingly, Asset4 evaluates more than 750 individual data points. Every data point matches a single question concerning the fulfillment of a specific item according to environmental, social, economic and governance issues. The information from the answers is aggregated in several stages to indicators, to pillars and finally to the average CSP rating. The scores are updated on an annual basis and range from 0 to 100, with a higher score indicating a higher level of CSP. The rating universe of Asset4 even includes a firm after a bankruptcy, a merger or another cause of delisting. Thus, the data set is free from survivorship bias.

3.2 Control variables

To capture the well-documented effects of predicting credit ratings, we control for several variables. We provide a detailed description of all of these variables in Table A.7 in the online appendix.

Following Standard & Poor's (2013) and Merton (1974), we include three-year averages of the operating margin, the long-term debt, the total debt and the interest coverage ratios. The interest coverage ratio is transformed, as suggested by Blume *et al* (1998). Since negative values can be caused either by low interest payments or by high negative earnings, the magnitude of negative values for interest coverage is not meaningful, and therefore we set these values to zero. The distribution of the interest coverage ratio is heavily skewed, and the marginal effect of changes may be small if interest coverage is already at a high level. Accordingly, we cap the three-year average at 100. To capture the nonlinear shape of interest coverage C_{it} for a company i in year t, we apply the decomposition to four subvariables, c_{it}^{A}, c_{it}^{B}, c_{it}^{C}, c_{it}^{D}, defined in Table 1.

Further, we also include market capitalization, because bigger firms tend to have superior credit ratings (Altman *et al* 1977) and because Asset4 scores show a market cap dependence (Ioannou and Serafeim 2012). Moreover, since all claims on assets must earn the same compensation per unit of risk (Campbell *et al* 2008; Friewald *et al* 2014; Merton 1974), we also control for systematic risk (market model beta) and idiosyncratic risk. In addition, the dividend policy of a firm has an impact on credit risk (Hoberg and Prabhala 2009). As profitable firms are less likely to default, we expect a positive correlation between the market-to-book ratio and credit ratings, and thus include the market-to-book ratio (Pástor and Pietro 2003). Following DeAngelo *et al* (2006), retained earnings are a proxy of a company's life-cycle phase. Mature, stable firms generally observe better ratings (Fons 1994). Thus, we also include retained earnings as a control. In addition, Tang (2009) finds that upgraded firms have more capital expenditure than downgraded firms. Hence, we expect a positive correlation between capital expenditure and credit ratings.

Moreover, firms with a weak credit risk profile tend to have precautionary savings (Acharya *et al* 2012), which is why we also use the cash balance as a control. Rampini and Viswanathan (2013) have documented the impact of tangibility on credit risk.

Further, Bangia *et al* (2002) find evidence that S&P credit ratings change procyclically. Hence, we control for business-cycle effects as represented by the gross domestic product (GDP) growth rate. We further use year dummies to control for remaining systematic effects (Elton *et al* 2001). We follow Dimitrov *et al* (2015) in that our main models, like the standard model in the literature, come without industry fixed effects. Our underlying idea assumes the industry-specific influence factors so far considered in the other controls. However, we consider industry effects in our robustness checks in Section 7.

3.3 Descriptive statistics

We lag the ES variables and controls by one period compared with the credit ratings. The estimation set contains credit ratings covering the years from 2003 through 2013 and independent variables between 2002 and 2012. Out-of-sample predictions for credit ratings in the period 2014–17 are based on independent variables from 2013 to 2016. Table 2 presents the descriptive statistics of the credit rating variable, sorted by region (North America versus Europe) and subperiod (2003–13 versus 2014–17). Rating class BBB shows the largest number of observations in both regions.

Both mean and median ES, ENV and SOC scores are lower in North America than in Europe (see Table 3), indicating a weaker overall CSP. The standard deviation in the ES, ENV and SOC scores is higher in North America than in Europe. Thus, the

TABLE 2 Total number of firms and observations per rating class, including the partial quantity of rating upgrades and downgrades for the samples of North America and Europe.

		North America			Europe		
	Period	**Total**	**UG**	**DG**	**Total**	**UG**	**DG**
AAA	I	49	4	0	6	1	0
	II	7	3	0	0	0	0
AA	I	109	12	0	92	12	0
	II	32	2	0	9	3	0
A	I	906	61	6	443	45	1
	II	265	28	2	123	13	0
BBB	I	1600	62	46	535	35	18
	II	599	33	11	204	6	5
BB	I	850	54	63	142	8	16
	II	462	34	17	73	6	5
B	I	257	5	48	35	1	8
	II	208	11	31	38	1	3
C	I	14	0	5	5	0	2
	II	35	0	8	7	0	1
Total	I	3785	198	168	1258	102	45
	II	1608	111	69	454	29	14
# firms		724			218		

"UG" stands for "Upgrade" and "DG" stands for "Downgrade". The table shows credit ratings from 2003 to 2013 (period I) for the coefficient estimation and from 2014 to 2017 (period II) for the out-of-sample prediction. Independent variables are lagged by one year compared with the credit ratings. We use S&P long-term borrower credit ratings reflecting the obligors' creditworthiness over a long time horizon (greater than one year).

CSP shows greater variability in North America than in Europe. The measures for CSP follow left-skewed distributions. A reason for this may be the fact that companies with weak CSP ratings are less likely to provide the data required to obtain an ES rating. Hence, the proportion of weakly performing companies in the database is less than in the basic population, causing this skewness.

The most significant regional difference in descriptive statistics is that the mean firm size in the European sample is bigger than in North America. This difference can be explained by the wider availability of credit ratings of smaller firms in North America. Moreover, the macroeconomic situation measured by the GDP growth rate shows a high degree of deviation. While the level of GDP growth in the estimation period and the out-of-sample period in North America is 1.6% and 2.2%, respectively, the respective numbers in Europe are 0.4% and 0.8% lower.

4 METHODOLOGY

Based on the approach of Kaplan and Urwitz (1979), its continuation by Blume *et al* (1998) and its application in many studies (see, for example, Alp 2013; Baghai *et al* 2014; Becker and Milbourn 2011; Dimitrov *et al* 2015; Jiang *et al* 2012), we estimate a threshold model based on an unobserved linking variable y_{it}^*, which represents the creditworthiness of a firm i for a year t,

$$y_{it}^* = \boldsymbol{x}_{i,t-1}'\boldsymbol{\beta} + \varepsilon_{it}, \tag{4.1}$$

where $\boldsymbol{x}_{i,t-1}$ represents the vector of observed explanatory variables of firm i in year $t-1$, and $\boldsymbol{\beta}$ is a vector of slope coefficients. The variable R_{it} is the rating category of firm i in year t. The linking variable y_{it}^* is continuous, and its range comprises the set of real numbers. In our study, we consider seven different levels of credit ratings (ie, AAA, AA, A, BBB, BB, B and C). $R_{it} = 7$ if in year t firm i has a rating of AAA, $R_{it} = 6$ if AA, $R_{it} = 5$ if A, $R_{it} = 4$ if BBB, $R_{it} = 3$ if BB, $R_{it} = 2$ if B and $R_{it} = 1$ if C. Thus, the first stage of our estimation maps the credit ratings into a partition of the unobserved linking variable y_{it}^* as follows:

$$\text{if } y_{it}^* \in [\mu_{j-1}, \mu_j), \text{ then } R_{it} = j \text{ for } j = 1, \dots, 7, \tag{4.2}$$

where μ_j are partition points independent of time t and $\mu_0 := -\infty$ and $\mu_7 := \infty$. Thresholds are not given ex ante but instead determined in the statistical procedure of estimating the model.

Following the assumption that ε_{it} is normally and independently distributed with a mean of 0 and a variance of 1, which is ensured in the estimation procedure, we calculate the probabilities for the different rating classes (given $\boldsymbol{x}_{t-1}$) according to

$$P(R_{it} = j \mid \boldsymbol{x}_{i,t-1}) = \Phi(\mu_j - \boldsymbol{x}_{i,t-1}'\boldsymbol{\beta}) - \Phi(\mu_{j-1} - \boldsymbol{x}_{i,t-1}'\boldsymbol{\beta}), \quad j = 1, \dots, 7. \tag{4.3}$$

We use the panel structure of the data for the model estimation. Both a certain rating (ie, a realization of R_{it}) and realizations of the input variables are ascribed to each company for each year during the observation period. To represent the state of information when predictions for the following period are calculated, all influencing factors are lagged by one period. Table 4 provides an overview of the input factors, boundaries and outputs of the estimated models. We estimate models for the North American and European samples as well as their merged data set. We focus on three different specifications of the CSP model: the ES model, the ENV model and the SOC model. Each variant includes the corresponding Asset4 score, as indicated by their name plus control variables. In the merged estimation, a region dummy and an interaction term between the region and CSP are also considered. The estimation is

TABLE 3 Descriptive statistics for the Asset4 scores and control variables for the North American and European samples of explanatory variables. [Table continues on next two pages.]

(a) Asset4 variables

		North America					Europe				
	Period	**Mean**	**25%**	**Median**	**75%**	**SD**	**Mean**	**25%**	**Median**	**75%**	**SD**
ES score (%)	I	51.516	24.120	48.850	80.090	28.093	81.668	75.840	88.440	92.979	15.778
	II	57.699	32.270	60.333	84.480	27.458	85.885	84.907	91.410	93.653	13.237
ENV score (%)	I	50.538	18.820	48.070	82.850	31.098	81.390	76.817	89.805	93.287	17.912
	II	56.660	26.307	61.740	87.862	30.464	85.989	84.867	91.965	93.990	14.280
SOC score (%)	I	52.437	25.210	51.790	80.820	28.726	82.024	74.972	89.075	94.648	17.024
	II	58.673	34.505	63.615	84.347	27.642	85.722	85.177	91.155	94.138	14.363

(b) Control variables

		North America					Europe				
	Period	**Mean**	**25%**	**Median**	**75%**	**SD**	**Mean**	**25%**	**Median**	**75%**	**SD**
Interest coverage (%)	I	11.108	2.640	5.510	12.080	16.972	8.668	2.870	4.920	8.588	13.388
	II	8.648	1.800	4.290	9.592	14.894	8.160	1.820	4.285	8.325	14.108
Operating margin (%)	I	13.805	7.410	12.320	19.010	8.518	11.924	5.332	10.490	16.398	8.055
	II	13.259	6.397	11.870	18.605	8.801	10.830	4.500	8.550	15.428	8.154
Long-term debt (%)	I	39.532	24.970	37.430	51.680	19.200	39.736	26.797	38.610	51.028	16.842
	II	46.004	31.887	44.665	57.555	19.497	41.544	28.895	38.995	53.940	16.364

TABLE 3 Continued.

(b) Control variables (cont.)

		North America					Europe				
	Period	**Mean**	**25%**	**Median**	**75%**	**SD**	**Mean**	**25%**	**Median**	**75%**	**SD**
Total debt (%)	I	42.702	28.110	40.720	55.230	19.013	45.764	33.315	44.535	57.545	16.103
	II	48.518	33.778	47.400	60.540	19.518	46.829	33.733	44.515	59.460	15.818
Size (US$ billions)	I	18.141	3.407	6.993	16.525	36.779	26.484	5.466	11.166	28.862	39.369
	II	21.981	2.924	7.774	20.054	45.205	27.660	5.379	13.592	31.325	38.744
Beta	I	1.057	0.732	1.011	1.320	0.412	0.936	0.689	0.904	1.161	0.330
	II	1.143	0.850	1.089	1.394	0.412	0.978	0.784	0.966	1.145	0.284
Idiosyncratic risk (%)	I	1.677	1.116	1.500	2.038	0.743	1.477	1.063	1.325	1.758	0.564
	II	1.632	1.031	1.373	1.980	0.807	1.376	0.947	1.175	1.611	0.575
Dividend payer (%)	I	0.755	1.000	1.000	1.000	0.430	0.910	1.000	1.000	1.000	0.286
	II	0.727	0.000	1.000	1.000	0.446	0.833	1.000	1.000	1.000	0.374
Market/book (%)	I	2.768	1.430	2.180	3.450	1.961	2.342	1.260	1.960	3.050	1.425
	II	2.867	1.330	2.170	3.572	2.260	2.113	1.040	1.655	2.820	1.436
Retained earnings (%)	I	0.259	0.094	0.258	0.425	0.256	0.163	0.046	0.143	0.275	0.165
	II	0.207	0.009	0.195	0.405	0.293	0.206	0.076	0.198	0.348	0.180

TABLE 3 Continued.

(b) Control variables (cont.)

		North America					Europe				
	Period	**Mean**	**25%**	**Median**	**75%**	**SD**	**Mean**	**25%**	**Median**	**75%**	**SD**
Capital expense (%)	I	5.136	2.510	4.550	7.630	3.058	5.076	2.910	4.680	6.928	2.673
	II	5.105	2.290	4.400	7.940	3.145	4.342	2.433	3.725	5.860	2.473
Cash holdings (%)	I	0.090	0.023	0.063	0.137	0.081	0.093	0.046	0.078	0.124	0.060
	II	0.086	0.024	0.064	0.132	0.076	0.106	0.063	0.097	0.137	0.059
Tangibility (%)	I	0.358	0.143	0.296	0.557	0.247	0.326	0.164	0.312	0.470	0.194
	II	0.386	0.129	0.319	0.651	0.280	0.309	0.144	0.272	0.479	0.196
GDP growth (%)	I	0.016	0.016	0.023	0.029	0.018	0.012	0.002	0.017	0.026	0.022
	II	0.022	0.018	0.022	0.025	0.005	0.014	0.008	0.014	0.022	0.011
# observations	I	3785					1258				
	II	1608					454				

Our sample is divided into the estimation period 2002–12 (period I) and the out-of-sample prediction period 2013–16 (period II).

TABLE 4 Overview of the estimated model specifications.

Variable category	Variable	Regional models: Base	Regional models: CSP	Merged: CSP
CSP variables	Asset4 score		x_0	x_0
	Interaction North America & Asset4 score			x_1
Region variable	North America dummy			x_2
Control variables	Interest coverage A	x_1	x_1	x_3
	Interest coverage B	x_2	x_2	x_4
	Interest coverage C	x_3	x_3	x_5
	Interest coverage D	x_4	x_4	x_6
	Operating margin	x_5	x_5	x_7
	Long-term debt	x_6	x_6	x_8
	Total debt	x_7	x_7	x_9
	Market capitalization	x_8	x_8	x_{10}
	Beta	x_9	x_9	x_{11}
	Idiosyncratic risk	x_{10}	x_{10}	x_{12}
	Dividend payer dummy	x_{11}	x_{11}	x_{13}
	Market/book	x_{12}	x_{12}	x_{14}
	Retained earnings	x_{13}	x_{13}	x_{15}
	Capital expense	x_{14}	x_{14}	x_{16}
	Cash balance	x_{15}	x_{15}	x_{17}
	Tangibility	x_{16}	x_{16}	x_{18}
	GDP growth	x_{17}	x_{17}	x_{19}
	Dummy for year 1 (following years analog)	x_{18}	x_{18}	x_{20}
Boundaries	Lower boundary for rating AAA	μ_6	μ_6	μ_6
	Lower boundary for rating AA	μ_5	μ_5	μ_5
	Lower boundary for rating A	μ_4	μ_4	μ_4
	Lower boundary for rating BBB	μ_3	μ_3	μ_3
	Lower boundary for rating BB	μ_2	μ_2	μ_2
	Lower boundary for rating B	μ_1	μ_1	μ_1
Output	Linear predictor	y^*	y^*	y^*
	Rating class	R	R	R

The CSP models for isolated estimation of the regions North America and Europe include Asset4 and control variables, while the base model includes only controls between their independent variables. The boundaries required to assign rating classes, depending on the linear predictor, are the output of the regression. The Asset4 score represents the equal-weighted ES rating, the ENV score or the SOC score.

carried out by utilizing the maximum likelihood method referring to ordered probit models (McKelvey and Zavoina 1975; Venables and Ripley 2002). To account for the panel structure, we pool the observations and cluster standard errors on the firm level, which is appropriate for short panels. Wald p-values are calculated following the approach of Huber (1967) to reveal coefficient significance. Moreover, we include estimated thresholds for the various rating levels and the McFadden R^2 goodness-of-fit statistics. As the link function in ordered probit models limits the interpretation of the estimated coefficients, we also calculate marginal effects at the means to capture the impact of a marginal change in the ES score on the credit rating prediction, all other things being equal. The calculation is based on all independent variables being fixed at their means. To computationally derive the marginal effects, we follow Greene (2011).[3] One objective of credit portfolio models is to predict future credit ratings appropriately. To determine the quality of the credit rating prediction of our model specifications, we calculate the Somers's D values as a measure of the correlation between actual and predicted ratings.[4]

---

[3] Let $\boldsymbol{x}$ be the matrix of the independent variables. Then, the marginal effects give an indication of the extent to which the probability of a firm being assigned to a certain rating class changes based on the first derivative of the probabilities in (4.3):

$$\begin{aligned}
\frac{\partial P(R_{it} = 1 \mid \boldsymbol{x}_{i,t-1})}{\partial \boldsymbol{x}_{i,t-1}} &= -\phi(\mu_1 - \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})\boldsymbol{\beta}, \\
\frac{\partial P(R_{it} = j \mid \boldsymbol{x}_{i,t-1})}{\partial \boldsymbol{x}_{i,t-1}} &= [\phi(\mu_{j-1} - \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}) - \phi(\mu_j - \boldsymbol{x}'_{it}\boldsymbol{\beta})]\boldsymbol{\beta} \quad \text{for all } j \in \{2,3,4,5,6\}, \\
\frac{\partial P(R_{it} = 7 \mid \boldsymbol{x}_{i,t-1})}{\partial \boldsymbol{x}_{i,t-1}} &= \phi(\mu_6 - \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta})\boldsymbol{\beta}.
\end{aligned}$$

[4] According to Somers (1962), D is a measure of ordinal association. For actual ratings Z and predicted ratings Y, Newson (2001) defines Somers's D as

$$D_{YZ} = \frac{\tau(Z,Y)}{\tau(Z,Z)}, \tag{4.4}$$

with Kendall rank correlation coefficient

$$\tau = \frac{N_{\mathrm{C}} - N_{\mathrm{D}}}{n(n-1)/2}. \tag{4.5}$$

Kendall's τ is calculated by taking the difference in the number of concordant pairs N_{C} and the number of discordant pairs N_{D} as well as the sample size n. Two pairs (z_i, y_i) and (z_j, y_j) are called concordant if the ranks of both elements agree, eg, $z_i > z_j$ and $y_i > y_j$, or if $z_i < z_j$ and $y_i < y_j$. By contrast, two pairs are determined as being discordant if $z_i > z_j$ and $y_i < y_j$ or if $z_i < z_j$ and $y_i > y_j$. Somers's D can take values from -1 (only disagreeing pairs) to $+1$ (only agreeing pairs). In this context of measuring how predicted and actual credit ratings are associated, a Somers's D value of $+1$ expresses the optimal case in which all predictions are actually confirmed.

5 EMPIRICAL TESTS

Table 5 reports the results of the regional and the merged models. The estimation window for credit ratings in all of these probit models ranges from 2003 to 2013.

For the North American sample, each of the three CSP measures has a significantly positive coefficient. Thus, all else being equal, firms with a high level of CSP in the significant specifications have a higher probability of obtaining better credit ratings than firms with a low level of CSP. In the European sample, only the two specifications of the model with the ES and SOC scores show significance of the respective CSP measure at a 1% level. The ENV pillar provides no significant explanatory benefit for Europe in contrast to North America. The control variables display reasonable signs in the regressions, consistent with the findings of the literature cited in Section 3.2. To rule out the regionally different relations between CSP and credit ratings, we also provide a merged model of the North American and the European sample with an interaction term between region and CSP. The interpretation of the interaction effect is more difficult, as the marginal effect can even be of opposite sign (Ai and Norton 2003; Karaca-Mandic *et al* 2012). The results of our merged estimation are consistent with the isolated regional estimations. ES and SOC scores are relevant at the same level in both regions, as the coefficients of CSP scores are significant, while the interaction term between North America and CSP scores is not. The ENV score, in contrast, is only relevant in North America, as we document no significant coefficient for ENV in general but a significant interaction term between the North America dummy and the ENV score.

As the interpretation of coefficients is limited in terms of their magnitude, we estimate the marginal effect on the credit rating prediction of a change in the CSP scores and present the results in Table 6. For North America, we observe significant marginal effects for all three CSP scores. The lower and upper triangular matrixes for each score in Table 6 show a clear pattern, indicating that an increase in CSP scores significantly increases the probability of firms receiving a higher rating level and reduces the probability of the firms experiencing a rating downgrade. In particular, the marginal effects represent the difference in predicted probabilities for each rating class if, all else being equal, the mean CSP scores increase by 1%.

A detailed consideration of the diagonals shows that firms currently rated AAA, AA or A also show a significantly higher probability of being classified at the current rating level again, while it is less likely for firms currently rated BBB or worse to remain at the current rating level if the respective score is increased. For the ES measure, for instance, the predicted probability of an A-rated North American firm remaining in the A-rating category is 0.277% higher for a firm that has a 1% higher ES score than for a firm that is otherwise identical regarding the levels of the control variables. The probability of a BBB-rated firm obtaining a rating upgrade increases

TABLE 5 Estimation results of the ordered probit model for the North American and European data sets covering the years 2003–13. [Table continues on next page.]

	North America				Europe				Merged		
	Base	**ES**	**ENV**	**SOC**	**Base**	**ES**	**ENV**	**SOC**	**ES**	**ENV**	**SOC**
Coefficients											
Asset4 variable		0.014***	0.011***	0.012***		0.015***	0.006	0.018***	0.010***	0.005*	0.012***
Asset4 × North Am.									0.003	0.006**	−0.001
North Am. dummy									−0.632**	−0.879***	−0.460
Interest coverage A	0.209***	0.212***	0.213***	0.210***	0.286***	0.294***	0.289***	0.296***	0.239***	0.239***	0.236***
Interest coverage B	0.045	0.053*	0.051*	0.053*	0.065	0.061	0.061	0.068	0.052**	0.050**	0.054**
Interest coverage C	0.078***	0.084***	0.082***	0.083***	0.103***	0.109***	0.105***	0.111***	0.088***	0.086***	0.088***
Interest coverage D	0.005	0.005	0.005	0.005	−0.012	−0.013*	−0.012*	−0.013*	0.002	0.002	0.002
Operating margin	0.000	0.008	0.007	0.006	0.031***	0.035***	0.033***	0.034***	0.014***	0.013***	0.012***
Long-term debt	−0.050***	−0.044***	−0.045***	−0.045***	0.000	0.002	0.001	0.001	−0.024***	−0.024***	−0.025***
Total debt	0.030**	0.024**	0.024**	0.025**	−0.017	−0.019	−0.018	−0.018	0.005	0.005	0.006
Size	0.074***	0.060***	0.063***	0.061***	0.079***	0.070***	0.075***	0.069***	0.059***	0.062***	0.059***
Beta	−0.179	−0.189*	−0.201*	−0.173	−1.518***	−1.524***	−1.510***	−1.561***	−0.473***	−0.483***	−0.467***
Idiosyncratic risk	−1.646***	−1.631***	−1.619***	−1.647***	−1.499***	−1.515***	−1.509***	−1.502***	−1.578***	−1.568***	−1.587***
Dividend payer	1.080***	0.998***	1.006***	1.017***	1.204***	1.176***	1.190***	1.168***	0.989***	0.996***	1.007***
Market/book	0.090***	0.081***	0.084***	0.081***	0.089*	0.090*	0.091*	0.082	0.082***	0.085***	0.081***
Retained earnings	1.493***	1.481***	1.506***	1.456***	1.675***	1.572***	1.614***	1.606***	1.479***	1.511***	1.462***
Capital expense	−0.016	−0.010	−0.010	−0.013	−0.002	0.001	−0.001	0.003	−0.016	−0.016	−0.019

TABLE 5 Continued.

	North America				Europe				Merged		
	Base	ES	ENV	SOC	Base	ES	ENV	SOC	ES	ENV	SOC
Coefficients (cont.)											
Cash holdings	−0.345	−0.866	−0.862	−0.703	1.182	0.934	1.100	0.834	−0.378	−0.353	−0.267
Tangibility	0.337	0.182	0.155	0.264	1.463***	1.315***	1.397***	1.296***	0.560***	0.547**	0.637***
GDP growth	5.681***	0.841***	1.036***	2.186***	19.238***	20.911***	19.851***	21.326***	8.534***	7.733***	8.760***
Time dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Lower boundaries											
AAA–AA	12.488	12.208	12.284	12.212	14.388	14.854	14.541	15.045	11.988	11.805	12.110
AA–A	10.957	10.638	10.721	10.647	10.861	11.313	11.012	11.490	10.014	9.837	10.141
A–BBB	7.584	7.171	7.261	7.203	7.502	7.935	7.649	8.095	6.658	6.489	6.797
BBB–BB	3.924	3.435	3.539	3.476	3.640	4.047	3.781	4.187	2.960	2.802	3.105
BB–B	0.431	−0.058	0.052	−0.025	0.323	0.727	0.464	0.860	−0.505	−0.654	−0.369
B–C	−4.065	−4.490	−4.401	−4.453	−2.714	−2.277	−2.557	−2.156	−4.655	−4.825	−4.516
Goodness-of-fit											
McFadden R^2	0.367	0.375	0.374	0.373	0.334	0.337	0.335	0.340	0.362	0.361	0.361
# observations	3785	3785	3785	3785	1258	1258	1258	1258	5043	5043	5043

Estimation is carried out by utilizing the maximum likelihood method referring to ordered probit models. Coefficients of all variables are displayed, including the significance level, marked by asterisks. They are regarded as being statistically significant at a level of 1% (***), 5% (**) or 10% (*) when the p-value is below these levels. The lower boundaries of the rating categories corresponding to those in Section 4 are also displayed.

by 0.330% per 1% ES score (equaling the sum of probabilities to obtain the AAA, AA or A grade). In terms of absolute values, this emerges as an average saving of approximately US$1.45 concerning Basel III economic capital per loan nominal of US$1000 based on the ES score's change by one standard deviation (ie, a saving of 14.5bps).[5] For the European sample, the results for the ES score reveal that firms with a higher score have, all else being equal, a higher probability of remaining in the current credit rating class (firms with current credit ratings AAA, AA or A) or an increased probability of a rating migration into a better credit rating class (firms across all current credit ratings). European firms with a current credit rating of BBB, BB, B or C are less likely to remain in the current credit rating class. For instance, firms with a current credit rating of BBB exhibit a decrease by 0.169% in the probability of remaining in credit rating BBB and a decrease by 0.085% in the probability of experiencing a downgrade by one notch to noninvestment grade, as well as an increase by 0.244% in the probability of experiencing an upgrade to level A.

Table 7 shows the results of the analysis of the prediction quality. A positive value in the Delta column indicates that incorporating the respective CSP scores into the baseline model increases the prediction quality. For instance, considering the ES score in our credit risk model increases the Somers's D of the North American sample by 0.82%. We apply the Wilcoxon–Mann–Whitney (WMW) test on the probabilities that the actual credit rating is predicted and find a p-value of 0.01%, indicating that the increase is different from zero. For the two single pillar scores, we also find reasonable increases in Somers's D. In particular, for the ENV score, the increases of 0.80% in Somers's D are higher than for the SOC model (with 0.59%). In the European sample, none of the CSP models shows any relevant improvement in Somers's D compared with the base model. Following the improvement in prediction quality for North American ES and SOC and the significant coefficients for both measures for both regions, we would expect similar results in Europe, but we find lower ones. This deviation may depend on the different CSP distributions among firms in these regions. According to Table 3, the mean ES score is 51.5%, lower in North America than the European sample with 81.7%. Its standard deviation of 28.1% is higher than its counterpart, 15.8%, in Europe. The median of 88.4%, compared with the lower mean, indicates an ES distribution for Europe that is skewed to

[5] This calculation is based on the internal ratings-based approach for general corporates described in Basel Committee on Banking Supervision (2017) on a time horizon of one year. We assume the loss given default (LGD) rate to be 40%, analogous to the supervisory LGD for unsecured corporate exposure, which refers to the foundation approach. The required probabilities of default are provided by US average historic one-year corporate rating transition rates (1981–2016) according to Vazza and Kraemer (2017).

the left. The SOC distribution is similar. In order to explain regional deviations in the predictive performance, the variability of the single CSP dimensions requires further considerations, which we present in the following section.

6 DISCUSSION

As the ES score is an aggregation over the dimensions of ENV and SOC, we next take a more thorough look at the effects of these two dimensions.

6.1 Environmental regulation and credit ratings

We find that the environmental performance has explanatory power for credit ratings in North America. Firms with good environmental performance are more likely to be rewarded with a better credit rating (Tables 5 and 6). The prediction quality also increases by 0.8% (Table 7). In contrast, there is no observable relevance of environmental performance for Europe (Tables 5–7). To explain the observed difference, we consider the structure of the ENV score. It consists of three categories: resource reduction, emission reduction and product innovation. Geological conditions may influence the awareness of resource and emission reduction. North America is endowed with an abundance of natural resources. In contrast, Europe lacks such a variety of commodities. As a result, the necessity to use resources economically is higher in Europe. Further differences exist in the legislation of the two regions. Environmental regulations are weaker in the United States than in European countries such as Denmark and Sweden (Johnstone *et al* 2012). The Kyoto Protocol of 1997 is an example of the willingness to accept binding environmental protection agreements. It states a reduction goal of 8% for greenhouse gases for the European community, but less (7%) for the United States, which has, however, never ratified it (UNFCCC 1998). The United States has not yet adopted the Doha Amendment to the Kyoto Protocol. The different geographical and political circumstances between North America and Europe are reflected in the different average ENV scores. Dorfleitner *et al* (2018) show that, for the United States, high ENV scores can predict positive earnings surprises in later periods, which can partly explain the positive effect on creditworthiness that we find for North America. Finally, firms in North America have a higher degree of freedom to differentiate themselves from their peers regarding environmental issues compared with Europe, which results in explanatory and prediction power improvements only for North America.

6.2 The benefits of social politics

In the credit rating regressions (Tables 5), the coefficient for social performance is significant in both North America and Europe at a 1% level. Moreover, analysis of

TABLE 6 Marginal effects at means based on the estimation for panels North America and Europe covering the years from 2003 to 2013. [Table continues on next page.]

	North America						
	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**C**
ES							
AAA	1.321***	1.840***	–2.517***	–0.627***	–0.016***	–0.001***	0.000**
AA	0.691***	1.595***	–1.044***	–1.205***	–0.035***	–0.001***	0.000**
A	0.118***	0.413***	2.771***	–3.074***	–0.221***	–0.007***	0.000**
BBB	0.016***	0.062***	1.718***	–0.440***	–1.304***	–0.051***	–0.001***
BB	0.001***	0.004***	0.137***	2.749***	–2.085***	–0.795***	–0.011***
B	0.000***	0.000***	0.013***	0.501***	2.791***	–3.189***	–0.117***
C	0.001***	0.002***	0.074***	2.038***	–0.766***	–1.328***	–0.020***
ENV							
AAA	1.035***	1.474***	–1.973***	–0.522***	–0.014***	0.000***	0.000**
AA	0.552***	1.269***	–0.812***	–0.979***	–0.029***	–0.001***	0.000**
A	0.096***	0.334***	2.214***	–2.459***	–0.179***	–0.006***	0.000**
BBB	0.013***	0.050***	1.374***	–0.336***	–1.058***	–0.042***	0.000***
BB	0.001***	0.003***	0.113***	2.213***	–1.689***	–0.632***	–0.008***
B	0.000***	0.000***	0.010***	0.399***	2.243***	–2.561***	–0.093***
C	0.000***	0.002***	0.061***	1.642***	–0.628***	–1.061***	–0.016***
SOC							
AAA	1.178***	1.625***	–2.222***	–0.566***	–0.015***	0.000***	0.000**
AA	0.607***	1.401***	–0.874***	–1.100***	–0.033***	–0.001***	0.000**
A	0.107***	0.370***	2.454***	–2.725***	–0.199***	–0.006***	0.000**
BBB	0.015***	0.056***	1.531***	–0.400***	–1.157***	–0.045***	–0.001***
BB	0.001***	0.003***	0.122***	2.434***	–1.850***	–0.701***	–0.010***
B	0.000***	0.000***	0.012***	0.445***	2.469***	–2.823***	–0.103***
C	0.000***	0.002***	0.066***	1.805***	–0.684***	–1.171***	–0.018***

the marginal effects reveals a distinct increase in the probability of both regions either maintaining their current rating or even migrating to a better rating class (Table 6). As firms profit from high CSP by being able to hire better qualified employees (Turban and Greening 1997), it seems intuitive that this is true in both North America and Europe. Besides improvements in explanatory power, good social performance in North America results in significantly better credit rating predictions; the implied increase in prediction quality amounts to 0.6%, while in Europe it is only 0.1% (Table 7). Although prediction quality improves only in North America, we can confirm the impact of social performance for the European sample based on the coefficient estimation and the marginal effects. The relevance of the different

TABLE 6 Continued.

	Europe						
	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**C**
ES							
AAA	0.000*	0.080	−1.778***	−0.085**	−0.002*	0.000	0.000
AA	0.145*	2.668***	−1.621***	−1.163***	−0.028***	−0.001**	0.000
A	0.026*	0.775***	2.568***	−3.206***	−0.157***	−0.006**	0.000
BBB	0.004*	0.141***	2.439***	−1.693***	−0.854***	−0.035**	−0.002
BB	0.000*	0.010**	0.274***	3.385***	−3.168***	−0.475***	−0.027*
B	0.000	0.001**	0.019**	0.846**	2.473***	−2.968***	−0.372*
C	0.000*	0.005**	0.149***	3.157***	−2.425***	−0.837***	−0.050*
ENV							
AAA	0.663	0.094	−0.720	−0.036	−0.001	0.000	0.000
AA	0.057	1.028	−0.616	−0.457	−0.011	0.000	0.000
A	0.010	0.300	0.997	−1.243	−0.063	−0.002	0.000
BBB	0.002	0.056	0.948	−0.658	−0.333	−0.014	−0.001
BB	0.000	0.004	0.110	1.305	−1.229	−0.181	−0.010
B	0.000	0.000	0.008	0.323	0.968	−1.155	−0.144
C	0.000	0.002	0.059	1.224	−0.946	−0.321	−0.019
SOC							
AAA	2.228**	−0.009	−2.119***	−0.098**	−0.002**	0.000	0.000
AA	0.176**	3.274***	−2.014***	−1.401***	−0.033***	−0.001**	0.000
A	0.031**	0.944***	3.136***	−3.918***	−0.187***	−0.007**	0.000
BBB	0.005**	0.168***	2.979***	−2.074***	−1.034***	−0.042***	−0.002*
BB	0.000*	0.011***	0.325***	4.157***	−3.873***	−0.588***	−0.033*
B	0.000*	0.001**	0.023**	1.006***	3.092***	−3.660***	−0.462**
C	0.000*	0.006***	0.176***	3.849***	−2.934***	−1.036***	−0.061*

The marginal effects are listed in per mill and show the impact of an increase in the ES scores by one percentage point, all else being equal, on the predicted probabilities of occurrence for the various rating classes. The row sum for each panel is zero since the sum of the predicted probabilities across all rating classes equals one, and changes in probabilities hence equal zero. Coefficients are marked as significant at a level of 1% (***), 5% (**) or 10% (*) when the p-value is below these levels.

prediction qualities in SOC score for North America and Europe is underpinned by the varying score levels in the two regions. Referring to the descriptive statistics of our sample, the SOC score in North America is distinctly lower than it is in Europe (58.7 versus 85.7), while the variance is higher (standard deviation 27.6 versus 14.4). The literature has, up to now, identified differences between both regions regarding many aspects of the SOC score, such as employment quality, health and safety, training and development, diversity and opportunity, community and product responsibility. An important indicator is the social expense of a country: the United States

TABLE 7 Somers's D values for panels of North America and Europe for predictions in the period 2014–17.

	North America		
	WMW p-value	**Somers's D**	**Delta**
Base model		0.5968	
ES model	0.0001	0.6050	0.0082
ENV model	0.0005	0.6048	0.0080
SOC model	0.0001	0.6027	0.0059
# observations	1608		
	Europe		
	WMW p-value	**Somers's D**	**Delta**
Base model		0.5695	
ES model	0.9778	0.5705	0.0010
ENV model	0.9530	0.5705	0.0009
SOC model	0.9396	0.5703	0.0008
# observations	454		

We use Somers's D to measure the correlation between predicted ratings and actual ratings. D can take values from -1 to $+1$, where the latter is the optimal case in which all predictions are confirmed. We show the differences between ES models' Somers's D and those of the base models in order to illustrate the improvement ascribed to CSP. The Wilcoxon–Mann–Whitney (WMW) test provides p-values to evaluate whether the probabilities of correct predictions are significantly higher in the CSP models than in the benchmark model.

and Canada spent 19.2% and 17.0% of their GDP in 2014, respectively. European countries such as France and Germany tend toward a higher level of expenditure (31.9% and 25.8%, respectively; see OECD (2016)). Unlike Europe, North America, and the United States in particular, lacks a comprehensive labor market policy (Matten and Moon 2008), as well as labor unions with strong negotiating power (Du Caju *et al* 2009), proper employment protection and mandatory health protection (Pfeffer 2010). Further, the World Economic Forum (2017) gender gap report, which analyzes the emancipation of women and men regarding economic participation and opportunity, educational attainment, health and survival, and political empowerment, shows that the United States ranked forty-ninth (worldwide), while many European countries do better (such as Germany, ranked twelfth). Strategies for human rights protection are more common in Europe than in North America (Welford 2005). Firms in North America have a higher degree of freedom to differentiate themselves from their peers concerning social issues. Further, the diversification of the European firms in terms of social performance decreased between the in-sample period and the out-of-sample period (standard deviation 17.0 versus 14.4). The high basic level and decreasing variability of social scores in Europe have led to a certain degree of similarity between firms, which explains the insignificant results in our prediction models despite their increased explanatory power.

7 ROBUSTNESS CHECKS

This section contains several robustness checks to rule out the fact that our results may be driven by our methodological framework, the sample selection, a missing data bias, a local bias or a time bias.

We analyze whether the general risk level of firms affects the impact of CSP ratings on credit ratings and cluster firms into an investment-grade group and a non-investment-grade (speculative) group. The marginal effects in Table 6 show that the impact of CSP on credit rating predictions differs across rating classes. The better the initial rating class of a firm is, the lower the conditional probability by which an increase in the ES score predicts a better credit rating (except for rating class C). The results for single pillars (ENV and SOC) show similar patterns to the ES results. In Europe, we also find evidence comparable to North America for the SOC score, which alone appears significant there. Also, previous studies show differences between the impact of CSP on the cluster of investment-grade ratings and non-investment-grade ratings. In the United States, for specific CSP factors (such as the percentage of independent directors from the corporate governance dimension), the significance of the impact on credit ratings diminishes when restricting the sample to investment-grade bonds only (Ashbaugh-Skaife *et al* 2006). Therefore, we reestimate our models based on the investment-grade subsample of our analysis. Tables B.9 and B.10 in the online appendix contain the results. The levels of significance in the investment-grade subset decrease, certainly to some extent, due to the smaller number of observations. We document the improvements in prediction quality in North America that are lacking compared with the full sample (Panel A of Table B.10), while the improvement from the entire sample is 0.82% (see Table 7).

Since endogenous CSP ratings may generate reverse causality issues, we replace the firm ES ratings by industry-based ES rating ranks and rerun our analyses. Table B.8 (in the online appendix) presents an overview of the industry classes of our data. The absolute range of possible CSP activities can vary across industries; to address systematic differences across industries, we follow Utz (2018) and modify the Asset4 CSP ratings. First, firms in each industry are ranked by their Asset4 score. Second, we assign to each firm its percentile score within the respective industry. Hence, the best CSP performing firm in each industry holds a value of 1 and the lowest CSP performing firm holds a value of 0. Overall, the credit-rating explanation quality in terms of the significance of explanatory variables remains the same (see

Table B.9 in the online appendix) and thus the implications are compatible with our main results. The Somers's D improvement has a lower magnitude (see Table B.10 in the online appendix). However, the decrease compared with the standard case amounts only to 0.13% for ES when we exclude utilities, since the model cannot capture all the relevant effects of this industry class. The results show that the loss of information (the original distance between firms concerning their CSP ratings) results in a smaller improvement in the prediction quality.

Further, we also check the robustness by adding industry fixed effects, which the standard model in the literature does not contain (Dimitrov *et al* 2015). The significance levels of the coefficients for North America and Europe remain unchanged. The Somers's D values decrease at first glance but remain comparable if we again exclude utilities (see Table B.10 in the online appendix).

We capture a possible effect in the results of excluded observations due to the lack of control variables data. Instead of discarding these observations, we substitute the missing values with the mean, according to the mean imputation method (Schafer 1997). When rerunning our regressions, we find the same significance of the Asset4 scores (see Table B.9 in the online appendix) and even higher Somers's D improvements (see Table B.10 in the online appendix).

Next, we restrict our North American sample (5393 observations) to a US sample (4849 observations) under the assumption that the underlying drivers for credit rating predictions are more homogeneous inside the domestic market compared with the region. Again, we observe no relevant changes in the significance of the CSP scores (see Table B.9 in the online appendix). The impact of CSP on the credit-rating prediction quality even increases, to 1.0% for ENV and 0.7% for SOC (see Table B.10 in the online appendix).

Moreover, we check the robustness of results according to changes in the observation period for the parameter estimation panel and the prediction data set. The main results rely on a period covering observations between 2003 and 2013, while the prediction data set includes observations from 2014 to 2017. To increase the number of observations in the prediction data set, we reduce the years considered in the estimation data set in favor of the prediction data set. We choose two estimation data sets lasting until 2011 and 2012 with a prediction data set commencing accordingly in 2012 and 2013, respectively. Overall, there is no major difference in the significance of the scores (see Table B.9 in the online appendix) and the improvement in credit-rating prediction quality by considering CSP (see Table B.10 in the online appendix).

Finally, we check whether multicollinearity impairs our regressions for all main regressions and robustness checks. In all the regional regressions, the variance inflation factor (VIF) of the CSP variable is below 1.63, which implies that we do not have a multicollinearity problem. Naturally, the VIF values are higher in the merged sample because the interaction term is strongly correlated with the CSP score, as are most observations for North America. However, as this regression is only there to capture the significance of the difference between Europe and North America, this does not compromise the validity of our results.

8 CONCLUSION

One central question in the finance literature addresses the prediction quality of credit ratings (Blume *et al* 1998; Kisgen 2006). CSP is an additional informational proxy for factors that reduce firm risk, as shown in several studies (see, for example, Kim *et al* 2014; Utz 2018). Therefore, we investigate whether, how and the extent to which the integration of CSP measures in credit rating predictions improves their quality. The relationship between CSP and credit ratings is significantly positive in North America, ie, high CSP performance goes along with better credit ratings. In addition, out-of-sample predictions are improved by 0.8%. In Europe, the social score adds informational power to a basic prediction model, while the environmental performance does not. In general, the differing regional impact of the environmental performance is presumably due to regional differences in both economic areas, such as stronger existing legal and cultural frameworks. To embed our results in a theoretical framework, our findings show supporting evidence of the risk mitigation view of high CSP.

We resolve the contrary results of earlier studies by generating comparable findings for an international sample consisting of North America and Europe. Our results are based on a consistent identification of the explanatory power and the quantification of the prediction quality of CSP dimensions for both regions. In particular, our study builds on the findings of Stellner *et al* (2015) with an analysis of the single performance of the environmental and social dimensions as well as their aggregate. While Stellner *et al* (2015) show that (aggregated) CSP has no impact on the credit ratings of European firms, we ascertain that the social performance is a significant explanatory factor for credit ratings. Moreover, we confirm the findings of Jiraporn *et al* (2014) that North American firms with high CSP obtain better credit ratings, although our study uses the methodologically more sophisticated Asset4 scores (see Humphrey *et al* 2012). We complement this study by quantifying the improvement of the prediction quality, in particular, by 0.8 percentage points for North America; this is economically significant, as there is less economic capital required. Overall, we find supporting evidence for the impact of CSP performance on credit ratings being independent of the sustainability rating agency in North America and Europe.

Since the country level of CSP is important in the relationship between CSP and creditworthiness (Stellner *et al* 2015), future research may extend our study to different regions. This extension is particularly interesting since Utz (2018) finds evidence for crash risk such that – consistent with our results – the risk mitigation view holds in North America and Europe; however, the overinvestment hypothesis applies in the Asia-Pacific region. As this study focuses rather on the technical effects for credit risk models, a further extension of this research could be to dig deeper into the economic channels through which the observed effects causally emerge.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ACKNOWLEDGEMENTS

We thank Kenneth French for providing the factor returns used in our model estimation. We also appreciate the valuable comments of an anonymous referee that helped us to strengthen the message and increase the validity of our paper.

REFERENCES

Acharya, V., Davydenko, S. A., and Strebulaev, I. A. (2012). Cash holdings and credit risk. *Review of Financial Studies* **25**(12), 3572–3609 (https://doi.org/10.1093/rfs/hhs106).

Ai, C., and Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters* **80**(1), 123–129 (https://doi.org/10.1016/S0165-1765(03)00032-6).

Alp, A. (2013). Structural shifts in credit rating standards. *Journal of Finance* **68**(6), 2435–2470 (https://doi.org/10.1111/jofi.12070).

Altman, E. I., Haldeman, R. G., and Narayanan, P. (1977). Zeta analysis: a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* **1**(1), 29–54 (https://doi.org/10.1016/0378-4266(77)90017-6).

Ashbaugh-Skaife, H., Collins, D. W., and LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics* **42**(1), 203–243 (https://doi.org/10.1016/j.jacceco.2006.02.003).

Attig, N., El Ghoul, S., Guedhami, O., and Suh, J. (2013). Corporate social responsibility and credit ratings. *Journal of Business Ethics* **117**(4), 679–694 (https://doi.org/10.1007/s10551-013-1714-2).

Aupperle, K., Carroll, A., and Hatfield, J. D. (1985). An empirical examination of the relationship between corporate social responsibility and profitability. *Academy of Management Journal* **28**(2), 446–463. URL: http://www.jstor.org/stable/256210.

Baghai, R. P., Servaes, H., and Tamayo, A. (2014). Have rating agencies become more conservative? Implications for capital structure and debt pricing. *Journal of Finance* **69**(5), 1961–2005 (https://doi.org/10.1111/jofi.12153).

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance* **26**(2), 445–474 (https://doi.org/10.1016/S0378-4266(01)00229-1).

Basel Committee on Banking Supervision (2017). Basel III: finalising post-crisis reforms. Standard, December 7, Bank for International Settlements, Basel. URL: https://www.bis.org/bcbs/publ/d424.htm.

Bauer, R., and Hann, D. (2010). Corporate environmental management and credit risk. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.1660470).

Bauer, R., Derwall, J., and Hann, D. (2009). Employee relations and credit risk. Working Paper, Social Science Research Network (https://doi.org/10.2139/ssrn.1483112).

Becker, B., and Milbourn, T. (2011). How did increased competition affect credit ratings? *Journal of Financial Economics* **101**(3), 493–514 (https://doi.org/10.1016/j.jfineco.2011.03.012).

Blume, M. E., Lim, F., and MacKinlay, A. C. (1998). The declining credit quality of US corporate debt: myth or reality? *Journal of Finance* **53**(4), 1389–1413 (https://doi.org/10.1111/0022-1082.00057).

Brammer, S., and Millington, A. (2008). Does it pay to be different? An analysis of the relationship between corporate social and financial performance. *Strategic Management Journal* **29**(12), 1325–1343 (https://doi.org/10.1002/smj.714).

Cai, Y., Pan, C. H., and Statman, M. (2016). Why do countries matter so much in corporate social performance? *Journal of Corporate Finance* **41**, 591–609 (https://doi.org/10.1016/j.jcorpfin.2016.09.004).

Campbell, J. Y., Hilscher, J., and Szilagyi, J. (2008). In search of distress risk. *Journal of Finance* **63**(6), 2899–2939 (https://doi.org/10.1111/j.1540-6261.2008.01416.x).

Chatterji, A. K., and Levine, D. (2006). Breaking down the wall of codes: evaluating non-financial performance measurement. *California Management Review* **48**(2), 29–51 (https://doi.org/10.2307/41166337).

Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (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).

Chen, H., Kacperczyk, M., and Ortiz-Molina, H. (2011). Do nonfinancial stakeholders affect the pricing of risky debt? Evidence from unionized workers. *Review of Finance* **16**(2), 347–383 (https://doi.org/10.1093/rof/rfq028).

Cornell, B., and Shapiro, A. C. (1987). Corporate stakeholders and corporate finance. *Financial Management* **16**(1), 5–14. URL: https://www.jstor.org/stable/3665543.

DeAngelo, H., DeAngelo, L., and Stulz, R. M. (2006). Dividend policy and the earned/contributed capital mix: a test of the life-cycle theory. *Journal of Financial Economics* **81**(2), 227–254 (https://doi.org/10.1016/j.jfineco.2005.07.005).

Dhaliwal, D. S., Li, O. Z., Tsang, A., and Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: the initiation of corporate social responsibility reporting. *Accounting Review* **86**(1), 59–100 (https://doi.org/10.2308/accr.00000005).

Dimitrov, V., Palia, D., and Tang, L. (2015). Impact of the Dodd–Frank act on credit ratings. *Journal of Financial Economics* **115**(3), 505–520 (https://doi.org/10.1016/j.jfineco.2014.10.012).

Doh, J. P., and Guay, T. R. (2006). Corporate social responsibility, public policy, and NGO activism in Europe and the United States: an institutional-stakeholder perspective. *Journal of Management Studies* **43**(1), 47–73 (https://doi.org/10.1111/j.1467-6486.2006.00582.x).

Dorfleitner, G., Halbritter, G., and Nguyen, M. (2015). Measuring the level and risk of corporate responsibility: an empirical comparison of different ESG rating approaches. *Journal of Asset Management* **16**(7), 450–466 (https://doi.org/10.1057/jam.2015.31).

Dorfleitner, G., Utz, S., and Wimmer, M. (2018). Patience pays off: corporate social responsibility and long-term stock returns. *Journal of Sustainable Finance and Investment* **8**(2), 132–157 (https://doi.org/10.1080/20430795.2017.1403272).

Du Caju, P., Gautier, E., Momferatou, D., and Ward-Warmedinger, M. E. (2009). Institutional features of wage bargaining in 23 European countries, the US and Japan. ECB Working Paper Series 974, European Central Bank, Frankfurt. URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp974.pdf.

El Ghoul, S., Guedhami, O., Kwok, C. C., and Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking and Finance* **35**(9), 2388–2406 (https://doi.org/10.1016/j.jbankfin.2011.02.007).

El Ghoul, S., Guedhami, O., and Kim, Y. (2017). Country-level institutions, firm value, and the role of corporate social responsibility initiatives. *Journal of International Business Studies* **48**(3), 360–385 (https://doi.org/10.1057/jibs.2016.4).

Elton, E. J., Gruber, M. J., Agrawal, D., and Mann, C. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance* **56**(1), 247–277 (https://doi.org/10.1111/0022-1082.00324).

Fama, E. F., and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics* **105**(3), 457–472 (https://doi.org/10.1016/j.jfineco.2012.05.011).

Fons, J. S. (1994). Using default rates to model the term structure of credit risk. *Financial Analysts Journal* **50**(5), 25–32 (https://doi.org/10.2469/faj.v50.n5.25).

Friewald, N., Wagner, C., and Zechner, J. (2014). The cross-section of credit risk premia and equity returns. *Journal of Finance* **69**(6), 2419–2469 (https://doi.org/10.1111/jofi.12143).

Frooman, J., Zietsma, C., and McKnight, B. (2008). There is no good reason not to be good. In *Proceedings for the Administrative Science Association of Canada (ASAC 2008), Halifax, Nova Scotia*, pp. 141–158. ASAC. URL: https://bit.ly/39q48CW.

Goss, A., and Roberts, G. S. (2011). The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking and Finance* **35**(7), 1794–1810 (https://doi.org/10.1016/j.jbankfin.2010.12.002).

Greene, W. H. (2011). *Econometric Analysis*. Pearson Higher Education, Harlow.

Hoberg, G., and Prabhala, N. R. (2009). Disappearing dividends, catering, and risk. *Review of Financial Studies* **22**(1), 79–116 (https://doi.org/10.1093/rfs/hhn073).

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 221–233. University of California Press, Berkeley, CA.

Humphrey, J. E., Lee, D. D., and Shen, Y. (2012). Does it cost to be sustainable? *Journal of Corporate Finance* **18**(3), 626–639 (https://doi.org/10.1016/j.jcorpfin.2012.03.002).

Ioannou, I., and Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies* **43**(9), 834–864 (https://doi.org/10.1057/jibs.2012.26).

Jiang, J. X., Stanford, M. H., and Xie, Y. (2012). Does it matter who pays for bond ratings? Historical evidence. *Journal of Financial Economics* **105**(3), 607–621 (https://doi.org/10.1016/j.jfineco.2012.04.001).

Jiraporn, P., Jiraporn, N., Boeprasert, A., and Chang, K. (2014). Does corporate social responsibility (CSR) improve credit ratings? Evidence from geographic identification. *Financial Management* **43**(3), 505–531 (https://doi.org/10.1111/fima.12044).

Johnstone, N., Haščič, I., Poirier, J., Hemar, M., and Michel, C. (2012). Environmental policy stringency and technological innovation: evidence from survey data and patent counts. *Applied Economics* **44**(17), 2157–2170 (https://doi.org/10.1080/00036846.2011.560110).

Kang, C., Germann, F., and Grewal, R. (2016). Washing away your sins? Corporate social responsibility, corporate social irresponsibility, and firm performance. *Journal of Marketing* **80**(2), 59–79 (https://doi.org/10.1509/jm.15.0324).

Kaplan, R. S., and Urwitz, G. (1979). Statistical models of bond ratings: a methodological inquiry. *Journal of Business* **52**(2), 231–261 (https://doi.org/10.1086/296045).

Karaca-Mandic, P., Norton, E. C., and Dowd, B. (2012). Interaction terms in nonlinear models. *Health Services Research* **47**, 255–274 (https://doi.org/10.1111/j.1475-6773.2011.01314.x).

Kim, Y., Li, H., and Li, S. (2014). Corporate social responsibility and stock price crash risk. *Journal of Banking and Finance* **43**, 1–13 (https://doi.org/10.1016/j.jbankfin.2014.02.013).

Kisgen, D. J. (2006). Credit ratings and capital structure. *Journal of Finance* **61**(3), 1035–1072 (https://doi.org/10.1111/j.1540-6261.2006.00866.x).

Liang, H., and Renneboog, L. (2017). On the foundations of corporate social responsibility. *Journal of Finance* **72**(2), 853–910 (https://doi.org/10.1111/jofi.12487).

Lodge, G. C. (1990). *Comparative Business–Government Relations*. Prentice Hall, Englewood Cliffs, NJ.

Luo, X., Wang, H., Raithel, S., and Zheng, Q. (2015). Corporate social performance, analyst stock recommendations, and firm future returns. *Strategic Management Journal* **36**(1), 123–136 (https://doi.org/10.1002/smj.2219).

Maignan, I. (2001). Consumers' perceptions of corporate social responsibilities: a cross-cultural comparison. *Journal of Business Ethics* **30**(1), 57–72 (https://doi.org/10.1023/A:1006433928640).

Maignan, I., and Ralston, D. A. (2002). Corporate social responsibility in Europe and the US: insights from businesses' self-presentations. *Journal of International Business Studies* **33**(3), 497–514 (https://doi.org/10.1057/palgrave.jibs.8491028).

Matten, D., and Moon, J. (2008). "Implicit" and "explicit" CSR: a conceptual framework for a comparative understanding of corporate social responsibility. *Academy of Management Review* **33**(2), 404–424 (https://doi.org/10.5465/amr.2008.31193458).

McKelvey, R. D., and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology* **4**(1), 103–120 (https://doi.org/10.1080/0022250X.1975.9989847).

Merton, R. C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* **29**(2), 449–470 (https://doi.org/10.2307/2978814).

Newson, R. (2001). Parameters behind “nonparametric” statistics: Kendall’s tau, Somers’ D and median differences. *Stata Journal* **1**(1), 1–20 (https://doi.org/10.1177/1536867X0200200103).

OECD (2016). *OECD Factbook 2015–2016: Economic, Environmental and Social Statistics*. Organisation for Economic Co-operation and Development (https://doi.org/10.1787/18147364).

Oikonomou, I., Brooks, C., and Pavelin, S. (2014). The effects of corporate social performance on the cost of corporate debt and credit ratings. *Financial Review* **49**(1), 49–75 (https://doi.org/10.1111/fire.12025).

Orlitzky, M. (2008). Corporate social performance and financial performance: a research synthesis. In *The Oxford Handbook of Corporate Social Responsibility*, Crane, A., McWilliams, A., Matten, D., Moon, J., and Siegel, D. S. (eds), pp. 113–134. Oxford University Press (https://doi.org/10.1093/oxfordhb/9780199211593.003.0005).

Pástor, L., and Pietro, V. (2003). Stock valuation and learning about profitability. *Journal of Finance* **58**(5), 1749–1789 (https://doi.org/10.1111/1540-6261.00587).

Pfeffer, J. (2010). Building sustainable organizations: the human factor. *Academy of Management Perspectives* **24**(1), 34–45 (https://doi.org/10.5465/amp.24.1.34).

Rampini, A. A., and Viswanathan, S. (2013). Collateral and capital structure. *Journal of Financial Economics* **109**(2), 466–492 (https://doi.org/10.1016/j.jfineco.2013.03.002).

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. CRC Press, Boca Raton, FL.

Schneider, T. E. (2011). Is environmental performance a determinant of bond pricing? Evidence from the US pulp and paper and chemical industries. *Contemporary Accounting Research* **28**(5), 1537–1561 (https://doi.org/10.1111/j.1911-3846.2010.01064.x).

Sharfman, M. P., and Fernando, C. S. (2008). Environmental risk management and the cost of capital. *Strategic Management Journal* **29**(6), 569–592 (https://doi.org/10.1002/smj.678).

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review* **27**(6), 799–811 (https://doi.org/10.2307/2090408).

Sotorrío, L. L., and Sánchez, J. L. F. (2008). Corporate social responsibility of the most highly reputed European and North American firms. *Journal of Business Ethics* **82**(2), 379–390 (https://doi.org/10.1007/s10551-008-9901-2).

Standard & Poor’s (2013). Corporate criteria: ratios and adjustments. URL: https://bit.ly/2D7cTGc.

Stellner, C., Klein, C., and Zwergel, B. (2015). Corporate social responsibility and eurozone corporate bonds: the moderating role of country sustainability. *Journal of Banking and Finance* **59**, 538–549 (https://doi.org/10.1016/j.jbankfin.2015.04.032).

Tang, T. T. (2009). Information asymmetry and firms’ credit market access: evidence from Moody’s credit rating format refinement. *Journal of Financial Economics* **93**(2), 325–351 (https://doi.org/10.1016/j.jfineco.2008.07.007).

Thomson Reuters (2011). Asset4 ESG data glossary. URL: https://bit.ly/32F4gvc.

Turban, D. B., and Greening, D. W. (1997). Corporate social performance and organizational attractiveness to prospective employees. *Academy of Management Journal* **40**(3), 658–672 (https://doi.org/10.2307/257057).

UNFCCC (1998). Kyoto Protocol to the United Nations Framework Convention on Climate Change. Treaty Document, UNFCCC Secretariat, Bonn. URL: https://treaties.un.org/doc/source/RecentTexts/kyoto-en.htm.

Utz, S. (2018). Over-investment or risk mitigation? Corporate social responsibility in Asia-Pacific, Europe, Japan, and the United States. *Review of Financial Economics* **36**(2), 167–193 (https://doi.org/10.1016/j.rfe.2017.10.001).

Vazza, D., and Kraemer, N. (2017). 2016 annual global corporate default study and rating transitions. Research Insights, Standard & Poor's. URL: https://bit.ly/3eZUFTM.

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th edn. Springer (https://doi.org/10.1007/978-0-387-21706-2).

von Arx, U., and Ziegler, A. (2014). The effect of corporate social responsibility on stock performance: new evidence for the USA and Europe. *Quantitative Finance* **14**(6), 977–991 (https://doi.org/10.1080/14697688.2013.815796).

Welford, R. (2005). Corporate social responsibility in Europe, North America and Asia: 2004 survey results. *Journal of Corporate Citizenship* **17**, 33–52. URL: www.jstor.org/stable/jcorpciti.17.33.

World Economic Forum (2017). Global gender gap report 2017. Technical Report, WEF, Cologny-Geneva.