CISEF: a Composite Index of Social, Environmental and Financial Performance

Chrysovalantis Gaganis\textsuperscript{a}, Fotios Pasiouras\textsuperscript{b}, Menelaos Tasiou\textsuperscript{c}, and Constantin Zopounidis\textsuperscript{d,e}

\textsuperscript{a}Department of Economics, University of Crete, Rethymno, Greece
\textsuperscript{b}Montpellier Business School, Montpellier, France
\textsuperscript{c}Portsmouth Business School, Portsmouth, UK
\textsuperscript{d}School of Production Engineering & Management, Technical University of Crete, Crete, Chania, Greece
\textsuperscript{e}Audencia Business School, Nantes, France

Abstract

We propose a holistic evaluation framework that banks could adopt when screening corporate entities. The framework is based on the development of a composite indicator of social, environmental and financial performance. This integrated view of evaluation is conceptually aligned with ISO standards and empirical proposals of academics and market practitioners in defining inclusive performance. We complement those proposals with a methodological framework that permits incorporation of a plethora of viewpoints in the evaluation process, reflecting the expectations of the various stakeholders in the environment of a bank. We further enhance the evaluation process with analytics of a more detailed hierarchical view of performance, fine-tuning of the evaluation process, and provide implications and suggestions to the senior executive team of a bank’s clients.

Keywords: Multiple-Criteria Analysis · Holistic Firm Evaluation · Composite Indicators.

Corresponding Author: Constantin Zopounidis
E: kostas@dpem.tuc.gr; Tel.: +30 28210 37236;
Mailing Address: School of Production Engineering and Management,
Technical University of Crete, 73100, Chania, Greece.

Chrysovalantis Gaganis
E: c.gaganis@uoc.gr

Fotios Pasiouras
E: f.pasiouras@montpellier-bs.com

Menelaos Tasiou
E: menelaos.tasiou@port.ac.uk
1 Introduction

Banks have a multi-faceted functionality that in its most simplistic form can be thought of as a way of bridging depositors and borrowers, channelling capital deposited by the former into a type of loan extended to the latter. When the borrower is a corporate entity, a bank adopts a “special” role, that of a ‘corporate quasi-insider’ (Thomas and Wang, 2004). That is, a bank becomes an agent possessing inside information during the screening process of a company, which takes place prior to the release of any loan. Indeed, as Goss and Roberts (2011, p.1795) put it: “banks are given access to information about the firm that may not be available to outsiders”, which gives them great powers in a monitoring process, even to the capacity of monitoring a firm’s corporate governance -at least for bank-dependent firms (Ahn and Choi, 2009). The evaluation framework used by the bank is crucial, as it naturally affects its risk-return portfolio. In an evaluation framework, however, what often goes unnoticed is that several stakeholders in the environment of a bank are concerned one way or another, and they inherently affect the bank’s evaluation perspective. Understanding and taking into account the stakeholders’ objectives has implications that touch upon both conceptual and methodological grounds of an evaluation framework, which -as we will forthwith explain- is increasingly under the spotlight of regulators. To this end, this paper presents an evaluation framework that is inclusive from two perspectives: stakeholders’ objectives and their representation in the framework.

It is easy to understand the complexity behind an evaluation framework if one realises the multiple agents associated with the operations of the typical banking institution. First, there is the inner environment of a bank; depositors, who want security and a small premium to deposit their capital. These provide the necessary resources for a bank to start its intrinsic processes. At the same time, shareholders invest in a bank, influence its objectives and expect a flow of profits in return. On the other side of the equation lie regulators, who need to ensure no wrongdoings and a manageable risk that creates sustainable growth and does not trigger systemic incidents. And of course, the general society in which a bank operates, including individuals and firms looking to borrow capital to finance their needs or projects. This environment gets more complicated if one considers agency conflicts between these stakeholder groups. One can thus imagine that, from a bank’s perspective, its objectives and the evaluation of entities it services must reflect or respect its stakeholders’ views, and in such a mix that the bank deems realistic, desirable, and of course profitable. In other words, the questions a bank has to answer when it develops its evaluation framework could be: what would each stakeholder want and thus focus on, and whose view is to be reflected on the evaluation framework?

There is no simple answer given the complexity and variability behind different stakeholders’ objectives. From a purely unidimensional and simplified -yet mainstream- point of view, a bank could solely screen an entity according to its financial fundamentals. This would ensure capacity to repay the loan on time and set a respective price according to the underlying risk. After all, “for most entities, a good borrower is simply the one who pays back his loans” (Gutiérrez-Nieto et al., 2016, p. 691). However, this objective could fail a portion of the bank’s stakeholders. Depositors could be interested in the ethical integrity behind the lending process (Cowton, 2002; San-Jose et al., 2011), highlighting the need for responsible lending (Gutiérrez-Nieto et al., 2016), which strengthens trustworthiness between

2
depositors and banks (Cowton, 2002); that is a very important feature in banking (Jansen et al., 2015; Fungáčová et al., 2019) given the systemic incidents potentially triggered in the lack thereof. The need for responsible lending regards also policymakers whom, after the pressure of the wider public criticising integrity/responsibility issues could confront banks one way or another (Cowton, 2002). In fact, recent initiatives such as the 2019 recommendations of the Central Banks and Supervisors Network for Greening the Financial System (NGFS) and the 2018 EU Action Plan promote the inclusion of climate risks into the financial institutions’ risk management policies and the potential integration of climate risk into prudential regulations and capital requirements. In addition to that note, whilst some investors may share the ‘Friedman doctrine’\(^1\), many others have also ethical and social concerns (Hart and Zingales, 2017), particularly those investing in the long-run (Schoenmaker and Schramade, 2019; Edmans, 2011, 2012).

Therefore, an assessment framework that will consider simultaneously the financial creditworthiness and the socio-environmental performance of the borrowers could be of substantial value to bank managers. To this end, a bank could augment its unidimensional (financial-centered) evaluation framework with attributes related to corporate social responsibility (CSR)\(^2\), which has gained considerable traction in recent years. For example, there have been initiatives, such as the United Nations’ Environmental Programme ‘Finance Initiative’ (UNEP FI)\(^3\), or the ‘Equator Principles’ (EP)\(^4\) on which a few banks around the world signed on voluntarily, promoting more inclusive -from the viewpoint of criteri- evaluations. From a reputation perspective, this could be used by banks to signal responsible conduct (Scholtens and Dam, 2007).

All the above become even more important in the light of recent regulatory initiatives in the banking sector, for example by the Bank of England (2019) and the European Banking Authority (2019), aiming to introduce stress-tests to assess the impact of climate risk on banking institutions. While it appears that the Central Banks have not yet decided to ask additional capital related to climate risk, this does not necessarily mean that it will not happen in the near future, since there are ongoing discussions and references to this issue (European Commission, 2018; CPR II/CRD V; European Banking Authority, 2019). Hence, it might not be long into the future until banks that do not pass climate risk tests will be asked to hold more capital or remove certain assets from their portfolios to lower their exposure to climate change. This will have implications for banks and the selection of corporate customers as it will become more expensive to finance climate-damaging activity. Finally, regulatory standards and expectations aside, another concrete benefit for banks adopting a holistic such framework would stem from their advisory role in offering customized and industry-specific specific advisory services to their clients to improve

---

\(^1\) Based on the influential article of Friedman (1970) that companies should only focus on maximising value (see Macintosh, 1999, for a review of the “Berle-Dodd” debate started in the 1930s, based on which views Friedman argued in favour of the former).

\(^2\) There is no single definition of CSR (see Heal, 2005, p.393 for a discussion), although a general notion that we adopt in this study is that given in Ferrel et al. (2016, p. 585) that firms are socially responsible when they “promote efforts to help protect the environment, seek social equality, and improve community relations”.

\(^3\) This was focused on encouraging the implementation of sustainability into financial institutions’ relationships and trade with their customers (see UNEP, 1992).

\(^4\) An initiative based on a set of safeguarding policies by the International Finance Corporation (IFC), the World Bank's private sector financing division, aimed at identifying the magnitude of risk (classified in a scale of “A”, for high, to “C”, for low risk) of the social and environmental impact a project financed by the bank could have accordingly (see Wright, 2012).
their inclusive performance.

There have been several attempts in the past to create firm evaluation frameworks that are more holistic -in terms of inclusive criteria-, and we name a few here. From the viewpoint of financial markets, one of the most well-known and oldest initiatives adopting a composite indicator perspective is the Dow Jones Indices (DJI) ESG (which stands for Environmental, Social and Governance) score (S&P, 2019). This involves ESG related criteria on which publicly listed firms are benchmarked, and the top 75% form an alternative stock market index weighted by their market cap (‘Dow Jone Sustainability Index’; DJSI). However, both the index criteria and the parameters used to construct these ESG scores are opaque, with only information on the participating companies being shared. In the academic arena the studies of, among others, Oliveira et al. (2019); Puggioni and Stefanou (2019); Ortas et al. (2015); Cheung et al. (2010); Cai et al. (2011); Hillman and Keim (2001) use or develop a sustainability-related composite indicator (explicitly or implicitly). These indices, however, focus on performance related only to CSR aspects (thereby entirely neglecting financial performance\(^5\)), on either a single industry, or a single dimension of responsibility (e.g. the environmental or social). The closest study to our proposed theoretical framework is that of Gutiérrez-Nieto et al. (2016), who develop a credit score system for financial entities financing social projects. The novelty that we will present here, however, is posited on the methodological framework (to be presented in Section 2), which does not only present an inclusive evaluation (i.e. taking into account also non-financial criteria), but also takes into account the multiplicity of stakeholders in a bank’s environment. That is, even if a bank indeed adopts a more inclusive socio-economic framework to evaluate and screen entities based on, whose opinion should be incorporated in the evaluation? In our framework, conflicting (from a principal-agent viewpoint) shareholders’ views are taken into account in the aggregation process to reach a holistic evaluation that incorporates a multiplicity of viewpoints from a balanced growth view.

This study thus proposes an evaluation framework that measures firm performance, based on which banks may obtain ordinal (if ranking or choice problems are concerned) or cardinal (if relative or historic benchmarking in terms of magnitude is concerned) information on the relative position of their clients that can be used in their lending or investment processes. In particular, in terms of the theoretical framework, we adopt the view that CSR is enriching for firm performance, thus the adopted set of criteria is enlarged -compared at least to a unidimensional framework solely relying on financial fundamentals- to include environmental and social criteria. Whilst this may be in contrast with proponents of the value maximisation view, studies show that banks perhaps indeed value CSR, as firms with superior CSR performance are associated with inferior cost of bank loans (Goss and Roberts, 2011; Nandy and Lodh, 2012). In terms of the methodological framework, in order to take into account uncertainty, ill-determination in parameters and issues of representation of preferences or conflicting shareholders’ views involved in the evaluation process, we make use of a recently proposed method, namely the ‘Sigma-Mu efficiency analysis’ (Greco et al., 2019b). We further show how this method can be augmented to provide information tailored to the evaluation framework in question (from the viewpoint of the bank) or the client’s senior management.

\(^5\)Exception is the study of Oliveira et al. (2019), who include an indicator reflecting financial performance, albeit including only a profitability indicator (EBITDA).
In the remainder of this study, Section 2 discusses the proposed framework both from a conceptual and methodological point of view, the analyses and implications it may have for banks and the management of its clients, as well as close alternatives to the proposed framework. Section 3 presents a didactic illustration of the proposed framework, and Section 4 concludes the study.

2 Theoretical & methodological framework

2.1 Theoretical framework

2.1.1 Background discussion

As briefly discussed in the outset of this study, we adopt a ‘multidimensional’ view (we will interchangeably refer to this as ‘holistic’ or ‘integrated’ in line with terminology commonly referred to in the literature) of firm evaluation, in which not only financial fundamentals enter the evaluation framework, but also criteria related to corporate social responsibility. It is very difficult to define corporate social responsibility without being subjective or somewhat arbitrary, given the existence of no single official definition (Heal, 2005). However, the International Organization for Standardization (ISO) in co-operation with the Global Reporting Initiative (GRI) created a document (see ISO, 2010) drawing on international consensus from the broadest view of stakeholder groups, in order to ensure that the concept of evaluation of CSR is standardised, taking into account firm performance as a function of three dimensions: economic, environmental and social. According to the GRI, 75% of the largest 250 companies in the globe use the GRI principles (GRI, 2019). This is a deeply welcoming fact, and a great initiative from a firm perspective. At the same time over 100 financial institutions worldwide have currently signed up with the Equator Principles, which enhance the monitoring process of banks when screening the financing of firm projects with responsibility criteria. These two facts together show that both parties -i.e. banks and firms- start aligning with the notion of a more holistic evaluation, or at least its potential given the enlarged amount of information in firm disclosure.

As Oliveira et al. (2019) mention, despite the prominence of CSR in several literature strands and the existence of standardised frameworks, there appears to be a confusion when it comes to the broader interpretation of organisational outcomes. This is natural, as the existence of a variety of reporting frameworks, whilst very helpful and a welcoming initiative, does not provide any guidance as to a potential way of evaluation. A decision-maker is still confronted with a variety of information at hand, potentially more enlarged than ever. This is where the field of composite indicators steps in (for a detailed discussion, see Greco et al., 2019a).

Composite indicators are one of the most popular means used by academics, practitioners and policymakers to consolidate a multidimensional aspect of interest into a single metric (OECD, 2008). As their name suggests, they represent a synthesis of a set of elementary criteria to form a single metric, which acts as an estimate evaluation of an entity’s aspect, e.g. be it risk, performance etc. This is one of their main benefits (OECD, 2008) and a likely reason behind their popularity skyrocketing in recent times (Greco et al., 2019a). In this paper, we posit an evaluation framework that can be used by banks to screen companies from a multidimensional point of view, whilst at the same time taking into
account the multiplicity of stakeholders in its inner and outer environment. To do this, we develop a composite indicator of social, environmental and financial performance (CISEF). The procedure for the development of a composite indicator consist of two main parts: the theoretical, and the methodological framework. The remaining of this section explains how these two can be approached by a prospective bank in evaluating its corporate clients.

2.1.2 Theoretical framework

In the multidimensional framework of evaluation that we adopt, a corporate entity is screened on the basis of financial fundamentals and social responsibility criteria. In line with the ISO standards (ISO, 2010), our evaluation framework is split into 3 dimensions (visualised in Fig. 1), reflecting a hierarchy of criteria related to social, environmental and financial aspects. In a nutshell, the social and environmental dimensions -augmenting the unidimensional view of economic performance- encapsulate information on a firm’s practices of decent work and its contribution to the local development, as well as its considerate use of natural resources and dispersion of waste accordingly (Oliveira et al., 2019). Let us note here that, this is well in line with Ferrel et al.'s (2016) notion of CSR that we also adopt in this study. Moreover, the hierarchy of an integrated performance split into environmental, financial and social attributes is also in line with Schoenmaker and Schramade's (2019) thoughts of an ideal long-run evaluation framework shareholders should look at (what they term as a ‘new paradigm’) instead of a myopic short-term unidimensional view.

In regard to the nested criteria within each dimension, these arguably change depending on the type of agent adopting this evaluation framework. For instance, being tied to the responsibility principles, under the umbrella ‘Economic’, ISO (2010) contains clauses which are vaguely defined and are related to performance, presence in the market, economic impacts etc. An entirely different set of criteria could be looked into by banks, which could focus on a set of financial strength ratios, or by investors, who could focus on a set of financial market performance ratios. Indeed, all previously developed studies using a similar evaluation framework actually follow a variety of different criteria -attributed to the unavailability of data, or the convenient supply of frameworks, and hence data, from third party data providers (data bases)- to proxy the attributes in question. Moreover, different banks could have access to a variety of different criteria according to the types of clients they have, as well as due to standards attributed to country regulations. As the selection of indicators extends beyond the scope of this study, we refrain from being specific at a theoretical framework level (although we will specify a set of criteria that we use ourselves later in the evaluation phase), and we leave this to the expertise or experience of analysts using the outline of this framework.

6We would like to thank an anonymous reviewer for suggesting that corporate governance indicators should possibly be included as additional criteria to form a composite index of social, environmental, governance, and financial performance. Taking into account this comment, along with regulatory and practitioners’ references to ESG ratings, in additional analysis that is given as supplementary appendix, we re-estimate our composite index with the addition of corporate governance indicators. The Pearson’s correlation coefficient between the two composite indices (i.e. with and without the corporate governance indicators) equals 0.88.
As it will become more intuitive in the methodological framework (Section 2.2.1), in developing a composite indicator, choices have to be made at each and every stage, with the most pernicious options being in the weighting and aggregation stages (Greco et al., 2017a). The theoretical framework mainly regards the former, which revolves around deciding on who is the expert (if weights are subjectively assigned), and what do the weights represent (i.e. importance coefficients or trade-offs), the latter naturally pointing to an aggregation function that accommodates this option.

The common practice (and that one found in all previous integrated evaluation frameworks) is the use of a single weight-vector (commonly equal weights), decided by a single decision-maker (or a group of decision-makers that is instead summarised via means of an average, assuming a representative agent scenario). However, we argue that this is a somewhat arbitrary -perhaps authoritative- scenario, which places huge emphasis on who the decision-maker is and reveals a few conceptual issues with underlying assumptions unwittingly being made as a consequence. For instance, even if one is able to define a single decision-maker, a more pernicious question at this stage is whether that decision-maker is representative of the potential pool of other decision-makers; or even for herself.7

To give an example, one can imagine a single or a set of analysts declaring a set of preferences (say, weights), each element in which should be well-defined (i.e. precise and certain) and reflect other potential analysts or stakeholders the analysis could regard. With respect to the alleged existence of a well-defined vector of preferences, realistically, one should expect uncertainty to be back-tested for robustness purposes, e.g. with preferences being oscillated (Freudenberg, 2003), or preferences being expressed in an interval form from the very beginning. With respect to the representation issue, the set of parameters expressed by the analyst is realistically not representative of other stakeholder groups. For instance, imagine the existence of a well-defined representative set of preferences for shareholders. Do these necessarily align with the equivalent set of depositors, or regulators?

In our proposed framework, instead of using a single set of preferences in the evaluation phase, we adopt the view that a bank can amalgamate a plethora of viewpoints for each stakeholder group into the end metric. This is the novelty of our proposed framework (visualised in Fig.2), which we argue that it brings a more holistic bank evaluation framework. In Section 2.2.1, we show how this conceptual framework could be reflected in the methodological framework.

7As mentioned in Greco et al. (2019b), in the decision on a set of preferences, it is perhaps more realistic to encapsulate a set of parameter vectors -instead of a single one-, in order to account for uncertainty and ill-determination of preferences, or the fact that one can be seen as a multiplicity of shelves of which is composed Elster (1987).
2.2 Methodological framework

2.2.1 Developing a CISEF

The computational procedure of a composite index is essentially a means of aggregation, which can be seen as a function of elementary criteria and a set of preferences. The most common aggregation function in the development of composite indicators is that of a weighted arithmetic or geometric mean (see Greco et al., 2019a, for a review). Refraining from making a choice between the two at this stage, we give a general form of a quasi-arithmetic mean (Grabisch et al., 2009), i.e.:

\[
CI_i(x_i, w) = f^{-1}\left(\sum_{j=1}^{n} w_j f(x_{ij})\right),
\]

which boils down to the arithmetic weighted mean for \(f(x) = x\), and the geometric weighted mean for \(f(x) = \log x\), given that \(x_{ij} \in \mathbb{R}\) or \(x_{ij} \in \mathbb{R}^+\) accordingly. In terms of notation, \(CI\) annotates the developed composite indicator, \(I = \{1, \ldots, n\}\) is the set of firms to be evaluated by a bank, according to the set of dimensions \(J = \{1, \ldots, m\}\), the normalised values of which are annotated by \(x_{ij}\) (we elaborate on this further in this section). To each dimension \(j \in J\), a coefficient \(w_j\) is attached -denoting the importance of criterion \(j\)- (to be discussed further in this section), which in such compensatory aggregation settings takes the form of a trade-off (see Decancq and Lugo, 2013, for a discussion on the decomposition and meaning of trade-offs). Without loss of generality, a time dimension \(t \in T\) can be added in the case of panel data. Moreover, the cross sectional dimension \((i \in I)\) can be treated independently for each industry if that is desirable (we elaborate on this in Section 3). For reasons of simplicity, the provided notation describes the evaluation taking into a count a single industry, although, without loss of generality, the industry-specific set of firms can be annotated by \(I_d\), \(d \in D\) being the industry indexation.

Let us note here that other aggregation functions can also be used to synthesise a composite indicator. For instance, Greco et al. (2020) make use of aggregation functions stemming from methods such as the
PROMETHEE family (Brans and De Smet, 2016), which draws on non-compensatory principles. In this study, for simplicity, we will make use of the weighted arithmetic mean (hence \( f(x) = x \) in equation 1), being the most commonly encountered aggregation function (Greco et al., 2019a).

Any quasi-arithmetic aggregation function requires normalisation of the elementary indicators (OECD, 2008). We follow the normalisation procedure used for the ESG scoring indicators in the DJI (S&P, 2019). In particular, assuming a unipolar direction in a criterion that the larger it is the better, the normalised indicators \( x_{ij} \) can be obtained as follows:

\[
x_{ij} = G\left( \frac{z_{ij} - \bar{z}_j}{s_{z_j}} \right),
\]

(2)

where \( z_{ij} \) is the raw value of firm \( i \) on criterion \( j \), \( s_{z_j} \) is the standard deviation on criterion \( j \) and \( G(y) = \frac{2}{1 + \exp(-y)} - 1 \) is the sigmoid function that smoothens the normalised indicators in the \((-1, 1)\) range. The smoothening function \( (G(z)) \) eliminates extreme values, and the reason of the binding \((-1, 1)\) range is related to the normalisation of the final aggregation making use of the cumulative density function (see S&P, 2019, for a more detailed description of the DJI’s ESG scores methodological annex). In particular, following S&P (2019), when the evaluation of a firm is accomplished, the composite indicators (eq.1) are standardised (like in eq.2) and a probability score is calculated within industry and year, defined as the normal cumulative density function (with mean equal to 0 and standard deviation equal to 1), i.e.:

\[
P_{it} = F_{t}(\frac{CI_{it} - \overline{CI}_t}{s_{CI_t}}).
\]

(3)

This represents the percentage of an infinitely large population than each evaluated firm would be better.

In the evaluation framework presented up to this point, if one assumes a single decision-maker that is representative of all concerned parties, and a precise, well-defined set of parameters (\( \mathbf{w} \)); the evaluation framework is accomplished and \( P_{it} \) is the evaluation of firm \( i \) in year \( t \). In our proposed framework, the whole feasible space of weight vectors \( \mathbf{W} \) is taken into account to incorporate all potential stakeholders’ viewpoints into a single metric that acts as a holistic performance. To achieve this, we base this proposal on the methodological framework proposed by Greco et al. (2019b).

In particular, using the Stochastic Multiobjective Acceptability Analysis (SMAA) (Lahdelma et al., 1998; Lahdelma and Salminen, 2001), one may consider a probability distribution \( f_w \) over the space of all feasible weight vectors

\[
\mathbf{W} = \left\{ \mathbf{w} = [w_1, \ldots, w_m] : w_j \geq 0, \sum_{j \in d} w_j = 1 \right\},
\]

which, in the lack of information from stakeholders, remains unconditional, with uniform draws from this space conducted in the modelling phase. If, however, any information is provided in the form of ordered preferences, e.g. that dimension 1 is more important than dimension 2 and so on, the space of feasible weights can be transformed to a constrained example as follows:

\[
\mathbf{W}^c = \left\{ \mathbf{w} = [w_1, \ldots, w_m] : w_{j(1)} \geq w_{j(2)} \geq \ldots \geq w_{j(m)}, \sum_{j \in d} w_j = 1 \right\}.
\]
To give an example, based on prior expertise or observation, an analyst could declare that shareholders may be more interested in the financial performance rather than the social and/or the environmental, whereas the policymaker could be more interested in the environmental over the others, and so on. Moreover, in the presence of information from stakeholders or analysts, weight need not be precise, and they could for example be drawn on the basis of an interval, e.g. that $0.2 \leq w_j \leq 0.4$ (always subject to $\sum_{j \in J} w_j = 1$) to reflect uncertainty or a more diverse audience within and across each stakeholder group.

In a realistic concept, a bank could either survey its stakeholders to obtain a set of vector parameters (one per each surveyee), or simulate the space $W$ according to the analysts’ prior expertise. Understandably, for every $\mathbf{w} \in W$ the evaluation $CI_{it}$ may radically change, which can be thought of as a stakeholder-specific evaluation. SMAA permits consolidation of this evaluation of a multiplicity of stakeholders in an ordinal form, by considering probabilistic indicators, such as what is the probability an evaluated firm is positioned first, second and so on (what is called as ‘rank acceptability index’). However, that may be far from desirable in a composite indicator evaluation for two reasons. First, probabilities may be unacceptably small, particularly as the number of evaluated firms increases. Second, composite indicators are cardinal in nature (Booysen, 2002) and permit a more transparent comparison by showing the magnitude of a performance. To consolidate this depth of information into a single metric, drawing from the notion of the Markowitz mean-variance analysis framework (Markowitz, 1952), Greco et al. (2019b) propose a similar type of framework, according to which the mean ($\mu$) of the obtained results reflect the typical evaluation taking into account all stakeholders’ views, and the standard deviation ($\sigma$) reflects the typical deviance from this evaluation. To put this into our context, one could imagine $\sigma$ as a measure of inverse type of robustness, according to which a unit with a high value of $\sigma$ only satisfies a very small portion of stakeholders, and which should thus be taken into account and be minimised by the firm.

Assuming equal importance among all stakeholders, these two parameters are calculated as follows:

$$\mu_{it} = \int_{\mathbf{w} \in W} f_{\mathbf{w}}(\mathbf{w}) CI(x_{it}, \mathbf{w}) d\mathbf{w},$$  \hspace{1cm} (4)

$$\sigma_{it} = \sqrt{\int_{\mathbf{w} \in W} f_{\mathbf{w}}(\mathbf{w}) [CI(x_{it}, \mathbf{w}) - \mu_{it}]^2 d\mathbf{w}.}$$  \hspace{1cm} (5)

These two parameters are then fused into an efficiency score ($\delta$) that can be thought of as a ‘Benefit of the Doubt’ (BoD) approach (Melyn and Moesen, 1991; Cherchye et al., 2004, 2007) into what Greco et al. (2019b) term as the ‘sigma-mu plane’ - a scatter plot visualising the evaluated firms in question-, where $\mu_{it}$ is supposed to be maximised - as larger values denote better performance - and $\sigma_{it}$ is supposed to be minimised - as lower values denote more stable performance and thus satisfying a larger portion of stakeholders equally-. From the viewpoint of data envelopment analysis, this can be considered as a CCE model (Charnes et al., 1978) with unitary inputs, and with $\sigma_{it}$ being a type of ‘bad’ output (Cherchye et al., 2015). This is computed as follows:

---

8 Indeed, one could imagine a scenario according to which a bank could weigh one or certain stakeholders or groups of stakeholders higher than others.
$\delta_{it}^* = \max_\delta$
\[
\begin{aligned}
\text{s.t.} \\
\begin{cases}
\alpha_t \mu_{it} - \beta_t \sigma_{it} \geq \alpha_t \mu'_{it} - \beta_t \sigma'_{it} + \delta, \forall i \neq i', \text{with } i, i' \in I \\
\alpha_t, \beta_t \geq 0 \\
\alpha_t + \beta_t = 1
\end{cases}
\end{aligned}
\] (6)

What Greco et al. (2019b) term as a ‘local’ efficiency score, $\delta_{it}$ can be thought of as an inequality (in terms of how it satisfies the plurality of actors) adjusted performance on the basis of an (ideally) balanced-firm performance. Let us note here that this is conceptual similar to risk-adjusted performance ratios, such as the Sharpe ratio, although instead of a constant adjustment that is equal across all units, one gives the benefit of the doubt to all evaluated firms, by permitting them to choose their own $\mu - \sigma$ adjustment in such a way that maximises their scores. This is one of the reasons making BoD approaches very popular with decision-makers (Cherchye et al., 2007; OECD, 2008). Greco et al. (2019b) augment this with spatial information from the ‘sigma-mu’ plane. In particular, drawing from the ‘context-dependent’ analysis proposal of Seiford and Zhu (2003), it is possible to reveal a plethora of Pareto-Koopmans frontiers (PKF) in the sigma-mu plane, each of which can be thought of as a different level of competition. Considering the solution of (6) consists the first PKF, i.e. $PKF_1$, the $p$ PKF frontiers can be decomposed and collected in $PKF = \{PKF_1, PKF_2, \ldots, PKF_p\}$, with each unit $i \in I$ being benchmarked against all set of frontiers. In particular, for a unit $i$, the $k$-th frontier is calculated as follows:

$$
\delta_{itk}^* = \max_\delta
\begin{aligned}
\text{s.t.} \\
\begin{cases}
\alpha_t \mu_{it} - \beta_t \sigma_{it} \geq \alpha_t \mu'_{it} - \beta_t \sigma'_{it} + \delta, \forall i \neq i', \text{with } i, i' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h \\
\alpha_t, \beta_t \geq 0 \\
\alpha_t + \beta_t = 1
\end{cases}
\end{aligned}
\] (7)

Greco et al. (2019b) propose to aggregate for each unit this spatial information from the local efficiency scores (i.e. $\delta_{itk}$) into what they term as an overall ‘holistic’ evaluation ($sm_{it}$) that takes into account both the inequality (from stakeholders) adjusted measure of performance, and the spatial information of dominance and domination in the ‘sigma-mu’ plane. This is easily calculated as:

$$
sm_{it} = \sum_{h=1}^{p} \delta_{itk}
$$ (8)

To keep the consistency, we normalise the overall holistic scores as we did with the elementary criteria, i.e. by reporting the probability score defined as the normal cumulative density function of the standardised scores (as in eq.3).
2.2.2 Analytics, further analyses & potential managerial implications

The holistic evaluation metric computed in this process (eq. 8) can be used by the bank in several ways, permitting to delve into more detail in the evaluation framework, or simply to advice prospective client(s). We show a few examples in this Section, although these are far from exhaustive.

First, a composite indicator can be used for business intelligence (Barone et al., 2011). For instance, the bank can use the computed metric as a scorecard, in which an analyst may observe a cardinal (e.g. what is the estimate of the evaluation of a unit in terms of magnitude), or ordinal type of information (e.g. a ranking). However, in a hierarchical framework such as the one considered in this case (Fig. 1), it is difficult to know where exactly an exceptional or poor performance is attributed to. In the case of a single dimension, one can easily disaggregate the synthetic index back to its components and look at which dimensions can be improved. However, in a complex hierarchy consisted of many levels, this is somewhat more difficult; particularly considering the volume of preferences and spatial information incorporated in the computed metric. In this regard, the methodological framework presented in Section 2.2.1 could be augmented by coupling it with the Multiple Criteria Hierarchy Process (MCHP) (see Corrente et al., 2012), which is a well-appreciated method in the Multiple Criteria Decision Aiding (MCDA) domain (see e.g., among others, Angilella et al., 2013, 2018; Corrente et al., 2016, 2019; Arcidiacono et al., 2018). This can be easily done by applying the above methodological framework in each node of the hierarchy, i.e. including criteria specific to only a single dimension, e.g. partitioning the set $J$ into the three dimensions $J^s$, $J^e$ and $J^f$, and each subset can be partitioned into further sub-dimensions (if applicable).\footnote{In our considered case study, due to unavailability of data there are no sub-dimensions. Therefore the first level of hierarchy consists of the three dimensions, financial, social and environmental, and the second level simply consists of the normalised sub-indicators.}

Second, the evaluation process can be fine-tuned according to hypotheses being made on the basis of the analysts’ expertise. For instance, the evaluation presented up to this point assumes that each firm performs poorly and creates a bottleneck, relative to its counterparts, and provide more detailed recommendations/advice respectively.

![Figure 3. Hierarchical evaluation of CISEF](image)

This will provide for each firm a set of scores -aside from the overall- that show how each dimension or sub-dimension is evaluated. These can be used to understand in which component of the hierarchy a firm performs poorly and creates a bottleneck, relative to its counterparts, and provide more detailed recommendations/advice respectively.
firm $i$ is evaluated against a pool of firms for which there is data availability (in our case) or that are a bank’s clients. This renders the estimated metrics -up to a certain extent- conditional on the sample at hand. For example, a different sample could change the results due to the steps of both normalisation and evaluation taking into account the spatial information of the formed frontiers in the sigma-mu plane. This is reasonably encountered, perhaps being the norm, not only in the field of composite indicators, but in any evaluation framework. In terms of the sigma-mu plane, this can be -up to a certain extent- alleviated making use of robust frontier techniques\textsuperscript{10}, although one may argue that it is still bound to suffer from different estimators or rank reversals. If that is not acceptable, we propose that the evaluation framework could be adjusted as follows. Each unit $i \in I$ could be benchmarked against a fixed set of units $V = \{1, \ldots, v\}$ that could be decided by bank analysts. E.g. it could be a variety of ideal -such as industry leading firms- or anti-ideal -such as industry laggard firms- examples, chosen on the basis of analysts’ expertise and/or industry-wide commonly accepted examples.

In particular, let’s assume that analysts define a set $V$ consisted of three units, reflecting an ideal, an ordinary and an anti-ideal (poor) performance accordingly. These could be real, existing firms, or utopic ones created based on analysts’ beliefs. Understandably, it could be many more alternatives, or different types of alternatives, each tailored to the expertise and needs of an analyst. The evaluation presented up to this point changes as follows. For each alternative firm $a_i \in A, i \in I$, a new set $Z^i$ is formed, with $Z^i = V \cup \{a_i\}$. Then, the methodological framework is conducted $n$ times, one for each alternative $a_i$, extracting each time only the evaluation of the alternative of interest. Let us note here that this is in line with the notion of central profiles in sorting MCDA models (Zopounidis and Doumpos, 2002).

Last, but not least, as composite indicators can be fairly considered black boxes in terms of how the weights are actually transformed in the final metric due to correlation effects (see e.g. Paruolo et al., 2013; Becker et al., 2017, for an interesting discussion and a measure of estimating the actual weights in the final construct), it makes it rather difficult suggesting improvements on the potential effects changes could have, \textit{ceteris paribus}, on the final firm performance. This is exacerbated here given that the final construct not only involves a plethora of viewpoints (in the form of a balance-adjusted performance), but also spatial information. We thus argue that, in line with the steps followed by Paruolo et al. (2013); Becker et al. (2017) in the estimation of ‘main effects’, a good statistical approximation can be achieved through parametric or non-parametric regression models. Assuming a model of the former type, the following meaningful form of regression model can be evaluated:

$$CISEF_{it} = \theta' z_{it} + \gamma' Y_t + \eta' F_{it} + \epsilon_{it}, \quad (9)$$

where $z_{it}$ contains the raw values of the elementary evaluation criteria, $Y_t$ contains a set of dummy variables capturing year fixed effects and $F_{it}$ contains frontier fixed effects. $\theta$ and $\eta$ are $m \times 1$ size vectors and $\gamma$ is a $T \times 1$ vector of parameters estimated by the model, and $\epsilon_{it}$ is the idiosyncratic error. Let us note here that, as composite indicators by nature have ceiling and floor effects, it is best that they are estimated using censored models, see e.g. McDonald and Moffitt (1980) (for the use of Tobit in the case of parametric models) and Lewbel and Linton (2002) (in the case of non-parametric models).

\textsuperscript{10}See e.g. Cazals et al., (2002); Daraio and Simar, (2005), for a discussion and proposals on robust frontier techniques and Greco et al., (2019b), for a discussion on how the latter can be embedded in the sigma-mu framework.
2.3 Alternative methodological choices & the suitability of the proposed one

One may wonder at this point what the alternative methodological choices in the evaluation scheme are, and what exactly makes the proposed set of of choices -as given in this Section- suitable in such an evaluation framework. We should hereby note that this subsection only refers to methodological choices and not other choices related to the conceptual framework.

Let us begin with the two critical steps of metric development: weights and aggregation. As argued in the study of Greco et al. (2019b), the ‘sigma-mu’ method is not based on any \textit{a priori} assumptions on the function of weights, whilst, as seen in Greco et al. (2020), the same also holds for aggregation techniques. The point is that the choice of the weight/aggregation functions is in the hands of the metric developer to make and should not be seen as drawbacks to the proposed method. On the contrary, they could be seen as a benefit, as the OECD (2008) suggests the use of uncertainty analysis to accompany any composite index, which is hereby inherently encapsulated into the end result. Thus, one could use not only different stakeholders’ views - as we have argued in the first part of the paper-, but different weighting techniques to elicit those -including other data-driven techniques (see Greco et al., 2019a)- and the same holds for different aggregation functions. For instance, in a radical uncertainty notion, these could even be left to the simulation environment of SMAA to be chosen via trigger mechanisms identical to those being used in uncertainty analysis frameworks (see e.g. Saisana et al., 2005).

Furthermore, let us look at the very first requirement we have set in this study: inclusion of stakeholders’ views. We have mentioned that the framework needs to be inclusive of its stakeholders, otherwise ‘representative agent’ issues may arise (see e.g. Greco et al., 2019b, for a discussion). In the literature of composite indicators, this has been traditionally solved with models that rely on SMAA (Lahdelma et al., 1998; Lahdelma and Salminen, 2001), which include a space of preferences $W$ in the evaluation phase results of which can be aggregated and presented in an expected utility form (if cardinal results are to be obtained) or an expected rankings (if ordinal results are preferred instead) (see Greco et al., 2019a, for a literature review on composite indicators). A more advanced version of this modelling aspect could involve multilinear utility functions (MLUFs) that augment the plain aggregation of the additive utility function, from a purely exogenous manner to also being able to deal with higher order interactions and other aggregation functions (see e.g. Montiel and Bickel, 2014, for a very interesting paper on dealing with these aspects). To the best of our knowledge, these are the only two types of methods incorporating a multiplicity of elicited preferences (or general spaces of preferences) into a utility score acting as an evaluation metric. In an abstract way, both aforementioned methods relate (or can directly be reduced) to the first parameter in our proposed framework, i.e. the parameter $\mu$. Put simply, that parameter shows the expected evaluation taking into account a multiplicity of viewpoints in the evaluation phase (either if one includes or not higher-order terms). In the particular socio-economic framework that we presented here, however, this is only one aspect that we deem important, which takes us to the second important feature: \textit{imbalance}.

Imbalance is captured by the second parameter in our proposed framework, i.e. $\sigma$, which is a measure of inverse robustness that shows the dispersion of the typical evaluation, and this has a twofold meaning. Firstly, that shows how imbalanced an evaluation could be when different stakeholders’ views were to be selected. For instance, a very high $\sigma$ could mean that different stakeholders could be satisfied (proxied by
the utility score) in a completely different manner from the same company due to them being focused on a different (sub-)set of attributes respectively. This parameter is of paramount importance in our framework as it goes on to show that a bank not only considers the views of all its stakeholders, but it also satisfies them in the best possible way; that is, by penalising a firm’s evaluation if it satisfies its multiple agents differently, which takes us to the second meaning of \( \sigma \); a proxy of an imbalanced firm profile. This is in line with the concept of a balanced firm as argued by Enderle and Tavis (1998). Quoting from their study: “The concept of the [balanced] firm [...] consists in balancing economic, social and environmental responsibilities and thus implies a circular interrelationship among them. No one of these realms of responsibility can be purely instrumentalized in favor of others [...]. This balanced concept should help to bridge the gap between internal and external assessment of companies”. In fact, Enderle and Tavis (1998) present this argument of ‘circular interrelationship’ in Fig. 1 (their Fig. 4, p. 1336) to show how a company could excel in any aspect of responsibility (economic: firm A; environmental: firm B; social: firm C).

Figure 1: Dimensional trade-offs and balanced growth (from Enderle and Tavis, 1998, p.1336)

Understandably, a firm manager is faced with many trade-offs when managing the different types of responsibilities at her disposal, and this is reflected in the firm's performance. Bringing this into our context, when that firm is to be evaluated, depending on which stakeholder’s view is reflected on the evaluation, the firm could be either over or under-represented in this evaluation. A balanced firm, however, will always keep its evaluation score intact. In other words, if an inclusive evaluation is not only about sheer uni-dimensional (predominantly financial one) development but about one that is balanced; then, by definition the evaluation will satisfy not only the firm’s inner and outer environment, but also the bank’s one when it evaluates the said firm - and that includes all stakeholder groups taken into account in this process. This is captured by the parameter \( \sigma \) in the proposed framework, which complements \( \mu \) in the evaluation by 'judging’ a firm not only by its ‘absolute’ expected performance (i.e. \( \mu \), taking into account all stakeholders’ views) but also by the imbalance it hides under that score or, put simply, by the inequality that exists as to which portion of stakeholders is satisfied by this firm in this socio-economic framework. In this case, a firm that systematically (i.e. across all dimensions being examined) performs
excellent is the one that achieves a top score. And the trade-off (i.e. between \( \mu \) and \( \sigma \)) is optimally being calculated in such a manner that gives the most rewarding evaluation to each firm in the notion of a mean-variance framework.

Of course, we should hereby note that, if one is not interested in adjusting the evaluation of a firm’s performance also by its variability, the framework could simply revert back to the single parameter \( \mu \), i.e. obtained using only the SMAA framework (Lahdelma et al., 1998); or, if a more advanced and general framework that takes into account also different types of aggregation and interactions among attributes, to revert back to the MLUFs (Montiel and Bickel, 2014).

Finally, as metrics are only as good as the assumptions used to develop them (otherwise ‘garbage-in garbage-out’ outcomes may arise; see Greco et al., 2019a, p.67, for a discussion on the plurality of weighting systems); it is important to note a few more things on the validity of the current model, and an alternative variation in the step of aggregation (eq.1) that could possibly be followed by practitioners applying this framework. Firstly, to the best of our knowledge, there is no study in the domain of composite indicators dealing with ‘endogeneity’ in its traditional meaning as this is used in the field of Econometrics (Wooldridge, 2010). That is because a composite index is the sum of its parts, (i.e. the elementary indicators aggregated to construct it). So there is no ‘error’ component (residual) in developing a composite index similar to a regression model. An issue that may arise, however, during the aggregation process relates to how correlation can lead to the alteration of nominal weights in the actual index (see Paruolo et al., 2013, for a very interesting study on measuring these effects). Despite correlation among our indicators being weak to moderate, if one wants to isolate correlation effects in the aggregation step, there are proposals using regression models and optimising algorithms to isolate such effects and try to find the optimal nominal weights that translate to the desired (declared by a DM) weights once the aggregation has taken place (see Becker et al., 2017, for a very interesting proposal). In our model, this would mean that, for every simulation, the vector of preferences \( \mathbf{w} \) in the space \( W \) had to be analysed using their proposed model, which, for an illustrative study like this, it would over-complicate the analysis, it would be computationally very demanding, and, for a portion of the simulated vectors, it could even be infeasible to obtain strictly positive weights. In a real world case study, particularly where correlations among indicators are rather high, it would certainly make more sense to include in the analysis. We hereby refrain from using this model for reasons of simplicity, whilst we present marginal effects of interactions among the obtained results and raw indicators used to obtain them, in order to partially account for such statistical interactions between indicators when trying to estimate the effects elementary indicators exert upon the composite index (see the final two paragraphs of Section 3 for an example).

\[ ^{11} \text{Indeed, the word 'endogenous' in the literature of composite indicators is primarily (if not solely) used to describe weights calculated via an optimising algorithm, such as a DEA or DEA-like framework (see e.g. OECD, 2008).} \]

\[ ^{12} \text{Of course, in order to provide a meaningful framework, a causal direction has to be assumed, in the way that indicators should comprise of outputs only and not inputs used to produce other outputs. Otherwise output-to-input ratio indicators could be used instead if that is desirable, which again boils down to a proper indicator selection in the conceptual framework.} \]

\[ ^{13} \text{Quoting from Becker et al. (2017, p.16): “In practice, the optimised weights may result in negative weights due to the correlation structure between indicators. It is not proposed here that the optimised weights are blindly applied; rather that they are used, among other considerations, as a learning tool to show how far weights are from the “optimal” values, which indicators' weights have little impact on the results, and even which indicators could be removed.”} \]
3 A didactic application of the proposed methodological framework

In this section, we give a didactic example of how a bank could use the framework detailed in this study to evaluate a set of firms. In Section 3.1, we give an overview of the variables we use, the sources of the data, as well as how we model the input parameters included in the methodological framework. Then, in Section 3.2, we give a description of the overall results (the full set of which is appended online as a supplementary appendix for the interested reader), and discuss an example of the further analyses a bank could conduct as these were discussed in section 2.2.2.

3.1 Variables, data and modelling of parameters

As its name suggests, the composite indicator of social, environmental and financial performance (CISEF) proposed in this study is based on three dimensions. This application serves as a didactic illustration of a similar such composite construct, mainly because the choice of indicators falling into the broad three categories can be proxied with a variety of frameworks that are potentially conditional on a country’s regulatory standards, and firms’ disclosure practices. Of course, in a realistic world scenario, a bank could have access to a variety of such data; yet, in empirical studies, the observation is that frameworks are bound to be subject to data availability. As this step extends beyond the scope of this study, we choose a variety of indicators to proxy for important aspects within each dimension that we primarily base on data availability. These are discussed as follows.

We obtain all data from the Refinitiv database (formerly known as ‘DataStream’), operated by Thomson Reuters. We base this choice on its wide gamut of data coverage for sustainability indicators in the ESG spectrum. The elementary criteria chosen to proxy the three dimensions of the CISEF are: *profitability, leverage and liquidity* (financial); *emissions* (environmental); and *product responsibility, community, human rights and workforce* (social). A hierarchical view is presented in Table 2. These indicators are selected in line with prior literature and Refinitiv’s own selection framework, designed in cooperation with analysts and market participants. To conserve space, we detail their description and reasoning in the online supplementary appendix.

Our sample consists of 4,455 unique firms, segmented into 9 industries and operating in 70 countries around the globe in the period 2004-2017. This converts into an unbalanced sample of 28,236 firm-year observations. Raw data were winsorised at their 1st and 99th percentiles to avoid extreme values (particularly apparent around the global financial crisis). Each firm was normalised and evaluated in

---

In particular, the elementary indicators proxying the social and environmental performance dimensions stem from the ASSET4 database, which analyses companies’ ESG profiles. This source has become quite popular for ESG representation according to their use in recent studies (among others, see e.g. Utz et al., 2014, 2015; Gasser et al., 2017; Aouadi and Marsat, 2018). Thomson Reuters collects data points on over 400 metrics from companies’ annual reports, 178 of which were chosen after consultation with analysts and clientele to form the three major categories (‘E’, ‘S’ and ‘G’). The bottom layer of the 178 indicators is essentially aggregated into pillars (calculated by Thomson Reuters as an equally-weighted average of each indicator’s rank score, i.e. $score = \frac{\text{no. of companies with a worst value} + \text{no. of companies with the same value}}{2 \times \text{no. of companies with a value}}$), the score of which varies in the $[0, 100]$ range, and is used as an indicator to proxy a specific aspect. Due to lack of access on the bottom layer 178 scores, we use the pillar ranking scores Thomson Reuters share in their ESG ASSET4 database.
comparison to only its industry peers. Demerjian et al. (2012) treat all observations as a cross-sectional pool (i.e. neglecting the time horizon), as in the presence of a few firms being benchmarked in DEA, a large portion of the firms could lie on the frontier; particularly in the presence of many criteria (Zhou et al., 2007). In the authors' words (Demerjian et al., 2012, p.1235): “Because DEA requires a sufficiently large number of observations to provide a valid estimation, we partition the sample by industry and not by time; it is likely that an industry’s business model will remain stable over time”. However, we find this to be unnecessary in our case for two reasons. First, this should not happen here as the final evaluation is based on a solely two criteria (i.e. $\mu$ and $\sigma$) whilst the minimum sample across segments and years is that of Telecommunication Services in 2004 ($n = 25$; see Table 1), which is still an adequate sample to conduct such an analysis. Second, whilst comparison in a time-window would have a merit indeed (as firm’s performance takes into account also comparison to past performance); in such a time period (2004-2017) we argue that is could be unreasonable comparing firms in 2017 against their peers back in 2004. Therefore, contrary to the authors, instead of estimating the CISEF only by industry, firms are being benchmarked against their peers throughout three-year rolling windows in the duration of our sample, i.e. 2004-2017. For instance, a firm evaluated in the year 2004 takes into account industry-peers (and itself) in the time window 2002-2004.

Finally, as far as the preferences are concerned, due to lack of information on the potential stakeholders’ opinions, or suitability of a different distribution; similar to other studies utilising SMAA in the field of composite indicators (see e.g. Doumpos et al., 2016, 2017; Greco et al., 2017b, 2019b; Angilella et al., 2018; Corrente et al., 2019, to name a few), we approximate the space $W$ using a uniform distribution to draw the weight vectors from in 10,000 simulations (respecting the conditions: $w_j \geq 0$ and $\sum_{j \in J} w_j = 1$ for each weight vector). Due to the nature of the hierarchical structure of CISEF, weights are generated at each level of the hierarchy. That is, in each iteration a set of weights is generated for the first level (3 dimensions), and a set of weights is generated for each dimension to split these weights further in the sub-dimensions, e.g. a set of weights for the financial (3) and another set for the social (4) criteria. The environmental dimension contains only one criterion, so it inherits the weight of the macro dimension in each iteration. Similar to Doumpos et al. (2016), whether simulated weights fell below the 1% threshold, the weight generating process was reiterated.

Table 1 shows the tabulation of firms per industry and year. Table 2 provides the descriptive statistics and Table 3 reports the Pearson’s correlation coefficients.

### 3.2 Empirical results

Starting with the performance of the CISEF and its main components, Fig. 4 portrays the time series of a typical (average) firm for each industry (2004 = 100). Some business segments appear to be overall more stable than others. Moreover the growth patterns seem to differ per industry in terms of each component

---

15 This is in line with the DJI ESG scores (S&P, 2019), as well as the study of Demerjian et al. (2012), who argue (p.1234) that “[they] estimate DEA efficiency by industry (based on Fama and French, 1997) to increase the likelihood that the peer firms have similar business models and cost structures within the estimations”. Arguably, this also extends to the potential existence of industry-specific regulation. In this regard, one may argue that an additional evaluation could be made in regard to country-specific evaluations. We refrained from doing so in this particular example as it extends beyond the scope of the didactic and descriptive example we posit in this section.
Table 1: Distribution of firms per industry and year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Materials</td>
<td>158</td>
<td>258</td>
<td>267</td>
<td>283</td>
<td>185</td>
<td>288</td>
<td>345</td>
<td>460</td>
<td>486</td>
<td>503</td>
<td>290</td>
<td>501</td>
<td>752</td>
<td>747</td>
<td>5,523</td>
</tr>
<tr>
<td>Consumer Cyclicals</td>
<td>77</td>
<td>145</td>
<td>151</td>
<td>178</td>
<td>101</td>
<td>180</td>
<td>238</td>
<td>366</td>
<td>377</td>
<td>392</td>
<td>171</td>
<td>311</td>
<td>502</td>
<td>473</td>
<td>3,662</td>
</tr>
<tr>
<td>Consumer Non-Cyclicals</td>
<td>110</td>
<td>258</td>
<td>267</td>
<td>283</td>
<td>185</td>
<td>288</td>
<td>345</td>
<td>460</td>
<td>486</td>
<td>503</td>
<td>290</td>
<td>501</td>
<td>752</td>
<td>747</td>
<td>5,523</td>
</tr>
<tr>
<td>Energy</td>
<td>62</td>
<td>90</td>
<td>96</td>
<td>121</td>
<td>63</td>
<td>110</td>
<td>141</td>
<td>231</td>
<td>235</td>
<td>242</td>
<td>112</td>
<td>191</td>
<td>322</td>
<td>327</td>
<td>2,538</td>
</tr>
<tr>
<td>Healthcare</td>
<td>48</td>
<td>72</td>
<td>74</td>
<td>84</td>
<td>56</td>
<td>98</td>
<td>131</td>
<td>163</td>
<td>164</td>
<td>170</td>
<td>89</td>
<td>172</td>
<td>239</td>
<td>205</td>
<td>1,765</td>
</tr>
<tr>
<td>Industrials</td>
<td>188</td>
<td>293</td>
<td>300</td>
<td>327</td>
<td>199</td>
<td>329</td>
<td>412</td>
<td>538</td>
<td>549</td>
<td>572</td>
<td>326</td>
<td>585</td>
<td>827</td>
<td>790</td>
<td>6,225</td>
</tr>
<tr>
<td>Technology</td>
<td>69</td>
<td>100</td>
<td>106</td>
<td>109</td>
<td>107</td>
<td>150</td>
<td>198</td>
<td>230</td>
<td>243</td>
<td>198</td>
<td>310</td>
<td>406</td>
<td>421</td>
<td>2,859</td>
<td></td>
</tr>
<tr>
<td>Telecommunications</td>
<td>25</td>
<td>40</td>
<td>41</td>
<td>49</td>
<td>27</td>
<td>69</td>
<td>86</td>
<td>99</td>
<td>98</td>
<td>104</td>
<td>47</td>
<td>113</td>
<td>143</td>
<td>134</td>
<td>1,075</td>
</tr>
<tr>
<td>Utilities</td>
<td>48</td>
<td>72</td>
<td>74</td>
<td>84</td>
<td>56</td>
<td>98</td>
<td>131</td>
<td>163</td>
<td>164</td>
<td>170</td>
<td>89</td>
<td>172</td>
<td>239</td>
<td>205</td>
<td>1,765</td>
</tr>
<tr>
<td>Column ∑</td>
<td>738</td>
<td>1,193</td>
<td>1,240</td>
<td>1,376</td>
<td>895</td>
<td>1,475</td>
<td>1,844</td>
<td>2,447</td>
<td>2,539</td>
<td>2,649</td>
<td>1,490</td>
<td>2,659</td>
<td>3,909</td>
<td>3,782</td>
<td>28,236</td>
</tr>
</tbody>
</table>

Table 2: Descriptive statistics of the elementary indicators

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Indicator</th>
<th>Obs.</th>
<th>Min</th>
<th>Q1</th>
<th>Mean</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>Product responsibility</td>
<td>28,236</td>
<td>0.00</td>
<td>28.79</td>
<td>52.00</td>
<td>50.41</td>
<td>77.27</td>
<td>99.84</td>
<td>27.66</td>
</tr>
<tr>
<td></td>
<td>Community</td>
<td>28,236</td>
<td>0.00</td>
<td>25.45</td>
<td>51.00</td>
<td>50.70</td>
<td>75.79</td>
<td>99.83</td>
<td>28.87</td>
</tr>
<tr>
<td></td>
<td>Human rights</td>
<td>28,236</td>
<td>0.00</td>
<td>29.59</td>
<td>51.50</td>
<td>43.24</td>
<td>75.58</td>
<td>99.75</td>
<td>25.67</td>
</tr>
<tr>
<td></td>
<td>Workforce</td>
<td>28,236</td>
<td>0.00</td>
<td>26.64</td>
<td>51.52</td>
<td>52.21</td>
<td>76.79</td>
<td>99.84</td>
<td>28.93</td>
</tr>
<tr>
<td>Environmental</td>
<td>Emissions</td>
<td>28,236</td>
<td>0.00</td>
<td>28.57</td>
<td>52.42</td>
<td>52.75</td>
<td>77.85</td>
<td>99.84</td>
<td>28.67</td>
</tr>
<tr>
<td>Financial</td>
<td>Profitability</td>
<td>28,236</td>
<td>-0.15</td>
<td>0.08</td>
<td>0.12</td>
<td>0.11</td>
<td>0.16</td>
<td>0.40</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Leverage</td>
<td>28,236</td>
<td>0.00</td>
<td>0.14</td>
<td>0.27</td>
<td>0.25</td>
<td>0.36</td>
<td>0.77</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Liquidity</td>
<td>28,236</td>
<td>0.35</td>
<td>1.05</td>
<td>1.85</td>
<td>1.45</td>
<td>2.10</td>
<td>7.98</td>
<td>1.12</td>
</tr>
</tbody>
</table>

forming the CISEF. For instance, Basic Materials, Energy and Telecommunication Services appear to be the most volatile in terms of financial performance. Industrials appear to be the less volatile in terms of environmental performance growth, although in this case low volatile change seems inappropriate, as there appears to be no significant improvement in terms of the average firm performance for the whole duration of the sample. As a last, but not least important descriptive remark, the performance change of the typical firm in the consumer non-cyclicals business segment seems to be the most noteworthy in terms of advancing performance in the social and environmental front.

Moving from the performance of the typical firm and into a more clear picture of performance per industry, Fig. 5 shows the whole distribution of evaluations. Interestingly, a common pattern appearing across all industries is that the performance of the financial dimension is almost bimodal. Most firms perform either quite adequately or quite poorly, with a more or less uniformly distributed middle ground in between. A common pattern also is observed in terms of the other two dimensions, albeit this time it is not bimodal. In particular, across all industries, the vast majority of firms is evaluated in the [0.3, 0.5] range -in terms of social and environmental performance-, with a feasible range of [0, 1], yet an observed one of [0.17, 0.84] among all firms, years and industries ([min,max] values).
Table 3: Pearson’s correlation coefficients

<table>
<thead>
<tr>
<th></th>
<th>Product responsibility</th>
<th>Community</th>
<th>Human rights</th>
<th>Workforce</th>
<th>Emissions</th>
<th>Profitability</th>
<th>Leverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product responsibility</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>0.3453***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human rights</td>
<td>0.4309***</td>
<td>0.3955***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workforce</td>
<td>0.4576***</td>
<td>0.3596***</td>
<td>0.5214***</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emissions</td>
<td>0.4649***</td>
<td>0.3429***</td>
<td>0.5089***</td>
<td>0.5874***</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profitability</td>
<td>0.0444***</td>
<td>0.0802***</td>
<td>0.0614***</td>
<td>0.0697***</td>
<td>0.053***</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Leverage</td>
<td>0.0041</td>
<td>0.0849***</td>
<td>0.0282***</td>
<td>-0.0212***</td>
<td>-0.0175***</td>
<td>-0.0562***</td>
<td>1</td>
</tr>
<tr>
<td>Liquidity</td>
<td>-0.0751***</td>
<td>-0.0403***</td>
<td>-0.0913***</td>
<td>-0.1048***</td>
<td>-0.0994***</td>
<td>-0.0999***</td>
<td>-0.1839***</td>
</tr>
</tbody>
</table>

*** p<0.01, ** p<0.05, * p<0.1.
Figure 4. Time-series CISEF & components’ performance per industry
On a final note on the descriptive statistics of the evaluation, Fig. 6 shows a heatmap of the typical firm performance across countries in 2017. Size shows magnitude, and colour illustrates the quartile. Firms in the EU seem to be overall the best-performing in terms of the holistic evaluation (CISEF), as they are consistently classified in the top quartile. In particular, in terms of ranking the typical firm performance in 2017, the top-5 consisted of France, Finland, Spain, Sweden and Switzerland. When breaking down the CISEF into its main components, EU firms seem to have a large dispersion in their financial performance that crosses all quartiles, yet they are more consistent - and again the most consistently high- in terms of social and environmental performance.

Turning to the prescriptive statistics and the implications these might have - as these were discussed in Section 2.2.2, we fitted a censored parametric model to disentangle the holistic performance (CISEF) into its raw, elementary criteria used to create it. The fitted model is a Tobit of the form $CISEF_{it} = \theta' z_{it} + \gamma' Y_t + \eta' F_{it},$ where $z_{it}$ contains the raw values of the elementary evaluation criteria, $Y_t$ contains a set of dummy variables capturing year fixed effects and $F_{it}$ contains frontier fixed effects. $\theta$ and $\eta$ are $m \times 1$ size vectors and $\gamma$ is a $T \times 1$ vector of parameters estimated by the model. It is censored with a lower and upper limit of 0 and 1 accordingly, linked to the minimum and maximum feasible values the 16 Out of the 70 countries in our sample, we have excluded countries with less than 4 firms (bottom 10% in terms of number of firms), as well as very small countries such as the Cayman Islands, British Virgin Islands and Jersey.
composite measure can take. As the evaluations were conducted within industries, we fit a Tobit model for each industry. Table 4 shows the results.

The point estimates (Table 4) or the 95% confidence intervals computed based on these (unreported to conserve space) can be used to predict the performance of a new client, or the estimated marginal change of a client’s score should something is to change. An example of the latter is as follows. Assuming that a bank’s client is in the Energy business sector, a standard deviation increase in its profitability (equal to 0.11) would increase its holistic overall performance, on average, by 0.01441 \(ceteris paribus\), which is 10% of that industry’s CISEF performance’s standard deviation (\(std_{CISEF_{Energy}} = 0.1421\)).

To tailor these results even more, one may model interaction effects from a statistical point of view, which offers several practical insights (Aguinis et al., 2017). Interactions show more tailored estimated effects of a variable at different values of another variable. Technically speaking, continuous by continuous interaction could easily be introduced in the regression model and be disentangled through marginal effects decomposition. To give an example of how these could work in this case, we take the same example of a firm in the Energy sector. In particular, suppose that a bank would like to

---

17 The adjusted R-square values obtained from an OLS fit of these models range from approximately 89 to 91%.
18 For a brief overview and a computational implementation through STATA’s postestimation, see Williams (2012).
provide advice to one of its clients. Let us assume that the client’s executive team prioritises financial and environmental means over social ones (that is assumed on the basis of simplification, using only a one-way continuous by continuous interaction). Suppose that the manager would like to see what the estimated score of her firm would be at different potential values of emissions score if the firm’s profitability was to decrease by a standard deviation. That would help the manager see what are the estimated trade-offs if she increases her firm’s environmental performance, ceteris paribus, in the future.

Assuming that this firm has (or is interested in achieving) an environmental performance in the top quartile, we only focus on interaction effects in the observable range of Emissions = [79, 99] in this industry. Note here that ‘Emissions’ is normalised as obtained from EIKON so that higher values reflect better performance and vice versa.

Table 5 shows the estimated unit impact (point estimate) of profitability on CISEF at different (random but equally spaced) values of Emissions in the top quartile, controlling for all other attributes. Notice how the coefficient of Profitability as a point estimate from 0.131 (Table 4, Column 4) on average across all firms is now quadrupled when it interacts with the Environmental dimension, showing that the aspect of profitability could have much larger effects on the estimated CISEF at top values (Q3) of the Environmental dimension. Fig. 7 visually delineates the contents of 5, showing how the slope of profitability (in determining CISEF; all else at their means) increases as performance on Emissions increase. To aid facilitation of this visual, as the values of Emissions increase, the fitted line’s colour becomes faded.

**Figure 7. Delineation of Marginal Effects: Profitability at different values of Emissions**

![Predictive Margins](image_url)
Table 4: Disentangling the CISEF into its elementary (raw) criteria - Tobit regressions

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>Basic Material</th>
<th>Consumer Cyclicals</th>
<th>Consumer Non-cyclicals</th>
<th>Energy</th>
<th>Healthcare</th>
<th>Industrials</th>
<th>Technology</th>
<th>Telecommunication Services</th>
<th>Utilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Responsibility</td>
<td>0.000565***</td>
<td>0.000685***</td>
<td>0.000464***</td>
<td>0.000772***</td>
<td>0.000555***</td>
<td>0.000868***</td>
<td>0.000601***</td>
<td>0.000517***</td>
<td>0.000681***</td>
</tr>
<tr>
<td></td>
<td>(5.37e-05)</td>
<td>(5.02e-05)</td>
<td>(6.79e-05)</td>
<td>(6.41e-05)</td>
<td>(7.55e-05)</td>
<td>(4.60e-05)</td>
<td>(6.33e-05)</td>
<td>(0.000117)</td>
<td>(8.06e-05)</td>
</tr>
<tr>
<td>Community</td>
<td>0.000608***</td>
<td>0.000576***</td>
<td>0.000721***</td>
<td>0.000838***</td>
<td>0.000738***</td>
<td>0.000454***</td>
<td>0.000806***</td>
<td>0.000262***</td>
<td>0.000566***</td>
</tr>
<tr>
<td></td>
<td>(4.97e-05)</td>
<td>(4.26e-05)</td>
<td>(6.17e-05)</td>
<td>(5.62e-05)</td>
<td>(7.06e-05)</td>
<td>(4.43e-05)</td>
<td>(6.21e-05)</td>
<td>(9.30e-05)</td>
<td>(7.14e-05)</td>
</tr>
<tr>
<td>Human rights</td>
<td>0.00109***</td>
<td>0.000848***</td>
<td>0.000840***</td>
<td>0.00130***</td>
<td>0.000918***</td>
<td>0.000929***</td>
<td>0.000747***</td>
<td>0.00111***</td>
<td>0.00116***</td>
</tr>
<tr>
<td></td>
<td>(6.24e-05)</td>
<td>(5.46e-05)</td>
<td>(7.67e-05)</td>
<td>(7.52e-05)</td>
<td>(8.89e-05)</td>
<td>(5.36e-05)</td>
<td>(7.46e-05)</td>
<td>(9.00e-05)</td>
<td>(8.70e-05)</td>
</tr>
<tr>
<td>Workforce</td>
<td>0.000477***</td>
<td>0.000643***</td>
<td>0.000701***</td>
<td>0.000604***</td>
<td>0.000513***</td>
<td>0.000641***</td>
<td>0.000515***</td>
<td>0.000767***</td>
<td>0.000474***</td>
</tr>
<tr>
<td></td>
<td>(5.75e-05)</td>
<td>(5.03e-05)</td>
<td>(6.96e-05)</td>
<td>(6.78e-05)</td>
<td>(7.93e-05)</td>
<td>(5.10e-05)</td>
<td>(7.45e-05)</td>
<td>(9.30e-05)</td>
<td>(8.38e-05)</td>
</tr>
<tr>
<td>Environmental</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emissions (lack thereof)</td>
<td>0.00300***</td>
<td>0.00296***</td>
<td>0.00297***</td>
<td>0.00265***</td>
<td>0.00307***</td>
<td>0.00262***</td>
<td>0.00265***</td>
<td>0.00303***</td>
<td>0.00298***</td>
</tr>
<tr>
<td></td>
<td>(6.00e-05)</td>
<td>(5.04e-05)</td>
<td>(7.13e-05)</td>
<td>(7.02e-05)</td>
<td>(7.74e-05)</td>
<td>(5.01e-05)</td>
<td>(7.34e-05)</td>
<td>(0.000125)</td>
<td>(8.29e-05)</td>
</tr>
<tr>
<td>Financial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Profitability</td>
<td>0.0525***</td>
<td>0.135***</td>
<td>0.164***</td>
<td>0.131***</td>
<td>0.0527***</td>
<td>0.149***</td>
<td>0.154***</td>
<td>0.0579**</td>
<td>0.292***</td>
</tr>
<tr>
<td></td>
<td>(0.0122)</td>
<td>(0.0123)</td>
<td>(0.0175)</td>
<td>(0.0166)</td>
<td>(0.00893)</td>
<td>(0.0161)</td>
<td>(0.0162)</td>
<td>(0.0229)</td>
<td>(0.0492)</td>
</tr>
<tr>
<td>Leverage</td>
<td>-0.142***</td>
<td>-0.0963***</td>
<td>-0.101***</td>
<td>-0.0461***</td>
<td>-0.0744***</td>
<td>-0.107***</td>
<td>-0.106***</td>
<td>-0.0950***</td>
<td>-0.0964***</td>
</tr>
<tr>
<td></td>
<td>(0.00865)</td>
<td>(0.00549)</td>
<td>(0.00817)</td>
<td>(0.00523)</td>
<td>(0.00858)</td>
<td>(0.00702)</td>
<td>(0.00817)</td>
<td>(0.0128)</td>
<td>(0.0139)</td>
</tr>
<tr>
<td>Liquidity</td>
<td>0.00219***</td>
<td>0.00138***</td>
<td>0.00240***</td>
<td>0.00278***</td>
<td>0.0026***</td>
<td>0.00188***</td>
<td>0.00191***</td>
<td>0.00169***</td>
<td>0.00617***</td>
</tr>
<tr>
<td></td>
<td>(0.000552)</td>
<td>(0.000758)</td>
<td>(0.000701)</td>
<td>(0.000626)</td>
<td>(0.000629)</td>
<td>(0.000610)</td>
<td>(0.000606)</td>
<td>(0.000672)</td>
<td>(0.000644)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.242***</td>
<td>0.201***</td>
<td>0.210***</td>
<td>0.158***</td>
<td>0.175***</td>
<td>0.223***</td>
<td>0.197***</td>
<td>0.246**</td>
<td>0.202***</td>
</tr>
<tr>
<td></td>
<td>(0.00992)</td>
<td>(0.00744)</td>
<td>(0.0104)</td>
<td>(0.0109)</td>
<td>(0.0113)</td>
<td>(0.00791)</td>
<td>(0.0108)</td>
<td>(0.0180)</td>
<td>(0.0147)</td>
</tr>
<tr>
<td>Year FE</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Observations</td>
<td>3,662</td>
<td>5,523</td>
<td>2,583</td>
<td>2,338</td>
<td>2,206</td>
<td>6,225</td>
<td>2,859</td>
<td>1,075</td>
<td>1,765</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses (clustered by firm). *** p<0.01, ** p<0.05, * p<0.1
Table 5: Marginal Effects: Decomposing the interaction between Profitability and Emissions

<table>
<thead>
<tr>
<th>Profitability (dy/dx) at:</th>
<th>$\beta_{\text{Profitability}}$</th>
<th>Std. Error</th>
<th>t-statistic</th>
<th>p-value</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Emissions = 79$</td>
<td>0.471</td>
<td>0.025</td>
<td>18.81</td>
<td>0.000</td>
<td>0.422 - 0.521</td>
</tr>
<tr>
<td>$Emissions = 81.5$</td>
<td>0.494</td>
<td>0.026</td>
<td>18.93</td>
<td>0.000</td>
<td>0.443 - 0.545</td>
</tr>
<tr>
<td>$Emissions = 84$</td>
<td>0.517</td>
<td>0.027</td>
<td>19.03</td>
<td>0.000</td>
<td>0.464 - 0.570</td>
</tr>
<tr>
<td>$Emissions = 86.5$</td>
<td>0.539</td>
<td>0.028</td>
<td>19.1</td>
<td>0.000</td>
<td>0.484 - 0.595</td>
</tr>
<tr>
<td>$Emissions = 89$</td>
<td>0.562</td>
<td>0.029</td>
<td>19.16</td>
<td>0.000</td>
<td>0.505 - 0.620</td>
</tr>
<tr>
<td>$Emissions = 91.5$</td>
<td>0.585</td>
<td>0.030</td>
<td>19.2</td>
<td>0.000</td>
<td>0.525 - 0.645</td>
</tr>
<tr>
<td>$Emissions = 94$</td>
<td>0.607</td>
<td>0.032</td>
<td>19.24</td>
<td>0.000</td>
<td>0.546 - 0.669</td>
</tr>
<tr>
<td>$Emissions = 96.5$</td>
<td>0.630</td>
<td>0.033</td>
<td>19.26</td>
<td>0.000</td>
<td>0.566 - 0.694</td>
</tr>
<tr>
<td>$Emissions = 99$</td>
<td>0.653</td>
<td>0.034</td>
<td>19.27</td>
<td>0.000</td>
<td>0.586 - 0.719</td>
</tr>
</tbody>
</table>
4 Conclusions

In this study, we propose a holistic evaluation framework that banks could adopt in their role as quasi-insiders. The framework is based on the development of a composite indicator based on social, environmental and financial performance that we name CISEF. This integrated view of overall performance is in line with ISO standards and empirical proposals of academics and market practitioners in measuring inclusive performance. However, we complement previous proposals with a methodological framework that deems the index more ‘holistic’, by incorporating not a single set of preferences, but a variety of such. In particular, the proposed methodological framework permits incorporation of a plethora of viewpoints -generally speaking the stakeholders in the environment of a bank- in the evaluation procedure. We enhance this process with proposals that further provide analytics to delve into a more detailed analysis of performance from a hierarchical point of view, fine-tune the evaluation process -providing a robust score that is invariant to the sample- or, simply, to offer advice to their customers along the three dimensions of financial, social, and environmental performance in the form of customer-based value-adding solutions that give banks a competitive advantage over rival counterparts. This can form the basis of an initial screening to alleviate adverse selection issues, but also of ongoing monitoring that will mitigate moral hazard issues.

This framework could be of value to both banks and their customers, and it therefore has various managerial implications. First, as it concerns banking institutions, it is in line with the notion of responsible lending, signals responsible conduct and can be appreciated by the variety of stakeholders in a bank’s environment, like depositors and investors. Additionally, it takes into account recent regulatory initiatives that stress the importance of the risks that emerge for banking institutions from climate change.

On the same note, the proposed framework is in accordance with recent suggestions from market participants like the the Global Association of Risk Professionals (GARP) Risk Institute. For example, Nelson (2020) argues that it might be important to know the greenhouse gas emissions of a portfolio, or whether that portfolio is aligned with the Paris Agreement, or how it could be affected by increasing physical risk. She also suggests building a global standard for ranking climate financial risk, with clear criteria that can be readily applied, enabling stakeholders to compare risk data easily across firms. Furthermore, Paisley and Nelson (2019, p.2) argue that “The treatment of climate risk at financial institutions has changed significantly over the past five years. Whereas it used to be viewed mostly as a reputational risk that could be addressed through the environmental, social and governance (ESG) agenda, climate change is now seen by many firms as a financial risk that needs to be integrated into existing risk management frameworks”.

The adoption of such a framework could also provide a competitive advantage to banks. As highlighted in Oracle (2019, p.5) “Corporate Customers are not just looking at expertise on specific financial domain instead they are looking for innovative ideas that are specific to their industry and enables them to tackle their unique business situations by looking at the problem from a different angle”. Therefore, to maintain their central role in corporate customers’ financial management and transactions, banks need to evolve to become customer-centric (Oracle, 2019). The literature has always emphasized that banks
play a monitoring role and they influence an array of their customers’ business decisions (Vashishtha, 2014; Marshall et al., 2014; Saunders and Song, 2018). Within this context, the proposed framework could be used not only as a screening tool prior to the lending decision (i.e. mitigating adverse selection problems), but also as post-monitoring tool that will be of value to the bank (i.e. mitigating moral hazard concerns) and will serve as the basis of advice to its customers. Offering advice to existing customers as far who to improve along the dimensions of CISEF could be seen as a value-adding service, allowing the bank to differentiate itself from its competitors. This competitive advantage would enable the bank to attract new customers, retain existing ones, and possibly sell-higher margin products to clients. In other words, it can serve as a marketing tool that can demonstrate the bank’s commitment to its clients.

Additionally, the adoption of the proposed framework by banking institutions could benefit their clients in at least one more way. If enterprises know that they will be evaluated against CRS criteria, then managers may take actions, to enhance their performance on such aspects, which have been associated with various benefits for firms like firm value (Ferrell et al., 2016), lower cost of equity (El Ghoul et al., 2011), lower idiosyncratic risk (Lee and Faff, 2009) and volatility (Boutin-Dufresne and Savaria, 2004), higher market-to-book ratios (Galema et al., 2008), long-run stock market performance (Edmans, 2011, 2012) and fewer capital constraints (Cheng et al., 2014). Therefore, if firms are unwittingly pushed to adopt such framework-as this will be part of their evaluation-, it could be a positive driving factor and a justifiable cause for all stakeholders in the environment of a firm that would otherwise oppose it on the basis of other, solely financial interests. So, both from the viewpoint of firms and banks, CSR additions to their financial prowess could be seen as a ‘risk-mitigation’ technique (Goss and Roberts, 2011).

Finally, our framework could be revised and extended towards various directions to support managerial decision making. One way could be to extend it to include dimensional insights alongside the use of composite indicators as a scorecard. For instance, consider that a firm manager wants to improve performance in a certain dimension, say Social. The censored regression model given as an example in Section 3 could be adjusted to a form of \( \hat{S}_{Social_{it}} = \theta' z^s_{it} + \gamma' Y_t + \eta' F^s_{it} \), where in that case \( z^s_{it} \) contains only the social-related attributes (i.e. Product responsibility, Community, Human Rights, and Workforce). This within-dimensional regression model, however, would need to be combined with a second macro-view of the hierarchical regression model presented in the application, i.e. to estimate how each dimension (financial, environmental and social) contributes to the overall CISEF, perhaps also with interactions between dimensions. This could show how each elementary indicator contributes to the dimensional score, and how this then interacts with other dimensions to increase the overall CISEF when trade-offs have to be made from a manager’s perspective.

Acknowledgements

Montpellier Business School (MBS) is a founding member of the public research center Montpellier Research in Management, MRM (EA 4557, Univ. Montpellier).

We would like to thank four anonymous reviewers that greatly helped us in improving the previous state of this manuscript. Thanks are also extended to participants of the 30th European (EURO) Conference on Operational Research. All errors remain our own.
References


European Banking Authority (2019). EBA action plan on sustainable finance. 6 December.


Supplementary Appendix:
Description of elementary criteria

Starting with the corporate sustainability attributes, we largely rely on ASSET4's coverage and choices of indicators falling into these two categories. Whilst the database provides these data in bulk in a trichotomic default categorisation, namely the ESG; we do not include the last category that proxies for corporate governance attributes ('G' of the 'ESG' framework). The reason boils down to our belief that the corporate governance of an entity is the mechanism through which the outcomes (e.g. financial, social and environmental performance) of the entity are achieved, and not yet another aspect of the inclusive evaluation of the entity we aim to measure. That is indeed empirically found to be true by Jo and Harjoto (2012), whose findings show a causal effect running from governance to CSR engagement. ¹

Turning to the included criteria, starting with the social performance dimension, the first proxy chosen from the ASSET 4 framework is that of product responsibility. This shows a rank score of an entity’s capacity to have policies in place on matters such as protecting its customers’ health and safety, personal data, and policies about general integrity, quality standards (ISO) and privacy. Additionally, the community score shows an entity’s good citizenship in many forms, including having policies on fair competition, having been receiving recognition, e.g. through awards and donations for good causes from its local surroundings, or financially aiding and supporting the communities in the vicinity. The human rights score rates an entity’s capacity to have policies in place avoiding the use of force or child labor, ensure the freedom of association of its employees, complying with UN’s declaration of human rights, and keeping the same standards with its partners too. Last proxy on this dimension is workforce quality and diversity. This index shows an entity’s capacity to have policies regarding its diversity and equal opportunities, and setting targets to achieve that, as well as policies on flexible working schemes, and employee training and promotion. Moreover, for an entity to score high on this dimension, it shall have displayed diversity and equal opportunities, as displayed by the reported percentages of employees being women, people with disabilities, as well as the amounts and hours that it spends (% of its assets) on training its employees, and the gap between the highest salary in the upper echelons (CEO or higher) and the average employee (ratio).

Continuing with the second responsibility dimension, i.e. that of environmental performance, the ASSET4 framework includes three pillars, namely environmental innovation, emissions reduction and

¹However missing from the framework that we will present, we do recognise that this view may differ from an application basis. We would like to thank an anonymous reviewer for suggesting that corporate governance indicators should possibly be included as additional criteria into a compositive index of social, environmental, governance, and financial performance. Taking into account this comment, along with regulatory and practitioners’ references to ESG ratings, we re-estimate our composite index with the addition of corporate governance indicators, and we provide the results -alongside those discussed in Section 3- in the online Appendix. The Pearson’s correlation coefficient between the two composite indices (i.e. with and without the corporate governance indicators) equals 0.88.
resources reduction. These cover several aspects of an entity’s environmental performance for having potential policies such as setting targets for water/energy efficiency, or using hybrid technology etc. We choose only the emissions reduction pillar, which includes meeting targets on environmental reduction, amounts of CO2 and CO2-equivalent emissions (% of sales), participating in an emissions trading initiative, environmental fines, tonnes recycled and tonnes of waste produced (% of sales) and investing in products and services to reduce future environmental risks. We do so on the basis of choosing a measurable output (i.e. environmental impact) rather than observing policies only being put in place.

Last, but not least, in the discussion that follows we outline the financial performance attributes. As our entities are segmented across different industries, we choose criteria that are broad enough to extend beyond industry-specific ones, but still cover the basic aspects of financial analysis of an entity. At the same time we intend to keep the number of criteria manageable not to lose a big portion of our sample due to missing data. The first included criterion is that of profitability. We compute an entity’s return on assets (ROA), a common proxy for profitability (Greene and Segal, 2004; Cho and Pucik, 2005; Athanasoglou et al., 2008; Enqvist et al., 2014; Gaganis et al., 2019), which is the ratio of net income divided by total assets. This indicator shows investors a picture of how effective an entity is at converting the money it invests (assets) into net income. The next criterion is leverage and it is commonly expressed as a debt-to-assets ratio (Ahn et al., 2006; Verwijmeren and Derwall, 2010; Bae et al., 2011). In particular, we compute leverage as the ratio of an entity’s total debt divided by its total assets. The higher this figure is, the more leveraged a company is and thus the higher is its risk. The third criterion is liquidity and is proxied by the current ratio (Dahlquist and Robertsson, 2001; Sharma and Kumar, 2011; Saleem and Rehman, 2011), an entity’s current assets divided by its current liabilities. It shows an entity’s capacity to have adequate coverage of its short-term debts with its short-term assets.
References


