Multi-criteria decision analysis: Methods to define and evaluate socially responsible investments

Tim Verheyden*
Lieven De Moor†

October 25, 2014

Abstract

Originally being a niche strategy followed by few investors, socially responsible investing (SRI) now represents a significant part of the assets under management. After summarizing empirical evidence on the performance of SRI funds, we present four challenges that are facing the further development of SRI and point to multi-criteria decision analysis (MCDA) as the methodological framework that could help overcome these challenges. A first group of challenges calls for the development of a social performance indicator, which can score and classify mutual funds with respect to social responsibility. Another challenge requires a transparent tool for retail investors interested in SRI to learn about their SRI preferences. Reviewing the three schools of available MCDA methods, we present a concrete approach for future research in building such a social performance indicator and a retail investor tool for SRI.

Keywords: Socially responsible investing; multi-criteria decision analysis; mutual funds
JEL-codes: G11

*Corresponding author, KU Leuven (Brussels Campus), Warmoesberg 26, 1000 Brussels, Belgium, tim.verheyden@kuleuven.be
†Vrije Universiteit Brussel, Faculty of Economic and Social Sciences and Solvay Business School, Pleinlaan 2, 1050 Brussels, Belgium, lieven.de.moor@vub.ac.be

1 Introduction

Socially responsible investing (SRI) has experienced a rapid growth over the past decade, reflecting the increasing awareness of investors for environmental, social and governance (ESG) issues. Different definitions of SRI have been suggested. Instead of looking at one static definition, we might consider the development of the SRI concept. The first generation of SRI simply consisted of the application of negative screens to the investment universe. In considering different investment opportunities, some criteria are established to screen out "sin stocks" or companies that are in discordance with a set of moral and/or ethical principles. In the second generation of SRI, the focus was more on adopting positive screens and a best-in-class approach. The combination of both positive and negative screens led to the third generation of SRI. The fourth and most recent generation of SRI includes shareholder activism, next to the application of positive and negative screens.

The origins of SRI go back to the moral principles adopted by religious organizations in considering investment alternatives. The definitive breakthrough for socially responsible investment came with the massive worldwide protest against the racist system of apartheid in South Africa. In recent years SRI has moved from niche to mainstream (KPMG & ALFI, 2013), as issues like global warming, the Kyoto Protocol, corporate governance, and community investing have gained significant attention from investors around the world. In addition, governments in western countries have taken many regulatory initiatives to stimulate SRI. Both elements create a pro-SRI environment in which SRI will continue to grow and establish its relative importance as an asset class (Renneboog et al., 2008). According to the Forum for Sustainable and Responsible Investment (2012) a little more than one out of every nine dollars (11.3%) under professional asset management in the United States is invested in the SRI universe. At the start of 2012, SRI assets managed by professionals stood at $3.744 trillion, a rise of more than 486 percent from the $639 billion in 1995. Over the same period, the broader universe of assets under conventional professional management rose only 376 percent. The latest SRI study by Eurosif (2012) demonstrates similar results. The combined growth of SRI strategies has outperformed the conventional market in Europe, despite the current financial turmoil.

The central aim of this paper is to open up alleys for future research, by pointing out some outstanding challenges in the SRI field and by presenting a methodological approach to address these challenges. We briefly summarize the empirical research on the financial performance of SRI funds and argue that the current dichotomic classification of a funds' social responsibility explains why previous research finds insignificant differences in returns between SRI and conventional funds. In order to enrich the academic state-of-the-art, we see a need for a proper framework to select, evaluate and categorize SRI funds in a more nuanced and continuous way. Such a framework can also benefit governmental agencies in regulating the SRI market and commercial banks in developing new SRI products. Additionally, from the 2012 Eurosif SRI study, we see that the European SRI retail market is particularly underdeveloped. The main part of SRI investments and growth come from institutional investors. One of the reasons is a lack of transparency for retail investors and the proliferation of different methods to determine the social performance of mutual funds. Clearly, there is a need for a transparent yardstick allowing retail investors to determine the social responsibility content of their investments. Also, retail investors need some guidance in getting to know their personal preferences with regard to social investing. For all of these challenges, we point to multi-criteria decision analysis (MCDA) as an interesting methodology to help define and

evaluate the social performance of mutual funds. We review the different schools of MCDA methods and present a concrete approach to apply MCDA to the modern challenges facing the further development of SRI as an asset class.

2 The SRI performance debate

As several review papers on SRI performance have been written (e.g. Margolis & Walsh, 2003; Orlitzky et al., 2003), it is not our aim to provide a complete overview of earlier research on this topic. Rather, we present a summary from the two fields where research has been conducted and the different findings that have been presented.

In strategic management science, the debate on corporate social versus corporate financial performance goes back to the opposing views of Friedman (1970) and Freeman (1984). In a New York Times Magazine article, Friedman (1970) makes the case for shareholder theory, which states that the sole responsibility of businesses is to maximize the value for its shareholders. In this view, which is also referred to as Friedman's doctrine, it is believed that society at large benefits most if companies simply focus on maximizing their own profits. Consequently, any corporate social responsibility (CSR) initiative is obsolete. Stakeholder theory, first proposed by Freeman (1984), takes a different view on the role of a business. The responsibility of a firm should not be limited to the shareholders, but should consider all the stakeholders. This is not only believed to increase overall welfare, but also the profitability of individual firms as the theory argues that a good relationship with all the stakeholders will improve long-term financial performance. Reconciliating the opposing views, Mackey et al. (2007) find that demand and supply conditions for socially responsible investment opportunities determine whether socially responsible decisions can lead to better financial performance.

The same question on the financial performance of SRI was posed in the field of financial economics. Mostly, researchers have implemented empirical models to compare the financial performance of SRI and conventional funds, controlling for different factors of risk. The first empirical studies go back to Moskowitz (1972) and Bragdon and Marlin (1972), who find a positive rank correlation between corporate social and financial performance. The sophistication of the applied methods has since then increased, and so has the quality of the results. Following the rank correlation tests, SRI performance was researched using the capital asset pricing model (CAPM; Sharpe, 1964), controlling only for market risk (e.g. Hamilton et al., 1993), and using performance ratios like the Sharpe index (e.g. Sauer, 1997). As more recent advances in empirical asset pricing got adopted across the field, the CAPM was gradually replaced by the Fama-French (1993) three-factor model and the Carhart (1997) four-factor model (e.g. Bauer et al., 2005). Today, most SRI performance research implements a conditional four-factor model, following Ferson and Schadt (1996), taking into account possible time-varying risk. Given these different possible approaches, Geczy et al. (2003) found that the cost of investing in SRI funds instead of conventional funds crucially depends on two elements: the belief held by the investor regarding the valid underlying asset pricing model, and the ability of fund managers to select stocks. The cost of investing in SRI when assuming the CAPM holds true and fund managers have no stock-picking skills is negligible. However, the cost is substantial when adhering to a four-factor model and assuming that fund managers have some skills in selecting stocks.

Regardless of the evolution in the methodological approach of researching the performance of SRI funds, inconclusive results have been found from the start. This led scholars to divide into believers and non-believers of SRI, each with their own set of arguments. The non-believers, who argue that SRI funds can only underperform traditional funds, mainly refer to modern portfolio theory (Markowitz, 1952). Because social screening adds constraints to the optimization problem of finding an efficient portfolio, some idiosyncratic risk cannot be diversified away. Consequently, SRI portfolios are not on the efficient frontier and thus will yield subpar risk-adjusted returns. The critique of the non-believers is also in line with the argument of Friedman (1970). Engaging in socially responsible activities increases the operational costs of a firm, which negatively impacts overall profitability. The believers contend that social responsibility is not a cost, but rather an investment as firms are an inherent part of their social environment (Granovetter, 1985). This view is supported by stakeholder theory (Freeman, 1984). Additionally, the believers refute the argument that social screening leads to inefficient portfolios. Even though the pool of potential stocks to diversify away idiosyncratic risk is smaller because of the additional constraints, the quality of this pool is believed to be higher as the screening process yields value-relevant information for the investor (Barnett & Salomon, 2006).

3 Challenges and issues for SRI

Despite the progress of SRI as an asset class, the research on SRI performance is still ongoing. We identify four important challenges for the continued growth of SRI, both in the academic and the professional world.

A first challenge concerns the methodological approach for the performance analysis of SRI funds in academics. From the review of methodologies to test for SRI performance, we learn that measuring risk-adjusted returns from asset pricing models is standard practice today. To compare SRI and conventional funds, a difference portfolio is usually constructed from a dichotomous dummy variable that indicates whether the fund is labeled socially responsible or not. The problem with the dummy approach is that it neglects possible heterogeneity among different SRI funds and that it does not take into account the multiple dimensions relevant to social responsibility. In reality socially responsible investors do not adopt a dichotomous classification approach and need to carefully examine the mutual funds' prospectus to examine if the fund's investment strategy and social responsible guidelines meet their individual ethical standards (Hollingworth, 1998). Hoggett and Nahan (2002) and Tippet (2001) show that this kind of SRI investment information might be hard to retrieve or even be unreliable. Barnett and Salomon (2006) already addressed the issue of dichotomy and found a curvilinear relationship between corporate social and financial performance. These findings even suggest that both views on SRI performance could coexist. Schwartz (2003) and Koellner et al. (2005) propose a general code of ethics for socially responsible investing regarding the information disclosure or transparency, the investment process and the credibility of information, but there is a lack of specific social responsible indicators that capture the multifaceted nature of SRI. A methodology to overcome the dichotomous measurement of social responsibility in asset pricing models is still lacking, although some earlier research was devoted to examining the effect of different types of SRI screens on flow-return relations (Renneboog et al., 2011). We believe that the development of a social responsibility indicator to score and/or classify funds based on multi-criteria decision analysis (MCDA) methodologies could overcome this issue. The MCDA field is devoted to the development of appropriate methodologies that can be used to support and aid decision

makers in circumstances where multiple conflicting decision factors (objectives, goals, criteria) have to be considered simultaneously. The application of MCDA to finance problems is not new (see e.g. Steuer & Na (2003) and Zopounidis & Doumpos (2002) for an overview of this line of research).

The lack of a proper regulatory framework in the certification of mutual funds is the second challenge. Although at the company level, several independent agencies[1] try to supply transparent and credible information about the social, labor and environmental performance of corporations throughout the world, few rating agencies monitor the process-oriented social responsibility value/authenticity of mutual funds. Most of these agencies only provide financial information about the funds (costs, performance, risk and liquidity) and conventional investment strategy information (type of security, country and industry allocation, financial investment objectives and fund composition). Supervising authorities are currently unable to adequately screen the design of ethical mutual funds to, for example, grant a certificate of "ethical authenticity" to funds or to promote ethical investments all together. As SRI is becoming more popular, this gives an incentive to investment institutions to label their mutual funds socially responsible, even though this is not really the case. The MCDA framework can again help to create a tool to classify mutual funds based on an assessment of a wide variety of underlying socially responsible criteria.

In the absence of a clear framework to define and categorize SRI funds, commercial banks also need to spend considerable time and resources to develop, implement and communicate an in-house SRI view. This forms the third challenge to the further development of SRI. A framework based on MCDA can help companies to construct SRI mutual funds in a more efficient, consistent and transparent way. Consequently, consumers would be less confused by standalone SRI definitions that differ from bank to bank and would be enabled to compare different SRI mutual funds in a straightforward manner. Assessing an investment alternative would then be possible along three dimensions: risk, return and social responsibility.

These first three challenges are rather similar, in the sense that they can be addressed in the same way. They all need an overarching framework that can help in scoring and classifying mutual funds based on social responsibility. More specifically, a social performance indicator can be helpful in discriminating between the social responsible design of mutual funds in a more continuous way, it would give regulators a tool to develop labels — based on categories or scores — for genuine SRI funds and it would facilitate the process of developing new SRI mutual funds for banks. This performance indicator needs to be as general as possible, taking into account views from all the different stakeholders and interest groups involved in the SRI field. A more concrete approach for developing this indicator is presented in section 5.

A final outstanding challenge for the field of SRI was revealed by the 2012 Eurosif report on the European SRI mutual fund industry. Even though the SRI market continues to grow, the retail segment remains underdeveloped as growth and volume in the SRI market predominantly comes from institutional investors. For the overall European mutual fund industry, 25% of assets under management are held by retail investors (European Fund and Asset Management Association, 2013). Retail investments in the European SRI mutual fund industry only amount to 6% of assets under management, which is illustrative of a large potential for growth (Eurosif, 2012). As a main reason for this underdevelopment of the European SRI retail mutual fund market, Eurosif (2012) points towards bad communication and a lack of transparency and clarification of SRI strategies,

which keeps many retail investors from investing in SRI funds. Again, MCDA could provide the framework to overcome this challenge, as the methodology can assist retail investors in handling extensive information in a transparent way. As MCDA is focused on accommodating better decisions, it can also assist retail investors in making wiser investment choices. Similar to the investment services directive by the European Commission to allow investors to better understand the risk they want to take, we believe the MCDA framework could be formalized into a "green" MiFID (Markets in Financial Instruments Directive) questionnaire, which could assist investors in better understanding their social responsibility preferences and increase the investor protection in Europe with respect to SRI (Davies et al., 2006). Note that the MCDA tool to address this challenge needs to be more tailored to the needs of individual investors, which sets the fourth challenge apart from the first three.

Different authors have attempted to define and evaluate SRI before. Within the field of operations research, Pérez-Gladish and M'Zali (2010) already constructed a first SRI indicator applying one MCDA method called the analytic hierarchy process (AHP). Although very interesting, their paper lacks methodological sophistication as it only considers and implements one MCDA method, whereas other methods might be more appropriate. Additionally, the input required from investors and/or experts is too detailed (e.g. different gradations in assessing SRI criteria), which makes the method harder to implement, more unreliable and susceptible to rank reversal issues, which we discuss in more detail further on in this paper. Within the field of financial economics, different authors have tried to include social responsibility in performance evaluation estimations (e.g. Renneboog et al., 2008). However, these attempts are rather limited in scope, as they typically reduce the concept of social responsibility to a dummy-variable, which is driven by the underlying decision of a financial institution to promote investment products as SRI or not, leading to a classification bias. Finally, the above-mentioned SRI challenges have also been extensively addressed in the strategic management literature (e.g. Waddock and Graves, 1997). Although insightful, these studies fail to incorporate the social responsibility topic in the financial performance methodology, which leads to weak and incomplete results on the performance of socially responsible investment strategies vis-à-vis conventional investment strategies. CSR evaluation techniques are also not appropriate to address the problem at hand since they focus on in-depth individual analysis of companies. As we are trying to address SRI performance and regulatory issues on the fund level, we are looking for aggregate fund-level assessments of social responsibility, which calls for a different approach and set of criteria. Furthermore, we aim for a quantitative instrument transforming qualitative criteria into an indicator score or category, as to integrate these into performance evaluation multi-factor models from the financial economics literature.

Clearly, MCDA could be instrumental in addressing all of the above challenges, and in improving earlier work on this topic. This is not to say that MCDA is the magic formula that will resolve each and every issue, but merely that it can provide the framework to help move the SRI field to the next level. In order to further explore and guide future research, we review the different schools of MCDA methodologies and show how these methods could be used as a tool to define and evaluate SRI funds.

4 Overview of MCDA methods

The development of MCDA, an advanced field of operations research, is based on the simple finding that a single objective, goal, criterion or point of view is rarely used to make real-world decisions. The MCDA field is devoted to the development of appropriate methodologies that can be used to support and aid decision makers in situations where multiple conflicting decision factors (objectives, goals, criteria) have to be considered simultaneously. Given the different dimensions to the concept of social responsibility, MCDA is also relevant in scoring and classifying SRI mutual funds.

Within a multi-criteria context, decision-making problems are realized in the following paradigm: a decision maker considers a set of alternatives and seeks to take an "optimal" decision considering all the factors that are relevant to the analysis. Since these factors usually lead to conflicting results and conclusions, the "optimal" decision is not really optimal in the traditional optimization perspective. Instead, it is a satisfactory non-dominated decision, i.e. a decision that is in accordance with the decision maker's system of values and is not dominated by other possible decisions.

Irrespective of whether the set of alternatives A is discrete or continuous, making a decision in a multi-criteria context requires the appropriate aggregation of all the pertinent decision factors, which are referred to as "evaluation criteria" or simply "criteria". Formally, a criterion is a non-decreasing real-valued function that describes an aspect of the global performance of the alternatives and defines how the alternatives are compared to each other.

In making a decision within the multi-criteria context the aggregation of the criteria is a crucial process. This aggregation can be performed in many different ways depending on the form of the criteria aggregation model. Within the MCDA field one can distinguish three main forms of aggregation models: outranking relations (relational form), utility functions (functional form) and decision rules (symbolic form). The construction of an aggregation model is mainly of interest in the case where A is discrete. In such a case the alternatives are clearly identifiable and consequently their performance on each criterion can be specified rather easily. In the case where A is continuous, however, this is not a straightforward process, simply because it is impossible to identify all the alternatives that are relevant to the analysis. In this case special interactive aggregation techniques have been developed in MCDA to allow the efficient search of the solution space.

In all cases, the aggregation of the criteria is performed so as to respect the decision maker's (DM) judgment policy. To ensure that this objective is achieved some information on the preferential system of the DM must be specified, such as the criteria weights. The required preferential information can be specified either through direct procedures in which a decision analyst elicits it directly from the DM, or through indirect procedures in which the DM provides examples of the decision situations that he/she faces and the decision analyst examines them to determine the required preferential parameters which are most consistent with the DM's global evaluations. The latter approach is known in the MCDA field as "preference disaggregation analysis" (Jacquet-Lagrèze & Siskos, 1982, 1983, 2001).

It is recognized that the MCDA models can be classified into three broad categories, or schools of thought: (1) value measurement models in which one decision option may be preferred to another and for which scores are developed initially for each individual criterion, and then synthesized in

order to effect aggregation into higher level preference models; (2) outranking models in which alternative courses of action are compared pairwise, initially in terms of each criterion, in order to identify the extent to which a preference for one over the other can be asserted. In aggregating such preference information across all relevant criteria, the model seeks to establish the strength of evidence favoring selection of one alternative over the other; (3) goal, aspiration or reference level models in which desirable or satisfactory levels of achievement are established for each of the criteria. The process then seeks to discover options, which are in some sense closest to achieving these desirable goals or aspirations. Note that softer methods outside of these three categories exist as well. For example, the even-swap method by Hammond, Keeney and Raiffa (2000), which provides a very practical and understandable way of dealing with value tradeoffs to make a decision.

A number of authors have highlighted the similarities of data envelopment analysis (DEA) and MCDA models, commenting principally from a theoretical perspective on the mathematical structure and methods for solution. Given these similarities it is possible that the two approaches could be viewed as competing. DEA could be described as an approach, which seeks to extract as much as possible from "objective" historical data, without resort to subjectivity. In contrast, MCDA actively seeks to elicit, understand and manage value judgments.

4.1 Value measurement methods

The idea behind value measurement methods is to formulate a quantitative score for every alternative based on an aggregate value judgment of the relevant criteria (see e.g. Belton & Stewart, 2002). This score could then be used to rank or classify alternatives. The most straightforward approach is to score each alternative on every individual criterion, and then calculate a weighted sum of these partial scores based on the DM's judgment of the relevant importance of each criterion.

4.1.1 Multi-attribute value theory

Multi-attribute value theory (MAVT) is an extension of basic value measurement methods that takes into account possible non-linearity of the preference functions. In a first step, a hierarchical value tree, which represents the hierarchy of relevant criteria in scoring different alternatives, has to be constructed. In constructing the value tree, it is important to consider the condition of preferential independence, which means that tradeoffs between different criteria should not depend on any other criteria. Once the alternatives have been determined, the second step consists of constructing a performance table by scoring the different alternatives with respect to the different criteria. If the decision maker feels comfortable with the alternatives and the criteria, the scoring process can be completed by direct assessment. If the decision maker has more difficulties with the scoring process, other scoring methods can be considered, e.g. indirect assessment, using qualitative scales or by pairwise comparison.

The distinctive feature of MAVT is in the elicitation of partial value functions for each criterion, which represent the utility derived by the DM from the performance of an alternative with respect to a single criterion. Deriving the true underlying value function is not straightforward. To help the DM in this complex process, value functions can be derived in an indirect way, for example, via standard differences or via bisection methods.

Next to determining the underlying partial value functions, weights of criteria need to be elicited. Again, different methodologies can be applied. Either the DM feels comfortable assessing the importance of criteria directly, or methods like preference disaggregation analysis are used (Jacquet-Lagrèze & Siskos, 1982, 1983, 2001). In a final step, the overall score for each alternative is calculated aggregating the partial value functions on the basis of the elicited weights for every criterion. Different forms of aggregation are possible (e.g. additive or multiplicative). Note that this final step is not the same as a simple weighted sum, since only the transformed utility values from the partial value functions are used for each criterion, and not the direct scores from the performance table itself.

Similar to MAVT is multi-attribute utility theory (MAUT), which is based on expected utility theory and calls for even stronger assumptions. The main advantage of MAUT is the possibility to take into account uncertainty and risk. However, this makes it even more complex to elicit final scores for the alternatives. Therefore, MAVT is still the preferred methodology in approaching real-life decision problems. For a more complete overview of these techniques, see for example Keeney and Raiffa (1993) and Wallenius et al. (2008) for some more recent accomplishments and future applications.

MAVT/MAUT is of particular interest to our decision problem because of the straightforward interpretation. In essence it can be considered a sophistication of the weighted sum approach using value/utility information, which is easy to understand for every investor. However, future research should consider the trade-off between the ease of understanding and the complexity of the underlying utility functions. Because the input for the model comes from a decision maker, i.e. the investor, simpler utility functions are always preferred to more complex functions to ensure practical implementation. A disadvantage of this method is the lack of a built-in methodology to determine criteria weights. No special software is required for implementation, as MAVT/MAUT can easily be programmed into a spreadsheet package.

4.1.2 Analytic hierarchy process

The analytic hierarchy process (AHP), presented in detail by Saaty (1980), is an elegant approach in its simplicity, for addressing and analyzing discrete alternative problems with multiple conflicting criteria. Like MAVT, the AHP starts by subdividing a problem into a hierarchy of overall objective criteria. As we work to build this AHP hierarchy, we increase our understanding of the problem as a whole. Particular about AHP is the use of pairwise comparisons to elicit the criteria weights from experts. Psychologists argue that it is easier and more accurate to express one's opinion on only two alternatives than simultaneously on all the alternatives. It also allows consistency cross checking between the different pairwise comparisons. Starting at the bottom level of the hierarchy, we conduct pairwise comparisons between the elements immediately below each other. Under real conditions, it is not difficult, based on the condition of transitivity, to identify improperly filled in questionnaires. The AHP method assesses the consistency of each expert's opinions and defines a consistency index (Saaty, 1980, 2005).

One of AHP's strengths is the possibility to evaluate quantitative as well as qualitative criteria and alternatives on the same preference scale. These can be numerical, verbal or graphical. The use of verbal responses is intuitively appealing, user-friendly and more common in our everyday lives

than numbers. Nevertheless, it may also allow some ambiguity in non-trivial comparisons, which has been criticized (Donegan et al., 1992). To derive priorities, the verbal comparisons must be converted into numerical ones. In Saaty's AHP the verbal statements are converted into integers from one to nine. Theoretically there is no reason to be restricted to these numbers and verbal gradation. Although the verbal gradation has been little investigated, several other numerical scales have been proposed. Clearly, the choice of the "best" scale is a very heated debate. Some scientists therefore argue that the choice of scale depends on the person and the decision problem (Harker & Vargas, 1987; Pöyhönen et al., 1997).

AHP still suffers from some theoretical disputes. Rank reversal is surely the most debated problem. This phenomenon is still not fully resolved and may never be because the aggregation of preferences transposed from scales of different units is not easily interpretable and even questionable according to Roy (1996). In this sense, the rank reversal problem is not specific to AHP, but to the normalization of scores. The assumption of preferential independence may also be a limitation of AHP (and other MCDA methods). The analytic network process (ANP), a generalization of AHP with feedbacks to adjust weights, may be a solution. However the decision maker must answer a much larger number of questions, which may be complex (Saaty & Takizawa, 1986). A simplified ANP, while still keeping its proprieties, would be beneficial for a wider adoption of the method (Ishizaka & Labib, 2011). The choice of a hierarchy and a judgment scale is also important and difficult. Problem structuring methods could help in the construction of AHP hierarchies, which is its less formalized aspect (Petkov & Mihova-Petkova, 1997; Petkov et al., 2007).

Several works can be found in the literature relating AHP with finance. Beyond improving the quality of the decisions, the AHP is shown as a useful tool to support the process of examining, justifying, negotiating, and communicating ethical decisions. Pérez-Gladish and M'Zali (2010) already showed how AHP can be implemented to address the problem at hand. Most valuable about the AHP method is the built-in procedure to determine criteria weights using pairwise comparisons. This procedure can be used to complement other methodologies as well. With simple dichotomous utility functions, the MAUT/MAVT approach becomes nested within the AHP method when using pairwise comparisons to determine criteria weights. A common critique on the AHP method is that it is rather simplistic and only useful for its pairwise comparison approach. The use of a method from another school of thought, together with the AHP weighting procedure must therefore be considered in future research as a potential MCDA approach.

4.1.3 Other value measurement methods

Besides MAVT and AHP, other value measurement models exist. Like AHP, MACBETH (Measuring Attractiveness by a Categorical Based Evaluation TecHnique; Bana e Costa & Vansnick, 1994) is based on pairwise comparisons to evaluate alternatives with respect to different criteria. The difference, however, is in the use of an interval scale instead of a ratio scale and the fact that the MACBETH method calls for a much greater number of pairwise comparisons to elicit criteria weights, which makes it less desirable for future implementation in practice. MACBETH is also harder to implement in practice as it can only be achieved through costly software, whereas methods like AHP can be implemented using free software. The linear programming algorithm behind MACBETH does not always yield consistent results and is more black box and harder to understand than most other value measurement methods. Inconsistencies are however easier to

spot in the MACBETH method. Therefore, we suggest that future research looks into the application of both AHP and MACBETH, considering criteria like stability intervals to decide which one works best. To further accommodate the elicitation of utility functions from decision maker preferences, disaggregation methods like UTA (UTilités Additives; Jacquet-Lagreze & Siskos, 1982) and UTADIS (UTilités Additives DIScriminantes; Jacquet-Lagrèze & Siskos, 1982; Zopounidis & Doumpos, 1999) have also been developed, but might be harder to implement for the purpose of our decision problem.

4.2 Outranking methods

4.2.1 ELECTRE

The ELECTRE (ELimination Et Choix Traduisant la REalité) family of methods by Roy (1985) is based on the concept of outranking: "one solution outranks another if it is at least as good as the other in most respects, and not too much worse in any one respect." Typical for the outranking approach is the unicriterion comparison of alternatives based on preference degrees. Such preference degrees express how a decision maker prefers one alternative to another, based on an underlying preference function. A typical ELECTRE analysis yields only an outranking relation of the different alternatives, and no concrete quantitative output. Therefore, it might be slightly less appropriate to build a social performance indicator. Just like AHP, ELECTRE also suffers from the rank reversal issue (Wang & Triantaphyllou, 2008). However, ELECTRE has been successfully implemented in financial research before. For instance, Martel et al. (1988) employ ELECTRE to study the limitations of conventional risk in being able to capture global risk in a portfolio context. Also influenced by ELECTRE is BANK ADVISER by Mareschal and Brans (1991), which has been successful in the banking industry. Members of the ELECTRE family are ELECTRE I, II, III, IS, IV and TRI.

An important limitation of the ELECTRE methodology is that it does not generally yield a scoring output. Therefore, application of the ELECTRE method would call for some additional transformation to obtain the scores or categories we need for further implementation in performance regressions. Different software packages are available to implement this method. ELECTRE is also considered a more outdated technique in comparison to PROMETHEE, which should be taken into account when comparing methods upon implementation. The need for information from the decision maker can also be quite considerable in the ELECTRE methodology.

4.2.2 PROMETHEE

Originally developed by Brans and Vincke (1985), PROMETHEE (Preference Ranking Organization METHod for Enrichment Evaluations) is also based on the theory of outranking relations. The outranking methods include two phases: the construction of an outranking relation, and the exploitation of this relation in order to assist the decision-maker. To capture the outranking relation, Brans et al. (1986) use six types of functions that cover most of the cases occurring in practical applications The basic principles of the PROMETHEE method in relation with other methods of the same field are the following: extension of the notion of criteria, a valued outranking relation and exploitation of the outranking relation. In the PROMETHEE method the valued outranking relation is less sensitive to small modifications and its interpretation is straightforward. The exploitation of the valued outranking relation of the PROMETHEE method refers to the case in

which the alternatives have to be ranked from best to worst. A typical PROMETHEE analysis yields a quantitative output in the form of net flows (valued outranking relation), which represent the relative preference of the DM for one alternative to another. This quantitative output could be transformed to an index, which makes PROMETHEE a viable candidate to build a social performance indicator. Again, rank reversal occurs as an issue when implementing the PROMETHEE method (e.g. Mareschal, Smet, & Nemery, 2008), which is inherent to the normalization of scores as mentioned before. As an example of the application of PROMETHEE in investment decision-making, we refer to Qu, Li and Pei (2012).

Implementation of PROMETHEE is possible using either commercial software packages (e.g. D-Sight) or free academic alternatives (e.g. Visual PROMETHEE). In contrast to the ELECTRE method, PROMETHEE yields a quantitative output in the form of net flows, which might need some transformation/scaling to get a desirable score/category for use in further financial analyses. The PROMETHEE analysis can also be used in conjunction with a graphical visualization procedure termed GAIA (Geometric Analysis for Interactive Aid), which provides a two-dimensional representation of the multi-dimensional problem (Brans & Mareschal, 1990). Such a visual aid is instrumental in presenting results in a compact way and in gaining more insights into how scores and rankings were calculated, and how performances of individual alternatives could be improved. A downside of PROMETHEE, just like ELECTRE and MAUT/MAVT, can be the greater need for input from the decision maker. Members of the PROMETHEE family are PROMETHEE I, II, III, IV, V, VI and GDSS.

4.3 Goal, aspiration or reference level models

TThe last school of MCDA methods covers a wide range of optimization methods, most of which require highly quantifiable inputs. The advantage of goal programming (GP) and aspiration methods is that multiple criteria can be incorporated into a model that can be solved using conventional (single criterion) optimization software. The disadvantage is that information about the decision maker's preferences is required a priori in the form of priority levels, importance weights, and goal target values. GP models can be divided into linear goal programming models, interactive multiple goal programming (IMGP) and interactive sequential goal programming (ISGP). These kinds of models are especially useful when the set of alternatives is continuous. For example, GP might come in handy when determining the optimal funds-of-funds strategy or when managing a portfolio of socially responsible investments (e.g. Hallerbach et al., 2004).

Next to the optimization methods that are also found in the broader field of operations research (OR), this third school of methods also includes reference level models like TOPSIS (Technique of Order Preference Similarity to the Ideal Solution). The central principle of TOPSIS is very simple, and also allows for the use of qualitative criteria, as long as they can be translated onto a numeric scale (Behzadian et. al, 2012). With TOPSIS, the distance from every considered alternative to a theoretically defined worst and ideal solution is calculated, using a particular distance function (e.g. Euclidean). One alternative is then preferred to another when it is both closer to the ideal solution, and further away from the worst solution. The best alternative is the one that minimizes the distance from the ideal and maximizes the distance from the worst solution.

Rather than the typical OR optimization methods, which call for continuous sets of alternatives,

TOPSIS provides a simple and flexible way to compute preference scores for discrete alternatives. With respect to our decision problem of scoring SRI funds, this seems to be the only appropriate method. TOPSIS provides a really straightforward and easy to understand way of obtaining SRI scores, which can be used in further financial performance research without need for adaptation. The amount of information needed from the decision maker is generally also quite low and no special software packages are required for implementation. However, TOPSIS sometimes yields illogical results, particularly given extreme performances on different criteria. The most appropriate distance function to calculate the overall scores can also be a topic of discussion. For an example of an investment decision application of TOPSIS we refer to Tsao (2003).

5 Application of MCDA framework to SRI issues

All of the challenges presented in section 3 can be addressed using MCDA methodologies. As pointed out before, the first three challenges are similar in the sense that they can be met by a social responsibility indicator to score and classify mutual funds. The fourth challenge, which considers the untapped potential of the retail side of the SRI market, requires a MCDA framework that can be tailored to the needs of individual clients to better understand the SRI concept and their social investment preferences.

5.1 Social performance indicator

Using the MCDA framework, a social performance indicator for mutual funds can be built. Instead of using a dichotomous classification, an indicator would make it possible to discriminate between mutual funds on the basis of their social performance in a more continuous way. This can be achieved either by calculating a social performance score, or by allocating mutual funds to different categories based on their social performance. Both options can be achieved using MCDA methods. Having a social performance indicator would allow scholars to have an even better look at the corporate social versus corporate financial performance relationship. Regulators could use the indicator to assess the design of SRI funds. Banks can be assisted by such an indicator to improve their line of SRI products. Overall, the indicator would increase the transparency and clarity of the supply of SRI products.

In developing the indicator, it is necessary to first define a relevant and consistent set of criteria for assessing the social responsibility of a mutual fund, taking into account views from different stakeholders. This set of criteria needs to be comprehensive and yet as compact as possible. The assumption of preferential independence also needs to be fulfilled. A useful tool to define the set of criteria is the value-focused thinking approach by Keeney (1992), which structures the process of defining relevant criteria. The definition of a valid set of criteria requires the collaboration of an expert panel representing the different stakeholders involved with the SRI decision process. To help find the relevant groups of stakeholders, Checkland's CATWOE model might be used (Checkland & Scholes, 1990). Applying the AHP to score mutual funds on social performance, Pérez-Gladish and M'Zali (2010) already developed a set of criteria. However, they only used one expert to establish the list of criteria, so future research needs to consult representatives from all different interest groups to ensure that the set of criteria is comprehensive and consistent with the different views from the field of SRI. A first example of a comprehensive set of criteria is displayed in Table 1.

Goal	Criteria	Subcriteria	Sub-subcriteria
Social performance indicator	Screening process and consistency	C1. Priority screening process	
		Independent data gathering and analysis of sustainability	C2. Data gathering and analysis of sustainability by independent external specialists (e.g. EIRIS) C3. Incorporation of SRI principles established by reputable organizations (e.g. UN SRI, Febelfin) C4. Information from stakeholders and relevant NGOs
		Positive selection criteria	C5. Best-in-class approach for criteria with respect to ESG criteria C6. Use of sector specific positive criteria C7. Investment is principally (>75%) in companies that invest in sustainable technologies
		Negative selection criteria	C8. Use of categorical rejects C9. Assessment by means of negative criteria C10. Exclusion of unsustainable technologies
		Monitoring and updates	C11. A research teams checks legal and regulatory developments, trends and behavior of companies such that criteria are in line with recent societal developments C12. Monitoring if portfolio is consistent with defined criteria (continuously, sector specific or occasion specific)
		Dialogue	C13. Companies are informed about conclusions selection methodology C14. Active engagement policy (constructive and critical dialogue with companies in portfolio) C15. Active voting policy (voting at companies' shareholder meetings)
	Transparency and control	Transparency	C16. Release of qualitative information about the screening process (e.g. applied screens) C17. Release of quantitative information about the screening process (e.g. scores) C18. Release of current portfolio C19. Compliance with external transparency guidelines (e.g. Eurosif/Belsif)
		C20. Board of experts	

Table 1: Example set of SRI criteria to build an SRI fund-level performance indicator

Once the consistent set of criteria is found, the alternatives need to be considered. This set of alternatives will depend on the application of the social performance indicator. For academic purposes, a scholar might have a good reason to consider only a small subset of mutual funds. For the purpose of government regulation, the set of alternatives might include all the mutual funds offered in a certain country or region. In applying the social performance indicator in the banking industry, the set of alternatives might consist of different potential mutual funds that the bank wants to release in the near future. Given that the set of alternatives can be change for the different purposes, it is of keen importance to build the social performance indicator in such a way that it is independent from the underlying set of alternatives. The choice of decision alternatives will also depend on the availability of data and transparency with respect to the investment process. For example, countries like Belgium, the Netherlands and the Scandinavian region have already implemented a great number of transparency guidelines, which makes it easier to measure social responsibility of investment vehicles in those countries.

The next step is to choose a particular MCDA model to build the social performance indicator. From the school of value measurement methods, different methodologies can be applied to the problem at hand. AHP has been successfully implemented in the past, although given a limited number of experts determining the set of criteria. MAVT can also be applied, as it is a more general form of the AHP, on the condition that enough information can be collected from the expert panel about the form of the partial utility functions. Another value measurement method that is feasible for this problem is MACBETH, which is similar to AHP but uses categorical instead of ratio scales. In the outranking school of MCDA methods, both PROMETHEE, together with its visual aid GAIA, and ELECTRE can be used as a tool to classify mutual funds based on outranking relations. The third class of models — goal, aspiration and reference-level methods — is generally less appropriate as a social performance indicator, as it typically requires an association of every criterion with quantitative and measurable attributes, which is not the case with some of the softer criteria related to SRI. However, the TOPSIS reference-level methodology allows for a very simple and efficient calculation of social performance scores without requiring too much quantitative input. The goal programming methods can be used either ex-ante to narrow down a broad set of alternatives, or ex-post to find the optimal way of implementing a fund-of-funds strategy. For all of these methodologies, it is important to consider underlying assumptions (e.g. rank reversal, preferential independence) and to end up with a straightforward quantitative output that can be used in performance regressions. This last condition might be harder to fulfill for the outranking methods.

To implement these different methodologies, and to find the most robust and qualitative one to score mutual funds on their social performance, a lot of information is needed as an input. This information should be collected from the expert panel of representative stakeholders. Therefore, a key success factor is to collect this information in the most efficient way possible, as it seems unlikely that such an expert panel can be consulted regularly. We suggest that a questionnaire, consisting of all the relevant questions to implement the different methodologies, is constructed. This might not be straightforward, as the different methods require quite different sets of information.

Once the criteria have been defined, a sample set of alternatives is constructed and all the necessary information is collected, different methods can be implemented to build the social performance indicator. Instead of choosing only one of the methods ex-ante, we suggest that future research has a look at every feasible model that can be applied to the problem. After the models are built,

robustness and sensitivity analyses on these different models can help to determine which model is most qualitative and robust in judging the social performance of a mutual fund. These robustness checks can either be executed using built-in software modules that alter input values and consider the changes in output, or using more statistical procedures like Gini's concept of transvariation (e.g. Van den Bossche et al. 2010). More concretely, future research could also look into the stability of obtained SRI scores given changes in the underlying criteria weights. For every criterion, a stability interval can be calculated, indicating to what extent the weight of a criterion can be changed without resulting in a change of SRI rankings. Ideally, the applied method should be highly robust for underlying changes in the weights, as these are very hard to elicit or determine for decision makers, i.e. investors. Furthermore, feedback from the expert panel can be used to determine what kind of model is most in line with the expertise from different stakeholders. Depending on the particular application of the model, several other criteria can be considered in assessing its appropriateness as well. For example, one should always take into account ease of implementation, as the ideal SRI indicator will be implemented on a larger scale both in the professional industry (e.g. asset management) and academia (i.e. future research). Models that might be more advanced from a technical point of view might be harder to implement because of a greater need for detailed input from the decision maker (e.g. elicitation of utility functions). Also, more advanced methods might be harder to understand by the decision makers, which might complicate the implementation. The MCDA method that provides the best trade-off between robustness from a technical point of view (i.e. has the largest average stability interval), and convenience from a practical point of view (i.e. ease of implementation and understanding) can then be applied as a social performance indicator to address the academic, regulation and commercial bank challenges presented in section 3.

5.2 "Green" MiFID

To address the underdeveloped state of the retail end of the SRI mutual fund market, the MCDA framework could also be adopted by banks to help their customers better understand the variety of SRI mutual funds being offered. The main difference with the social performance indicator is that this tool needs to be tailored to the individual needs of investors. Instead of looking for a consensus on the set of criteria and their importance from an expert panel of representative stakeholders, individuals now need to provide input that represents their preferences. Ideally, the best performing model found when constructing the social performance indicator could be used on the individual level as well, just by replacing input from the expert panel by individual information provided by investors. However, chances are that this approach would be too technical for a straightforward implementation in standard investment advisory practice. Therefore, we suggest that future research explores to what extent it is possible to transform the principles behind the best possible MCDA methodology into a standardized questionnaire that would be better suited to help clients in getting to know SRI products and their preferences towards these products in a better way. As an interesting example, we point to the European Commissions' directive called MiFID, which obliges banks to let investors complete a questionnaire to get to know their personal investment profile. As a result, a client interested in investing is sorted into a certain category, which comes with certain rules of protection (Davies et al., 2006). The directive provides transparency and tries to protect retail investors from investing in products that are incompatible with their preferences. In a similar way, we propose a so-called "green" MiFID to be created. Retail investors interested in SRI would need to complete a survey, the questions of which are based on the underlying MCDA framework,

to get more insight into their preferences with regard to social investing.

It is worth noting that this "green" MiFID is compatible with the social performance indicator in opening up the retail side of the SRI market. The social performance indicator at the supply side of the SRI market increases product transparency as it accommodates the comparison of different mutual funds with respect to their performance on ESG criteria. The "green" MiFID provides retail investors with a better insight into their personal SRI profile. Together, these two measures make it possible for individual investors interested in SRI to get a clear understanding of what products are best suited to match their preferences, which exactly addresses the main cause of the SRI retail market remaining underdeveloped.

6 Conclusion

We highlight four challenges that are facing SRI and a framework that could be used in future research to address these. A first group of challenges revolves around the need for a social performance indicator that is able to score and/or classify mutual funds with respect to their social responsibility. Such an indicator could enable scholars to better examine the relative performance of SRI mutual funds; it can help the regulatory authorities in developing a certificate for genuinely socially responsible mutual funds; and it accommodates the process for banks to develop new and transparent SRI products. Another challenge concerns the underdevelopment of the retail side of the SRI market. To tap the potential at this side of the market, there is the need for a tool that can help investors to better understand the SRI products and their own preferences with respect to these products.

Given these two groups of challenges, we point to multi-criteria decision analysis (MCDA) as the methodological framework that could be used in future research to develop appropriate solutions. Different MCDA methodologies can be implemented to build a social performance indicator that would address the first group of challenges. Using robustness analyses and feedback from an expert panel representing the main stakeholders in the SRI field, it should be possible to determine the best possible MCDA method to score and classify mutual funds. This would benefit academic research, the regulatory burden of governments and the efforts of banks to create new SRI products. From this MCDA indicator, it should also be possible to create a simple tool based on a standard questionnaire that can help investors to better understand their own preferences with regard to the social responsibility of financial products. Such a questionnaire is comparable to the current MiFID framework that is used in banks to help clients understand their risk preferences when investing in financial products.

Note that the indicator on the one hand, and the "green" MiFID tool on the other hand, are not unrelated. Ideally, they should be aligned as to stimulate the further growth of the SRI market. The indicator, among other things, can create more transparency in the supply of SRI financial products and makes it easier to discriminate between SRI mutual funds. The investor tool can help learn potential investors about their social performance preferences. Together, these two tools have the potential to open up new perspectives for the retail side of the SRI market, which could further add to the growth of the SRI investment universe.

References

Bana e Costa, C. and Vansnick, J. (1994). MACBETH - An interactive path towards the construction of cardinal value functions. *International Transactions in Operational Research*, 1(1):107–114.

Barnett, M. L. and Salomon, R. M. (2006). Beyond dichotomy : The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, 27(11):1101–1122.

Bauer, R., Koedijk, K., and Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7):1751–1767.

Behzadian, M., Otaghsara, K., Yazdani, M., and Ignatius, J. (2012). A state-of-the-art survey of TOPSIS applications. *Expert Systems with Applications*, 39(17):13051–13069.

Belton, V. and Stewart, T. J. (2002). *Multiple criteria decision analysis: An integrated approach*. Kluwer Academic Publishers, Boston.

Bragdon, J. and Marlin, J. (1972). Is pollution profitable? *Risk Management*, 19(4):9–18.

Brans, J. and Mareschal, B. (1990). The PROMETHEE methods for MCDM; the PROMCALC, GAIA and BANKADVISOR software. In Bana e Costa, C., editor, *Readings in multiple criteria decision aid*, pages 216–252. Springer Verlag, Berlin.

Brans, J. and Vincke, P. (1985). A preference ranking organisation method: the PROMETHEE method for multiple criteria decision-making. *Management Science*, 31(6):647–656.

Brans, J., Vincke, P., and Mareschal, B. (1986). How to rank and how to select projects: The PROMETHEE method. *European Journal of Operational Research*, 24(2):228–238.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.

Checkland, P. and Scholes, J. (1990). *Soft systems methodology in action*. John Wiley & Sons, Chichester.

Davies, R., Dufour, A., and Scott-Quinn, B. (2006). The MiFID: Competition in a new European equity market regulatory structure. In Ferrarini, G. and Wymeersch, E., editors, *Investor protection in Europe: Corporate law making, the MiFID and beyond*, chapter 6, pages 161–198. Oxford University Press, Oxford.

Donegan, H., Dodd, F., and McMaster, T. (1992). A new approach to AHP decision-making. *The Statistician*, 41(3):295–302.

European Fund and Asset Management Association (2013). Asset management in Europe: Facts and figures. Technical report, Brussels.

Eurosif (2012). European SRI Study. Technical report, Brussels.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Ferson, W. E. and Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, 51(2):425–461.

Forum for Sustainable and Responsible Investment (2012). Sustainable and responsible investing trends in the US. Technical report, Washington, D.C.

Freeman, R. (1984). *Strategic management: A stakeholder approach*. Pitman/Ballinger, Boston.

Friedman, M. (1970). The social responsibility of business is to increase its profits.

Geczy, C. C., Stambaugh, R. F., and Levin, D. (2003). Investing in socially responsible mutual funds. Working paper.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510.

Hallerbach, W., Ning, H., Soppe, A., and Spronk, J. (2004). A framework for managing a portfolio of socially responsible investments. *European Journal of Operational Research*, 153(2):517–529.

Hamilton, S., Jo, H., and Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, 49(6):62–66.

Hammond, J., Keeney, R., and Raiffa, H. (2000). The even-swap method for multiple objective decisions. In Haimes, Y. and Steuer, E., editors, *Research and practice in multiple criteria decision making*, pages 1–14. Springer, Berlin.

Harker, P. and Vargas, L. (1987). The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Management Science*, 33(11):1383–1403.

Hoggett, J. and Nahan, M. (2002). Ethical investment - Deconstructing the myth. *IPA Review*, 54(3):3–6.

Hollingworth, S. (1998). Green investing: A growing concern? *Australian CPA*, 68(4):28–30.

Ishizaka, A. and Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, 38(11):14336–14345.

Jacquet-Lagrèze, E. and Siskos, J. (1982). Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *European Journal of Operational Research*, 10(2):151–164.

Jacquet-Lagrèze, E. and Siskos, J. (1983). *Methode de decision multicritere*. Editions Hommes Et Techniques, Paris.

Jacquet-Lagrèze, E. and Siskos, J. (2001). Preference disaggregation : 20 years of MCDA experience. *European Journal of Operational Research*, 130(2):233–245.

Keeney, R. (1992). *Value-focused thinking: A path to creative decision making*. Harvard University Press, Cambridge.

Keeney, R. and Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value tradeoffs*. John Wiley & Sons, New York.

Koellner, T., Weber, O., Fenchel, M., and Scholz, R. (2005). Principles for sustainability rating of investment funds. *Business Strategy and the Environment*, 14(1):54–70.

KPMG and ALFI (2013). European Responsible Investing Fund Survey 2013. Technical report, Luxembourg.

Mackey, A., Mackey, T. B., and Barney, J. B. (2007). Corporate social responsibility and firm performance: Investor preferences and corporate strategies. *Academy of Management Review*, 32(3):817–835.

Mareschal, B. and Brans, J. (1991). BANKADVISER: An industrial evaluation system. *European Journal of Operational Research*, 54(3):318–324.

Mareschal, B., De Smet, Y., and Nemery, P. (2008). Rank reversal in the PROMETHEE II method: Some new results. In *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, pages 959 – 963, Singapore.

Margolis, J. D. and Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly*, 48(2):268–305.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Martel, J., Khoury, N. T., and Bergeron, M. (1988). An application of a multicriteria approach to portfolio comparisons. *Journal of the Operational Research Society*, 39(7):617–628.

Moskowitz, M. (1972). Choosing socially responsible stocks. *Business and Society Review*, 1:71–75.

Orlitzky, M., Schmidt, F. L., and Rynes, S. L. (2003). Corporate social and financial performance: A Meta-analysis. *Organization Studies*, 24(3):403–441.

Pérez-Gladish, B. and M'Zali, B. (2010). An AHP-based approach to mutual funds' social performance measurement. *International Journal of Multicriteria Decision Making*, 1(1):103–127.

Petkov, D. and Mihova-Petkova, O. (1997). The analytic hierarchy process and systems thinking. In *13th international MCDM conference*, pages 243–252, Cape Town. Springer.

Petkov, D., Petkova, O., Andrew, T., and Nepal, T. (2007). Mixing multiple criteria decision making with soft systems thinking techniques for decision support in complex situations. *Decision Support Systems*, 43(4):1615–1629.

Pöyhönen, M., Hamalainen, R., and Salo, A. (1997). An experiment on the numerical modelling of verbal ratio statements. *Journal of Multi-Criteria Decision Analysis*, 6(1):1–10.

Qu, S., Li, H., and Pei, Y. (2012). Decision making in investing: Application of interval-PROMETHEE based on the composite weight. *Journal of Theoretical and Applied Information Technology*, 45(1):239–245.

Renneboog, L., Ter Horst, J., and Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, 32(9):1723–1742.

Renneboog, L., Ter Horst, J., and Zhang, C. (2011). Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation*, 20(4):562–588.

Roy, B. (1985). *Méthodologie multicritère d'aide à la décision*. Economica, Paris.

Roy, B. (1996). *Multicriteria methodology for decision analyis*. Kluwer Academic Publishers, Dordrecht.

Saaty, T. (1980). Analytic hierarchy process. *Decision Analysis*, 50(1):579–606.

Saaty, T. (2005). The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making. In Figueira, J., Greco, S., and Ehrgott, M., editors, *Multiple criteria decision analysis: State of the art surveys*, pages 345–408. Springer.

Saaty, T. and Takizawa, M. (1986). Dependence and independence: From linear hierarchies to nonlinear networks. *European Journal of Operational Research*, 26(2):229–237.

Sauer, D. A. (1997). The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini equity mutual fund. *Review of Financial Economics*, 6(2):137–149.

Schwartz, M. S. (2003). The "ethics" of ethical investing. *Journal of Business Ethics*, 43(3):195–213.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Steuer, R. E. and Na, P. (2003). Multiple criteria decision making combined with finance: A categorized bibliographic study. *European Journal of Operational Research*, 150(3):496–515.

Tippet, J. (2001). Performance of Australia's ethical funds. *Australian Economic Review*, 34(2):170–178.

Tsao, C. (2003). Evaluating investment values of stocks using a fuzzy TOPSIS approach. *Journal of Information and Optimization Sciences*, 24(2):373–396.

Van den Bossche, F., Rogge, N., Devooght, K., and Van Puyenbroeck, T. (2010). Robust corporate social responsibility investment screening. *Ecological Economics*, 69(5):1159–1169.

Waddock, S. A. and Graves, S. B. (1997). The corporate social performance - financial performance link. *Strategic Management Journal*, 18(4):303–319.

Wallenius, J., Dyer, J., Fishburn, P., Steuer, R., Zionts, S., and Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science*, 54(7):1336–1349.

Wang, X. and Triantaphyllou, E. (2008). Ranking irregularities when evaluating alternatives by using some ELECTRE methods. *Omega*, 36(1):45–63.

Zopounidis, C. and Doumpos, M. (1999). Business failure prediction using the UTADIS multicriteria analysis method. *Journal of the Operational Research Society*, 50(11):1138–1148.

Zopounidis, C. and Doumpos, M. (2002). Multi-criteria decision aid in financial decision making: Methodologies and literature review. *Journal of Multi-Criteria Decision Analysis*, 11(4-5):167–186.