UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI
"M.FANNO"

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE
IN ECONOMICS AND FINANCE

TESI DI LAUREA

"ASSET ALLOCATION BENEFITS OF SRI: TESTING EFFICIENCY AND THE IMPACT ON PERFORMANCES"

RELATORE:

CH.MO PROF. MASSIMILIANO CAPORIN

LAUREANDO: EUGENIO ZILIOTTO

MATRICOLA N. 1082750

ANNO ACCADEMICO 2016 – 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

____

Contents

List of Figures

List of Tables

Introduction

Socially Responsible Investment (SRI) is an investment approach that combines investors' financial objectives with their concerns about environmental, social and governance (ESG) issues (Eurosif, 2016). Investors and asset managers achieve this goal by implementing SRI strategies. These involve avoiding companies engaged in activities considered harmful to society, selecting the best performing ones in terms of ESG criteria or collaborating with them in order to influence their behaviour and improve their ESG performance.

The origins of the phenomenon can be traced back to the eighteenth century, but it is in the last two decades that, especially in Europe and the United States, these SRI practices experienced an outstanding growth in terms of adoption rates. This expansion can be attributed to a strong demand from institutional investors, the launch of numerous international initiatives, legislations and an increasing attention from non-governmental organizations (NGOs). While SRI mainly developed as an equity focused investment style, its principles are now applied to a variety of asset classes like bonds, commodities and real estate.

One of the main obstacles that, especially in less recent years, has kept investors from adopting a socially responsible investment approach has been the belief that social and financial performances are negatively correlated. Companies with good Corporate Social Responsibility (CSR) records where thought to face competitive disadvantages, resulting in lower profits and lower returns for investors. Additionally, what most of the SRI strategies have in common is the exclusion of those assets that does not comply with the ethical and moral standards pursued by investors. As well known from Modern Portfolio Theory (Markowitz, 1952), by reducing the number of assets from their investment pool (or universe), investors give up on diversification opportunities, resulting in losses in terms of mean-variance efficiency.

The existing literature is abundant in studies aimed at verifying whether socially responsible investors are forced to pay a price in terms of financial performance in order to pursue their social goals. Most of them focused only on the first aspect of the problem; verifying whether the stocks of socially responsible companies underperform those of conventional ones. Findings, overall, were consistent with the so-called "no effect" hypothesis (Statman & Glushkov, 2009), according to which the social aspects does not significantly influence stock performances. Only few studies, instead, based on mean-variance analysis, verified the effects of the exclusion of non socially responsible assets from the investment universe. Most notably, those by Galema, Scholtens, and Plantinga (2009) and Herzel, Nicolosi, and Stărică (2012) employed spanning tests (Huberman & Kandel, 1987) and data relative to the North American equity market to verify whether the efficient frontier available to socially responsible investors significantly differs from that available to traditional investors. Their results showed that, when not imposing additional constraints, the hypothesis of spanning was almost always rejected, mostly because of losses in terms of risk reduction opportunities. These differences were no more significant when introducing restrictions to short-selling (positivity constraints), a condition faced by many investors.

Inspired by the aforementioned research, the main goal of this master thesis was to test the ex-ante mean-variance efficiency of equity portfolios to which SRI exclusions were applied. In particular, efficiency was tested on a rolling basis and adopting the methodology from Jobson (1982) and Gibbons, Ross, and Shanken (1989). The rolling approach, opposed to the one used in similar studies, which involved testing only a single time frame, allowed verifying whether the effects of the exclusions are constant through time or depend on market conditions. Also, instead of focusing on the North American equity market, like most of the existing SRI research, the data used was relative to the European region, more specifically, to the constituent companies of the STOXX Europe 600 index. The implementation of the SRI strategies, a combination of negative and positive screening, was possible thanks to Asset4's ESG data, retrieved, as for almost all empirical data used in this research, from Thompson Reuter's Datastream.

A significant part of this research was also aimed at testing the difference in realized performance, both in terms of Sharpe ratio and variance, of screened and non screened portfolios, obtained simulating different types of dynamic allocation strategies. The process involved performing the tests proposed by Ledoit and Wolf (2008, 2011).

Lastly, as in Herzel et al. (2012), the impact of excluding companies based on the three dimensions of sustainability—environment, social and governance—were singularly tested, repeating, for each one, the above mentioned methodologies.

Document Structure

The first chapter consists in an overview of the SRI phenomenon. After a first section reserved for the definitions, the chapter proceeds with the history of the phenomenon, from its roots, through the evolution phases, up to its current state, defining the demand drivers and listing the major players involved (asset owners, rating agencies, indices providers and NGOs). The chapter ends by describing the main strategies employed by socially responsible investors.

Chapter 2 contains a review of the existing literature. The publications here reviewed are all aimed at studying the difference in performance between SRI and conventional forms of investments. This chapter also contains two sections to introduce the main hypotheses and the methodologies most commonly employed to verify them.

Chapter 3 contains the methodology used in this research. It starts by briefly presenting Modern Portfolio Theory and the notions of efficient frontier and optimal portfolios. It then continues describing the test procedures employed for the empirical research, in particular: mean-variance efficiency test (with and without short-selling restrictions), Sharpe ratio and variance tests.

The fourth chapter thoroughly describes every single step of the empirical research. These include data retrieval and processing, the implementation of the SRI strategies, a descriptive analysis and the testing procedure. Results are then displayed and discussed with the aid of plots and tables.

The last chapter is reserved for the conclusions and recommendations for future research.

Chapter 1

Socially Responsible Investment

1.1 Definitions

According to Louche and Lydenberg (2010), Socially Responsible Investment (SRI) can be defined both as a *product* and a *practice*. It is a product in the sense that investors acquire, hold and sell stocks of companies that perform well in terms of environment, social and governance (ESG) factors as well as ethical factors. At the same time it is a practice through which investors select companies with good Corporate Social Responsibility (CSR) records and engage with the same companies in order to influence their behaviour and help improve their ESG performance.

The heterogeneity that characterises SRI has lead to the usage of many different terms to describe it; ethical investing, sustainable investing, green investing, ESG investing, impact investing, are just a few of them. Although some authors proposed to make a distinction between the many natures of SRI and to identify what can be considered as such (Sparkes, 2001), what most academics and practitioners seems to agree with is that the essence of SRI is, as stated by Sandberg et al. (2009), the "integration of certain non-financial concerns, such as ethical, social or environmental, into the investment process" (p. 521) and the pursuit of *long-term value creation* (Louche & Lydenberg, 2010).

Finding a valid definition for Socially Responsible Investment was not one of the main goals of this study. Instead, it was crucial to take an impartial stance on the phenomenon, free of any personal considerations and relying, instead, on the work of accredited authors and organisations. This chapter serves this purpose by describing the evolution of the SRI phenomenon, from its first manifestations up to current years.

1.2 Origins and Evolution

The origins of the SRI phenomenon can be traced back to the eighteenth century when religious institutions, such as the Society of Friend (Quakers) and the Methodists, were the first to apply social screening to their investments, trying to reconcile the values they preached with the reality of the business activity (Kinder & Domini, 1997; Louche, Arenas, & van Cranenburgh, 2012). One of the founders of the Methodist Church, John Wesley, in one of his sermons, "The use of Money", going as far as identifying specific products and businesses to avoid, admonished his people to abstain from making money and profiting by hurting others (Domini, 2001). Major players in the North American financial community of the first half of the 20th century, although not directly related to religious groups, followed policies of not investing in stocks of alcohol, tobacco and gambling companies (so-called "sin stocks") basing these decisions on their ethical and moral principles (Louche & Lydenberg, 2006). This is the case of Pioneer Investments, who, in 1928, launched what is considered to be the first SRI mutual fund. Louche and Lydenberg (2010) identified this period as the *Roots phase* of the SRI phenomenon.

From the 1970s up to late 1980s, in what can be defined as the *Development phase*, mostly in the United States, SRI started to evolve from an exclusive religious and faith based activity to become a more widespread process of promotion of corporations' social responsibility (Sparkes, 2001). Up until then, religious affiliations practised responsible investing by simply avoiding questionable companies, while the modern forms of SRI, emerged in the United States and Europe in those years, began to place increasing importance on the influence that investors, through shareholder activism, could have on the behaviour of corporations with regards of social and environmental issues (Louche & Lydenberg, 2006). These were also years characterized by political and protest movements (anti-Vietnam War in the United States and anti-apartheid in South Africa) that inevitably started to become a major drive in responsible investing (Louche & Lydenberg, 2006). The Pax fund, one of the first U.S. retail SRI mutual fund, launched in 1971 inspired by investors' fears of profiting from the Vietnam War (Sparkes, 2001).

Europe too, in this period, experienced the birth of the first SRI mutual funds. In 1965 the Church of Sweden established the *Ansvar Aktiefond Sverige*, the first European ethic fund (Louche et al., 2012). While, in 1984, in the UK, Friends Provident, a Quaker-affiliated insurance company, launched the Stewardship Unit trust (Louche & Lydenberg, 2006).

The 1990s where characterised by the *Transition phase*. Environmental concerns began to increase in importance (see The Kyoto Protocol, 1997) and, especially in Europe, numerous *green funds* emerged stressing the importance of identifying positive sectors and activities related to the environment (renewable energy, clean technologies, etc.). In this period, the number of agencies providing ESG ratings and the number of SRI indices experienced a rapid growth (Louche & Lydenberg, 2010).

A constantly growing interest by institutional investors is what mainly determined the *expansion phase*, during the first decade of the twentieth century. As SRI was losing part of its activist and religious connotations, new investment strategies, such as a Best-in-Class approach, started to be employed. The increasing accessibility and quality of companies' CSR reports, combined with different legislations and regulations put in place, especially in Europe, requiring pension funds to disclose their degree of ESG integration in their investment decisions, helped SRI to find acceptance in the mainstream financial community, becoming one of the fastest growing areas in finance (Sandberg et al., 2009).

Louche and Lydenberg (2010) identified the launch, in 2006, of the Principles for Responsible Investment (PRI) initiative as the event that marked the beginning of the *Mainstreaming phase*.

1.3 Current State

In recent years, the SRI phenomenon experienced a phase of outstanding growth. According to the *Global Sustainable Investment Review* from GSIA[1](2014), the Responsible Investment market grew from the 13.3 trillion dollars of the beginning of 2012 to the 21.4 trillions in the first months of 2014, reaching 30.2% of the professionally managed assets.

In those two years the country that experienced the fastest growth were the United States

Table 1.1: Growth of SRI assets by region (as of late 2014).

	2012		2014		
	SRI Assets	% of Total AuM	SRI Assets	% of Total AuM	Growth
Europe	$ 8,758	49.0%	$ 13,608	58.8%	55%
United States	$ 3,740	20.2%	$ 6,572	31.3%	76%
Canada	$ 589	11.2%	$ 945	17.9%	60%
Australia/NZ	$ 134	12.5%	$ 180	16.6%	34%
Asia	$ 40	0.6%	$ 53	0.8%	33%
Global	$ 13,261	21.5%	$ 21,358	30.2%	61%

Notes: Asset values are expressed in billions.
AuM: assets under mangement.
Source: GSIA (2014).

[1]The Global Sustainable Investment Alliance is a collaboration of membership-based sustainable investment organizations around the world. Its mission is to deepen the impact and visibility of sustainable investment organizations at the global level. Members of the GSIA are: European Sustainable Investment Forum (Eurosif), Responsible Investment Association Australasia (RIAA), UK Sustainable Investment and Finance Association (UKSIF), Forum for Sustainable and Responsible Investment (US SIF), Dutch Association of Investors for Sustainable Development (VBDO).

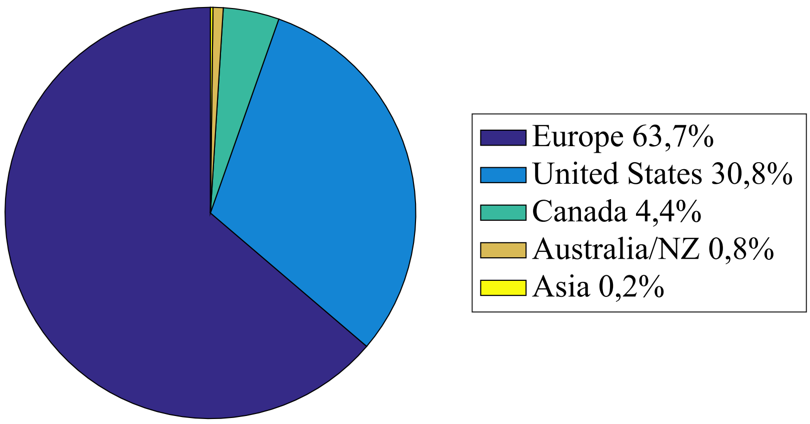


Figure 1.1: Proportion of global SRI assets by region (as of late 2014). Data source: GSIA (2014).

(76% increase), followed by Canada (60% increase) and Europe (55%).

These three regions combined accounted for 99% of responsible investments, globally. Europe was the main contributor with 13.6 trillion dollars of SRI assets under management (63.7% of the global total). Australia and Asia, although experiencing an average growth of 33%, accounted for only 1% of the global SRI market (0.2% for Asia).

1.3.1 Drivers of SRI Demand

The European Fund and Asset Management Association (EFAMA), on its *Report on Responsible Investment* (2014), identified the main drivers of SRI: strong demand from institutional asset owners, international initiatives, legislation (especially at the European level), and an always increasing attention from non-governmental organizations (NGOs) and media.

Furthermore, following Paredes-Gazquez et al. (2014), the main driving forces of ESG integration can be classified into 3 main groups: market pressures, group pressures and institutional pressures:

- Market pressures are the external forces promoting ESG integration and are mainly led by investors and analysts.
- Group pressures refer to initiatives undertaken by members of the financial market to integrate ESG information in the investment process.
- Institutional forces are those exerted by non-members of the financial market. These can be divided in “light forces”, such as voluntary adopted initiatives, or “strong forces” like regulation.

In their study, the authors also identified the barriers to integration: technical impediments and internal and external conventions.

Technical impediments are mainly related to the problems and difficulties in obtaining and processing ESG information. Rating agencies and indices providers play a major role in helping investors overcome these issues.

The term conventions here refers to common market practices and beliefs. The short-term focus and the search for quick returns by investors, coupled with a widespread belief that social and financial performance are negatively related, may have been an obstacle to the integration of ESG issues in the investment processes. Many studies (some of which are reviewed in the next chapter), in contrast, have proved that responsible investing does not destroy value, neither for companies nor for investors.

1.3.2 Asset Owners and Asset Classes

In the asset owners category it is crucial to distinguish between retail and institutional investors. Retail investors, active in the SRI field, are those individuals that, driven by their ethical and moral principles, seek to invest in companies with positive social and environmental records. Together with religious organizations, they were one of the driving forces of the SRI movement during its *development phase* (Louche & Lydenberg, 2006). Institutional investors, on the other hand, are large asset owners, such as pension funds and insurance companies, which, since the late 1990s have become a major factor in the development of the SRI phenomenon.

As of 2014, institutionally held assets accounted for 86.9% of the SRI market (Europe, U.S. and Canada only). Retail investing grew from 10.7% of 2010 to 13.1% of 2014. Europe is now experiencing a significant increase in favour of retail investors who, in late 2015, held 22% of SRI assets (against the 4% of 2013). This is mainly due to the launch of new products and a growing trend to focus on private clients, like High Net Worth Individuals (Eurosif, 2016).

While SRI mainly developed as an equity focused investment style, its recent expansion has lead asset owners to experiment and try to apply SRI principles to other assets classes. The most notable ones are bonds (mostly corporate but also sovereign), real estate, cash and commodities.

Focusing on the European SRI market, equity accounted for over 30% of SRI assets in 2015, an important decrease from the 50% of 2013. Bonds accounted for almost 64%, of which 51% were corporate and 41% sovereign. The remaining were real estate, private equity, hedge funds and commodities. The staggering growth experienced by the bond asset class is correlated to the recent surge in popularity of Green bonds[2].

[2]According to the International Capital Market Association's (ICMA) definition, Green Bonds are "any type of bond instrument where the proceeds will be exclusively applied to finance or re-finance in part or in full new

1.3.3 ESG Rating Agencies and Indices Providers

Crucial in the process of implementing SRI strategies is the role of ESG analysis performed by specialised rating agencies. By processing publicly available information reported by companies, NGOs and governmental organisations, these agencies provides investors with regularly updated data, starting point for comparing companies' ESG performances.

As reported by Novethic in their *Overview of ESG Rating Agencies* (2014), since the early 2000s, there has been a significant development of the ESG rating market, and it has recently undergone a phase of consolidation. The market is, in fact, becoming more and more concentrated around the largest agencies which, through partnerships or acquisitions of smaller and local ones, are now able to provide a wide range of services[3] at the international level.

International Rating Agencies

What follows is a brief list of the most prominent international ESG rating agencies:

- EIRIS Ltd., a subsidiary of the EIRIS foundation, was established in 1983 in the United Kingdom and, besides its London headquarters, it has offices in Paris, Boston and Washington. Their methodology consists in evaluating companies based on a set of 200 indicators covering environment, stakeholder, human rights, and governance issues. For each indicator, companies are assigned a grade expressing their level of commitment (limited, intermediate, good, advanced).

- Inrate is a Switzerland based agency and one of the oldest in continental Europe. Over the years it has established partnerships with universities, NGOs and other rating agencies. Its activity is mainly focused on small capitalization companies and emerging markets.

- MSCI ESG Research, subsidiary of the MSCI group (Morgan Stanley Capital International), is one of the world's largest provider of environment, social and governance research. It currently is the result of a series of acquisition of different ESG analysis agencies[4]. Each company in their analysed universe (more than 5000) is provided with a grade and a sectoral review, highlighting the best and worst ESG practices. MSCI has also devel-

and/or existing eligible Green Projects and which are aligned with the four core components' of the Green Bond Principles (GBP)."

[3]Novethic lists 11 different types of services: ESG analysis of companies, analysis of countries, analysis of supranational organisations and local governments, norm-based analysis, controversy alerts, engagement services, governance and proxy voting,evaluation of portfolios, requested rating, analysis of "green bonds", issuance of SRI indices.

[4]In 2010, MSCI acquired RiskMetrics which had previously bought ISS (Institutional Shareholder service), Innovest Strategic Value Advisors and KLD Research & Analytics.

oped a wide family of ESG indices, covering different geographical areas and a multitude of ESG thematics.

- Sustainalytics, a Netherlands based company founded in 2002, provides information on more than 4500 companies worldwide assessing their ESG performance using sector specific indicators. In 2009 it merged with the Canadian Jantzi Research Inc. and collaborates with STOXX Ltd. for its ESG and Sustainability index family.

- Vigeo, founded in 2002, is one of the European leading experts in the assessment of companies' practices and performances on ESG issues. In 2010 it created the Vigeo rating brand, through which it provides companies' ratings based on 38 ESG issues divided in six fields: environment, human rights, human resources, community involvement, business behaviour and corporate governance.

- Asset4 is a Swiss non-financial data provider founded in 2003 and acquired, in 2009, by Thompson Reuters. Being the primary data source for this thesis, a more detailed review of their methodology will be given in Chapter 4.

SRI Indices

Most of the aforementioned rating agencies also provide the service of creating and maintaining SRI indices. While the main purpose is to serve as benchmarks or reference for SRI funds and portfolios, they also play a crucial role in influencing ESG behaviour; companies are motivated to improve their effort in the ESG field in order to be included in these renowned indices.

The most famous is the Domini 400 Social Index (DS 400) by KLD Research and Analytics. Created in 1990, it was the first equity index compiled, starting from S&P 500 constituents, following ESG and ethical criteria. Since the acquisition, in 2010, of RiskMetrics by the MSCI Group, the index is known as the MSCI KLD 400 Social Index and is part of the MSCI's ESG index family.

Other important SRI index families are: Calvert Responsible Index series by Calvert Investments, FTSE4GOOD index series by the FTSE Group, Dow Jones Sustainability Index (DJSI) series and the STOXX ESG and Sustainability indices.

Among these, there are indices built focusing on specific themes (CO_2 emission, controversial weapons, etc.) and religious values (e.g. Catholic or Islamic indices).

1.3.4 Initiatives

The strong development of SRI and the rise in importance given to ESG related issues of the last decades has motivated important actors in the SRI field (associations[5], think tanks, NGOs, stock exchanges, national governments) to promote numerous initiatives. These are mainly aimed at encouraging CSR disclosure and at providing guidelines and best practices for companies and investors.

What follows is a list of the most relevant initiatives:

- The United Nations Environment Programme Finance Initiative (UNEP FI) is a partnership, started in 1992, between the United Nations and the global financial sector, now represented by more than 200 banks, insurer and investors from 51 countries. Its goal is to create an environment that helps financial institutions to integrate sustainable development policies into their operations.

- The UN Global Compact, described as a global multi-stakeholder network, aims at helping companies to meet their commitments to operate responsibly and support society. Companies are asked to embrace a set of core principles in the area of human rights, labour standards, environment and corruption.

- The UNEP FI and the Global Compact are both supporters of The United Nations Principles for Responsible Investment initiative (UN PRI), launched in 2006. It is based on 6 principles [6] and has the aim of promoting the incorporation of ESG issues into mainstream investment decision-making and ownership practices. It has almost 1500 signatories representing 60 trillion dollars of assets under management.

- The OECD Guidelines are recommendations addressed by governments to multinational enterprise (MNEs) operating in or from the adhering countries. They provide principles for responsible business consistent with applicable laws and internationally recognised standards.

- The Sustainable Stock Exchanges initiative (SSE), founded in 2009, is a global platform for exchanges to engage with policy makers, investors, companies and regulators with the aim of creating more sustainable and transparent capital markets. Exchanges, partners of

[5]Most of these associations are built on the model of the *social investment forum* (SIF).

[6]1.Incorporate ESG issues into investment analysis and decision making;2.Incorporate ESG issues into ownership policies and practices;3.Seek ESG disclosure;4.Promote the PRI principles within the financial industry;5.Work cooperatively to implement the PRI principles;6.Report on progress in implementing the PRI principles.

the SSE, are committed to promote improved ESG disclosure and performance among its listed companies [7].

- The Global Impact Investing Network is a nonprofit organization aimed at increasing the scale and effectiveness of impact investing. Its activity consists in providing critical infrastructure and in supporting activities, education, and research, in order to accelerate the development of a coherent impact investing industry. Since 2009 it promotes the IRIS initiative, a catalog of generally accepted performance metrics used by investors, investment funds, companies, and foundations to track their social, environmental, and financial achievements.

- In November 2004, Eurosif, together with its national members, created the European SRI Transparency Guidelines (Now known as Transparency code) aimed at increasing the accountability and transparency of SRI products to users. The code has become a widely used tool and it has been made mandatory for French and Belgium SRI funds. At the end of 2013 there were more than 500 signatories fund, representing an important share of the SRI market.

1.4 SRI Strategies

Following Louche and Lydenberg (2010), four main types of SRI strategies can be identified: avoidance, inclusion, relative selection, and engagement.

- Avoidance strategy consists in avoiding companies engaged in businesses or practices deemed unacceptable or harmful to society.

- Inclusion strategy, as opposed to Avoidance, is implemented by investing in companies involved in businesses or practices considered to be particularly beneficial to society.

- Relative selection is aimed at selecting the best performing companies in terms of ESG criteria in each sector or industry.

- Engagement implies active collaboration with companies in order to influence their behaviour and improve their ESG performance.

These strategies are not mutually exclusives; investors can implement them independently or combined.

[7] For a comprehensive list of partners see SSE's website at `http://www.sseinitiative.org/sse-partner-exchanges/list-of-partner-exchanges/` (accessed October 3, 2016).

1.4.1 SRI Strategies in Europe

Consistent with the aforementioned classification, Eurosif, in their 2016 study (Eurosif, 2016), distinguished between 7 categories of SRI strategies:[8]:

Exclusion or Negative Screening is a strategy which involves removing from the investment universe specific investments or classes of investments, such as companies, sectors or countries. There is not a single criterion for negative screening and motivations may vary from risk-management to value-based. The most common practice is the exclusion of companies which are involved in so-called "sin" activities. For example, for their ESG and Sustainability Index family, STOXX Ltd. excludes companies which are involved with alcohol, tobacco, gambling, weapons and pornography. As stated by Louche et al. (2012) the underlying principle is "do no harm".

Investors who are more concerned with environmental issues may decide to exclude activities such as production and distribution of agrochemicals, animal testing, nuclear energy production and genetically modified organisms (GMO) research. Faith driven investors, instead, may exclude companies producing abortifacients and contraceptives or involved with embryonic stem cell research.

Norm-based Screening is the assessment, for each company held in the portfolio, of its compliance with specific environmental, social and governance standards. These standards are most commonly based on internationals norms or treaties such as The United Nation Global Compact and the OECD Guidelines for Multinational Enterprises.

This strategy is most popular in the Nordic countries where, coupled with engagement activities, covers the majority of the SRI assets.

Engagement and voting by shareholders on ESG matters. It is a long-term process whose goal is to influence companies behaviour and increase disclosure. It involves dialogue and collaboration between management and investors (Eurosif, 2013).

ESG Integration is the explicit inclusion, by asset managers, of ESG risks and opportunities into traditional financial analysis and investment decisions based on a systematic process and apropriate research sources.

Best-in-Class typically involves selecting the top percentage of companies, within a sector or an investment universe (Best-in-Universe), based on ESG criteria. It is also referred to as positive screening.

[8]The strategies are listed by popularity as of 2016.

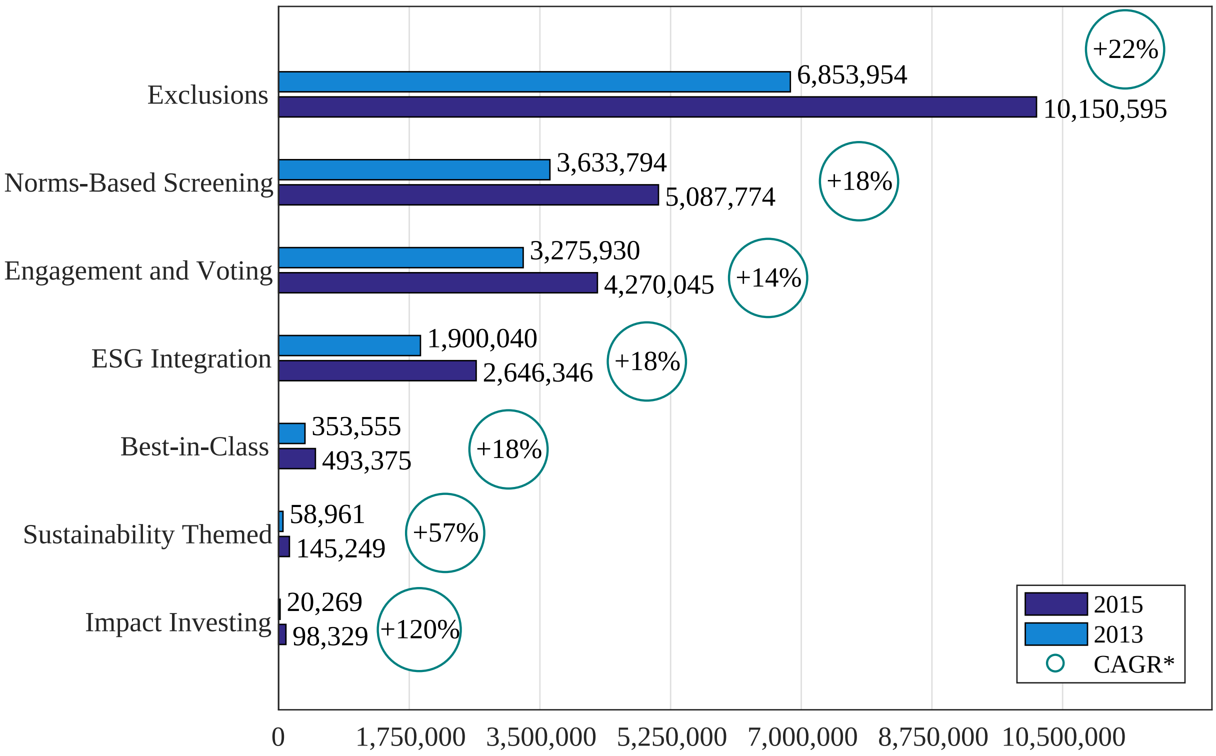


Figure 1.2: SRI assets per strategy (EUR in millions). Data source: Eurosif (2016). *Compounded Annual Growth rate.

Sustainability Themed is investing following specific themes related to ESG issues (climate change, energy efficiency, etc.). This type of strategy is usually aimed at supporting companies and industries in the process of transitioning to more sustainable practices. Themed investing may also provide some degree of de-correlation to other investments present in investors' portfolios.

Impact Investing is, according the Global Impact Investing Network (GIIN), the investment made into companies, organisation and funds with the intention to generate social and environmental impact alongside a financial return. Key aspects of Impact Investing are the intentionality, the measurability of the social impact and the expectation of financial returns.

As shown in figure 1.2, in late 2015, exclusions based strategies were the most adopted, covering around 10 trillion euro of assets under management. Norm-based screening and Engagement and Voting followed with 5 and 4.3 trillion euro, respectively. The least applied strategy, although experiencing an exceptional rate of growth (385% in the 2013-2015 biennium), was impact investing with 98 billion euro.

Chapter 2

Literature Review

This chapter provides a review of existing studies aimed at comparing the performance of Socially Responsible Investments versus conventional forms of investments, where conventional means that no SRI strategy is applied to the investment process.

2.1 Hypotheses

According to Hamilton, Jo, and Statman (1993) and, more recently, Statman and Glushkov (2009), there are three alternative hypotheses addressing the relative returns of stocks of socially responsible companies and conventional companies.

The reasoning behind these hypotheses could be explained in terms of competitive advantages or disadvantages, at company level, resulting from acting socially responsible (Porter & van der Linde, 1995; Walley & Whitehead, 1994). However, as Derwall et al. (2005) suggested, the measure in which ESG policies contribute to investment returns ultimately depends on the financial markets' ability to factor the financial consequences of corporate social responsibility into share prices.

2.1.1 Doing Good but not Well

The first hypothesis is the "doing good but not well" (Statman & Glushkov, 2009), where the risk-adjusted expected returns of socially responsible stocks are lower than the risk-adjusted expected returns of conventional stocks. This hypothesis is consistent with the idea that the market prices social responsibility characteristics and that socially responsible investors have an impact on stock returns. This type of investors are willing to sacrifice financial returns, to some extent, in exchange of social performance. By driving down the expected returns and, consequently, the cost of capital, they increase the value of socially responsible companies relative to the value of conventional companies. As a result, socially responsible companies are overvalued with respect to conventional ones.

2.1.2 Doing Good While Doing Well

The "doing good while doing well" hypothesis states that the expected returns of socially responsible stocks are higher than those of conventional stocks. This is possible if investors, and managers, consistently underestimate the benefits of acting according to socially responsible standards and, at the same time, overestimate its costs. Investors may also underestimate the probability that negative information may be released about non socially responsible companies. In this regard, Hamilton et al. (1993) made the example that conventional investors may be consistently underestimating the probability that oil companies will find themselves in trouble because of oil spills. If oil spills actually occurs, the companies stocks' prices will decline resulting in lower returns for conventional investors while not affecting the returns of socially responsible investors who refrained from investing in those companies.

2.1.3 No Effect

Last is the "no effect" hypothesis where there are no significant differences between the expected returns of socially responsible and conventional stocks. This may occur if the social responsibility components of the stocks are not priced by the market. Unlike the case of "doing good but not well", investors that sell stocks of non socially responsible companies find enough conventional investors ready to buy them that the prices do not drop. The "no effect" hypothesis might also be true if the effects of social responsibility consistent with the "doing good but not well" hypothesis are counterbalanced by those consistent with the "doing good while doing well".

2.1.4 Missed Diversification Opportunities

In addition to these three hypotheses, which mainly focus on the performance of companies' stocks, considerations regarding optimal portfolio construction and performance needs to be made. As Becchetti, Ciciretti, Dalò, and Herzel (2015) well explained, when socially responsible investors exclude assets from their investment universe, because of negative screening or Best-in-Class strategies, they introduce additional constraints in their optimal portfolio variance minimization problem, requiring the share invested in those assets to be zero. This entails that their efficient frontier shifts to the right and appears flatter than that of conventional investors.

For the same level of expected return, the risk (measured by the standard deviation) is higher, meaning that SRI strategies can be costly in terms of missed diversification opportunities.

The following section will be dedicated at describing the methodology used in the existing literature to test and verify the aforementioned hypotheses.

2.2 Methodologies

The most common approach used in the literature to test whether the stocks' performance of SR companies differ significantly from those of conventional ones is by means of Jensen's alpha (Jensen, 1968).

In the Capital Asset Pricing Model (CAPM) (Lintner, 1965; Sharpe, 1964), a single-factor model, the Jensen's alpha is obtained by regressing the stock's excess returns (returns minus the risk-free rate) on the excess returns of the market portfolio.[1] The constant in the regression, estimated through OLS, is the Jensen's alpha, which represents the part of stock's return which is not explained by its level of systematic risk (measured by its beta).

Multi-factor models, which take into account the exposure to additional risk factors, are also widely used. The most renowned is the Fama-French three factor model (Fama & French, 1993, 1996). Fama and French added two additional explanatory variables in the regression, the Small Minus Big factor (SMB) and the High Minus Low factor (HML). The SMB factor, computed as the difference in returns between a small capitalization and a large capitalization portfolio, measures the exposure to small size risk, while the HML factor, calculated as the difference in returns between a portfolio of companies with high book-to-market ratios and a portfolio of companies with low book-to-market ratios, captures the exposure to bankruptcy risk.

Carhart (1997) extended the three factor model by adding a momentum factor (MOM). Momentum is constructed as the difference in returns between a portfolio of stocks which exhibited relatively high returns in the past year and a portfolio of stocks which exhibited relatively low returns in the same time frame. It measures the tendency of stocks that performed well (badly) in the last year to continue to perform well (badly) in the future.

Spanning test (De Roon & Nijman, 2001; Huberman & Kandel, 1987) is often used in studies aimed at comparing the efficient frontier built on a restricted investment universe, in which assets are excluded, for example, based on their ESG performance, and the one corresponding to an unrestricted investment universe. It is a regression based test used to check if the difference between the efficient frontiers is statistically significant or only due to sampling error.

De Roon, Nijman, and Werker (2001) extended the spanning test in order to account for

[1] A broad market index is often used as a proxy for the market portfolio.

market frictions such as short-sale restrictions and transaction costs, while Zhou and Kan (2012) proposed a step-down procedure to test if the exclusion of some assets decreases diversification opportunities in terms of foregone returns (1st step) or foregone risk reduction (2nd step).

2.3 Results

In one of the first studies of its kind, Hamilton et al. (1993) measured the excess returns for 32 socially responsible equity mutual funds using monthly returns for the period 1981-1990. By using a single-factor model, they found that, for the majority of the SRI funds, the Jensen's alphas were not significantly different from zero. Based on those results they concluded that the market did not price social responsibility characteristics and that investors could have expected to lose nothing by investing in SRI mutual funds.

Diltz (1995), using a sample of 159 US firms and daily returns for the period encompassing the years 1989 to 1991, estimated the difference in Jensen's alpha for fourteen pairs of portfolios.[2] The author found that certain ethical screens, in particular good environmental performance, lack of military work, and lack of nuclear industry involvement, may enhance portfolio performance. Through cumulative average abnormal return (CAAR) analyses, Diltz also found that enhanced portfolio performance may come from environmental and charitable giving screens, while negative effects may be obtained from family benefits screen. Overall, the author interpreted his findings to be consistent with the notion that ethical screening neither helps nor hinders portfolio performance.

Stone, Guerard, Gultekin, and Adams (2001) studied how social screening impacted active portfolio management during period from 1984 to 1997. Using stocks data from three databases[3] and rating from KLD Research & Analytics, they found no statistically significant costs nor benefits to such practice.

Geczy, Stambaugh, and Levin (2003) studied whether investors pay a "price" by allocating their wealth to socially responsible equity mutual funds. The results strongly depended on what fraction of their portfolios investors decided to restrict to SRI funds and on their prior beliefs about pricing models and managerial skill. Investors who strongly believed in the CAPM and ruled out managerial skill (such as a market index investor) paid a small price for their SRI restrictions. Same result for investors whose allocation to SRI funds was small. Instead, the costs associated with socially responsible investing became more significant as investors beliefs shifted towards Fama-French or Carhart multi-factor models and when they relied on fund

[2]The first portfolio in each pair consisted of all sample firms which received the highest rating for a particular rating screen, while the second consisted of all sample firms which received the lowest rating.

[3]CRSP, COMPUSTAT, and I/B/E/S.

manager skill.

Bauer, Koedijk, and Otten (2005) used Carhart's four-factor model to test the significance of the difference in risk-adjusted returns between ethical and conventional mutual funds for the 1990-2001 period.[4] Although they found no evidence of statistically significant difference in returns between the two type of mutual fund, they documented that ethical mutual funds exhibited distinct investment styles.[5] They also investigated the relative performance of ethical mutual funds through time, finding that, after a period of under-performance in the beginning of the 1990's, they experienced a catching-up phase which brought their performance on par with that of conventional funds for the 1998-2001 period.

Derwall et al. (2005) focused on the concept of eco-efficiency[6]. They created two equity portfolios that differed in eco-efficiency ratings. After controlling for common risk factors, using a four-factor model and industry specific factors, they found that the high-rated portfolio significantly outperformed its low-rated counterpart over the 1995-2003 period. Their results remained significant also after accounting for different levels of transaction costs.

Barnett and Salomon (2006) hypothesized, based on modern portfolio theory and stakeholder theories, that the financial loss encountered by SRI funds due to poor diversification may be offset as the degree of social screening increases because better managed and more stable firms are selected. They found evidence for their hypothesis by means of an empirical test on a panel of 61 SRI funds from 1972 to 2000. Their results showed that, as the number of social screens used by a SRI fund increased, financial returns initially declined, but then, when the number of screens reached a maximum, these started to rise. Consistent with their hypothesis, they found a curvilinear relationship between financial and social performance.

Kempf and Osthoff (2007) investigated the impact of different socially responsible criteria (positive screening, best-in-class, negative screening) on the performance of stock portfolios. Using ratings from KLD Research & Analytics and stocks from the S&P 500 and DS 400 indices they built two portfolios, one with high SRI rated companies and the other with low rated companies.[7] Using the Carhart four-factor model, they measured portfolios performances during the period 1992-2004 finding that a long-short strategy, long on high rated companies and short on low rated ones, yielded a significantly positive alpha.[8] Checking for reasonable levels of transaction costs did not change their results. Another interesting result was that the two port-

[4]They used an international database containing 103 German, UK and US ethical mutual funds.

[5]They found that ethical funds are typically less exposed to market return variability and more growth-oriented (lower beta for the HML factor) compared to conventional funds.

[6]Eco-efficiency is here intended as the economic value a company creates relative to the waste it generates.

[7]Both a value-weighted and an equally-weighted (as a robustness check) weighting scheme were used for constructing the portfolios.

[8]The Best-in-Class screening approach was the one that lead to the highest alphas up to 8.7% per year.

folios differed in the exposure level to the HML factor, with the high-rated portfolio showing a lower factor-loading.

Schröder (2007) aimed its study at comparing the risk-return characteristics of SRI equity indices[9] with those of conventional indices (or benchmark indices). The analysis was conducted by regressing the excess returns of the SRI indices on the excess returns of the benchmarks, both in a single and a multi-equation framework. The results neither revealed a significant out-performance nor an under-performance of the SRI indices compared to the benchmarks (alphas in the regressions not significantly different from zero). Instead, most of the SRI indices exhibited higher risk compared to conventional counterparts (beta significantly bigger than 1).

Renneboog, ter Horst, and Zhang (2007) focused on the SRI mutual fund industry to test their hypothesis that ethical and social considerations have an effect on stock prices and that investors pay a price (the *price of ethics*, as they called it) for the use of SRI screening. Their sample was made of SRI and conventional equity funds from 23 different countries[10] and covered the period 1991-2003. By means of multi-factor analysis they found that the risk-adjusted returns of the average SRI fund in the UK and US were not statistically different from those of an average conventional fund. Instead, the average SRI fund in the majority of European and Asia-Pacific countries strongly underperformed its conventional counterpart. They also found evidence of what they called a “smart money” effect in the SRI fund industry; ethical investors, while unable to identify the SRI funds that would have outperformed their benchmarks in subsequent periods, were able to identify ethical funds that would have performed poorly. Lastly, in accordance with Barnett and Salomon (2006) study, the data showed that the performance of SRI funds increased with the number of SRI screens employed, supporting the hypothesis that screening processes generates value-relevant non-public information.

Statman and Glushkov (2009), in a study that strongly resembles that of Kempf and Osthoff (2007), both in methodology and data, analysed the effects of Best-in-Class and ethical screening strategies on portfolio excess returns. What partially differentiated their study from Kempf and Osthoff (2007), apart from the inclusion of three additional years of data, is how they treated the ratings provided by KLD Research & Analytics[11]. Despite the differences, their results were consistent with those obtained by the other authors; stock portfolios of high-rated companies out-performed portfolios of low-rated ones while the exclusion of stocks of companies involved

[9] 29 SRI equity indices representing different SRI screening procedures and covering different investment regions (global, Europe, and specific countries such as Australia, Sweden, the UK and the US).

[10] More specifically: Austria, Belgium, France, Germany, Ireland, Italy, Luxembourg, Netherlands, Norway, Sweden, Switzerland, UK (including Guernsey and the Isle of Man), USA, Australia, Canada, Cayman Islands, Japan, Malaysia, the Netherlands Antilles, Singapore, and South Africa.

[11] They excluded from their analysis the group of companies that had no indicators of strength and no indicators of concern during the year because that group likely included companies that KLD had not examined even if they were on its list while Kempf and Osthoff (2007) included them.

in “sin” activities determined a significant loss in performance. They concluded that the effects of Best-in-Class and ethical screening strategies tend to offset each other.

Galema et al. (2009), in a study related to that of Schröder (2007), used spanning tests to determine whether investors sustain a loss, in diversification terms, by excluding non socially responsible assets from their portfolio. In particular, by applying tests from De Roon et al. (2001) and Zhou and Kan (2012), they tried to answer three main questions:

1. Does restricting the universe to socially responsible stocks decreases diversification opportunities of investors in terms of foregone returns?

2. Does restricting the universe to socially responsible stocks decreases diversification opportunities of investors in terms of foregone risk reduction opportunities?

3. Does restricting the universe to socially responsible stocks decreases diversification opportunities of investors when they are subject to short sales constraints?

They used stock returns for more than 2000 North American companies for the period 1991-2004 and distinguished between five dimensions of social responsibility: environment, social, corporate governance, product and sin. For each dimension they built a SRI portfolio, made of only stock deemed socially responsible by KLD, and one of excluded assets.[12] Tests results evidenced that investors, when not imposing short-selling restrictions, were worse off for every dimension of social responsibility except for governance. However, the step-down procedure showed that these investors paid a price only in terms of foregone risk reduction opportunities and not in terms of foregone returns. For investors facing short-selling restriction the null hypothesis of spanning held for all 5 dimensions, meaning that these type of investors, such as many individuals, mutual funds and pension funds, were not worse off in terms of diversification opportunities.

Hong and Kacperczyk (2009) investigated the effects of social norms on equity markets by focusing only on “sin” stocks.[13] They hypothesized that investors, in particular institutional ones, subject to norms, pay a financial costs in abstaining from investing in these stocks. They indeed found that “sin” stocks are less held by norm-constrained investors, such as pension plans, compared to mutual or hedge funds and that they receive less analysts coverage than comparable stocks. Using data from 1965 to 2006 for NYSE, AMEX and NASDAQ stocks, they observed that a portfolio long in “sin” stocks and short on comparable non “sin” stocks had significantly positive excess returns after controlling for Fama-French’s three factors, Carhart’s momentum and firm characteristics. Their results support the hypothesis of “Doing good but not well”; the

[12] They built their portfolios by using both a value-weighted and an equally-weighted scheme.
[13] They considered companies involved in producing alcohol, tobacco and gaming.

exclusion of these companies from investors portfolios increases their cost of capital resulting in stock prices which are under-valued respect to their fundamentals.

Herzel et al. (2012), similarly to Galema et al. (2009), examined the effects of including socially responsible related constraints in optimal portfolio decision-making. Their investment universe was based on the components of the S&P 500 index from 1993 to 2008 and they focused on the three main dimensions of social responsibility: environment, social and governance. The SRI constraints consisted of eliminating from the investments set, for each one of the three dimensions, the worst performing companies.[14]. They found that SRI screening resulted in a small loss in terms of Sharpe ratio while it had a great impact on the market capitalization of the optimal portfolio. They performed spanning tests both with and without imposing short-sales restrictions and the results showed that, when short-selling was allowed, the null hypothesis of spanning was almost always rejected, meaning that the two frontiers, the one available to a SRI investor and the one available to a conventional investor, were significantly different. Opposite results in the case of short-sales restrictions, whereby the hypothesis of spanning was rejected only when excluding from the investment universe more than 30% of the worst companies based on environment ratings.

More recently, Becchetti et al. (2015) studied the performance of socially responsible and conventional funds in different markets during the 1992-2012 period. They adopted single and multi-factor models as well as a ”nearest neighbour” approach[15] to estimate the difference between the alphas for the two investment styles (socially responsible and conventional). They found that there was not a clear dominance of one investment style over the other for the entire period, meaning that the diversification cost, due to excluding certain assets, does not compromise the performance of SRI funds. Analysing different time segments they found that, during the period following the global financial crisis, SRI funds outperformed conventional ones playing a sort of “insurance role” against ethical risk factors.

2.4 Concluding Remarks

What most of these studies seems to agree with is that, exception made for some specific cases (time periods, geographical regions, screening methodologies), Socially Responsible Investments, whether in the form of funds, indices or portfolios, do not tend to exhibit significant performance differences with respect to their conventional counterparts. Instead, they tend to

[14]They repeated their analysis using four different thresholds excluding the worst 10%, 30%, 50% and an all-concern screening.

[15]The approach consisted in matching SRI and conventional funds that are as close as possible in terms of exposition to risk factors.

show a significantly different exposure to common risk factors (Fama-French) and different investment styles.

Furthermore, examining the effects of the exclusion of non-SRI assets from the investment universe, two main results needs to be highlighted:

1. Losses in terms of diversification opportunities are mainly due to foregone risk reduction opportunities and not in terms of foregone returns.

2. These losses tend to disappear when investors are faced with short-sales restrictions.

Overall, according to the studies here reviewed, the “no effect” hypothesis seems to be the most plausible one and, from a socially responsible investor point of view, this represents a positive result.

Chapter 3

Methodology

This chapter provides a description of the methodology employed in the empirical research.

3.1 Mean-Variance Frontier

First introduced by Markowitz, the concept of mean-variance frontier is a crucial component of Modern Portfolio Theory (MPT) and plays a fundamental role in portfolio selection problems. MPT is built upon the assumption that investors are risk-averse; between two portfolios, offering the same level of return, they will always choose the less risky alternative.

In this context, the efficient frontier represents the set of portfolios satisfying the condition that, for the same level of risk, no other portfolio exists offering a higher level of expected return.

3.1.1 Optimal Portfolios

Considering a set of N assets, let μ be the vector of expected returns and Σ the covariance matrix. A specific optimal portfolio is the solution to the following problem:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ \text{s.t.} \quad & w'\mu = \mu_{p,0} \\ & \sum_{i=1}^{N} w_i = 1, \end{aligned} \tag{3.1}$$

where $w'\Sigma w$ is the portfolio variance or σ_p^2, $w'\mu$ is the portfolio expected return or μ_p, w is the vector of asset weights (proportion of wealth invested in the available assets) and $\mu_{p,0}$ is the target return. The first constraint imposes that the portfolio return should match the target return, while the second requires that the available wealth is fully invested.

The solution to the optimization problem is given by

$$w^* = \frac{A\Sigma^{-1}\iota_N - B\Sigma^{-1}\mu}{\Delta} + \frac{C\Sigma^{-1}\mu - B\Sigma^{-1}\iota_N}{\Delta}\mu_{p,0}, \tag{3.2}$$

where

$$A = \mu'\Sigma^{-1}\mu \quad B = \iota_N'\Sigma^{1}\mu \quad C = \iota_N'\Sigma^{1}\iota_N$$

and

$$\Delta = AC - B^2,$$

with ι_N being a N-dimension vector of ones.

From 3.2 is possible to derive a relation between μ_p and σ_p^2 of the optimal portfolios:

$$\sigma_p^2 = \frac{C}{\Delta}\mu_p^2 - \frac{2B}{\Delta}\mu_p + \frac{A}{\Delta}.$$

This relation, which links a return level with the variance of the optimal portfolio, is the mean-variance frontier and is represented, in the (σ_p^2, μ_p) plane, as a hyperbola. Since standard deviation, or volatility, σ_p is expressed in the same units as the returns μ_p, is common practice to represent the frontier in the (σ_p, μ_p) plane instead.

The efficient frontier is the portion of the hyperbola that lies above its vertex. The portfolio located on the vertex is called the Global Minimum Variance (GMV) portfolio and it is the solution to the following problem:

$$\begin{aligned} &\min_{w} \quad w'\Sigma w \\ &\text{s.t.} \quad \sum_{i=1}^{N} w_i = 1. \end{aligned} \tag{3.3}$$

The weights of the GMV portfolio are given by

$$w_{GMV} = \frac{\Sigma^{-1}\iota_N}{\iota_N'\Sigma^{-1}\iota_N}.$$

The points lying below the vertex, while still solutions to the problem in 3.1, do not represent efficient portfolios; for the same level of risk, there exist a portfolio, on the efficient frontier, offering a higher level of return.

3.1.2 Efficient Frontier with a Risk-Free Asset

Besides investing only in risky assets, investors may also allocate part of their wealth on a risk-free asset, characterized by null standard deviation.

When the investment universe also includes a risk-free asset, the optimization problem expressed in 3.1 becomes

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ \text{s.t.} \quad & w'\mu + (1 - w'\iota_N)R_f = \mu_{p,0}, \end{aligned} \tag{3.4}$$

where w now represents the portion of wealth invested only in risky assets, $1 - w'\iota_N$ the portion of wealth invested in the risk-free asset and R_f the risk-free rate.

The set of efficient portfolios can be derived as before, and it is now represented, in the (σ_p, μ_p) plane, as a straight line originating from the $(0, R_f)$ point. This is the so-called Capital Market Line (CML) and its slope, which coincides with the Sharpe ratio of the efficient portfolios, is given by

$$\frac{\mu_p - R_f}{\sigma_p} = Sh_p.$$

There exists a value of $\mu_{p,0}$ for which the problem in 3.4 yields a solution, w, such that

$$w'\iota_N = 1.$$

In this case, the optimal portfolio is made of only risky assets and, therefore, it is also a solution to the problem in 3.1. Referred to as the Tangency, or Maximum Sharpe (MS) portfolio, it is located on the tangency point between the CML and the efficient frontier of risky assets only.

Exploiting the Sharpe ratio definition, the MS portfolio composition can be found by solving the following problem:

$$\begin{aligned} \max_{w} \quad & \frac{w'\mu - R_f}{w'\Sigma w} \\ \text{s.t.} \quad & \sum_{i=1}^{N} w_i = 1. \end{aligned}$$

The solution to which is given by

$$w_{MS} = \frac{\Sigma^{-1}(\mu - \iota_N R_f)}{\iota_N'\Sigma^{-1}(\mu - \iota_N R_f)}.$$

and its Sharpe ratio is equal to

$$Sh_{MS} = ((\mu - \iota_N R_f)'\Sigma^{-1}(\mu - \iota_N R_f))^{1/2}. \tag{3.5}$$

Each portfolio located on the Capital Market Line can be obtained as a combination of the risk-free asset and the Tangency portfolio.

3.1.3 Positivity Constraints

Solving the variance minimizing problems in 3.1 and 3.4 often leads to portfolios characterized by extreme positions (extremely high negative and/or positive asset weights) which are infeasible to real world investors. The most common limitation investors are faced with, whether institutional or retail, is the no short-selling restriction. This imposes not to assume negative positions (go short) on the available assets.

To accommodate for this restriction, the optimization problem in 3.1 is subject to an additional constraint:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ \text{s.t.} \quad & w'\mu = \mu_{p,0} \\ & \sum_{i=1}^{N} w_i = 1 \\ & w_i \geq 0 \qquad i = 1, ..., N. \end{aligned} \tag{3.6}$$

Because of the inequality constraint, the problem in 3.6 does not have an analytical solution as for the unconstrained case; it must be solved by using numerical methods (like the `quadprog` or `fmincon` functions in MATLAB).

The efficient frontier in the case of short-sales restrictions is lower bounded by the GMV portfolio, computed as in 3.3 with the addition of the positivity constraint, and upper bounded by a portfolio characterized by a 100% position in the asset with the highest expected return. Compared to the unrestricted one, because of limited diversification opportunities, the constrained efficient frontier is shifted to the right and down on the (σ_p, μ_p) plane.

The positivity constraint also affects the slope of the Capital Market Line; the Sharpe ratio of the MS portfolio with short-sales restriction will always be lower than or equal to the Sharpe ratio of its unrestricted counterpart.

3.2 Parameters Estimation

The composition of the optimal portfolios, solution to the variance minimizing problems seen in the previous section, depends on μ, the assets expected returns, and Σ, the covariance matrix. Since the true values are unknown, both μ and Σ need to be estimated from observed returns.

This section will be dedicated at describing two common estimation approaches.[1]

3.2.1 Sample Moments

Given a sample of a multivariate series of returns $\boldsymbol{R}_t, t = 1, \ldots, T$, estimates for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be computed as the sample moments

$$\widehat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{R}_t. \tag{3.7}$$

and

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{R}_t - \widehat{\boldsymbol{\mu}})(\boldsymbol{R}_t - \widehat{\boldsymbol{\mu}})' = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R}_t\boldsymbol{R}_t' - \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}'.$$

More specifically, the diagonal elements of $\widehat{\boldsymbol{\Sigma}}$ are the individual assets variance estimators $\widehat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}(R_{i,t} - \widehat{\mu}_i)^2$, and the off-diagonal elements are the covariances estimators between asset pairs $\widehat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}(R_{i,t} - \widehat{\mu}_i)(R_{j,t} - \widehat{\mu}_j)$.

Employing the sample covariance matrix may be problematic when the matrix dimension (which coincides with the number of assets N) is large compared to the number of observations (T). In these scenarios, the sample covariance matrix, although being an unbiased estimator, is estimated with a lot of error and may perform poorly (Fan, Fan, & Lv, 2008; Ledoit & Wolf, 2004a, 2004b). When N is greater than T the estimated covariance matrix becomes singular and, therefore, cannot be used as parameter in portfolio selection problems.

To overcome these issues is common practice to impose some structure on the estimator of $\boldsymbol{\Sigma}$. This can be achieved through the specification of a factor model.

3.2.2 Factor Models[2]

Factor models are used to explain assets returns by exploiting their exposure to a certain number of common variables called factors (or risk factors). Depending on the type of factors employed, it is possible to distinguish between macroeconomic factor models and fundamental factor models. The first category relies on observable macroeconomic variables such as GDP growth rate or inflation rate, while the second on observable asset characteristics like the book-to-market ratio.

[1] Just for this section, bold characters will be used in the notation to help differentiate vectors and matrices from scalar values.

[2] This section follows the approach used in *Statistics and Data Analysis for Financial Engineering* by Ruppert (2011).

A factor model is expressed as follows:

$$R_{i,t} = \alpha_i + \beta_{1,i}F_{1,t} + \beta_{2,i}F_{2,t} + \cdots + \beta_{K,i}F_{K,t} + \epsilon_{i,t} \quad i = 1, 2, \ldots, N, \tag{3.8}$$

where $R_{i,t}$ is now the excess return of the i-th asset at time t, $F_{1,t}, F_{2,t}, \ldots, F_{K,t}$ are the values of the K factors at time t, $\epsilon_{i,t}$ is a mean-zero variable that represents the individual risk of the ith asset and α_i is the part of the return not explained by the model (abnormal return). In factor models is common to assume that the ϵ terms are uncorrelated across assets, meaning that the cross-correlation between returns is only due to the factors.

The $\beta_{k,i}$ parameter is the factor loading and expresses the exposure, or sensitivity, of the i-th asset returns to the k risk factor. The factor loadings are unknown and must be estimated by means of times-series regression.

Expected Returns and Covariance Matrix Estimation

Using the factor model in 3.8, the expected return for asset i can be written as

$$E(R_{i,t}) = \alpha_i + \beta_{1,i}E(F_{1,t}) + \beta_{2,i}E(F_{2,t}) + \cdots + \beta_{K,i}E(F_{K,t}). \tag{3.9}$$

Let

$$\boldsymbol{\beta}_i' = (\beta_{1,i}, \beta_{2,i}, \ldots, \beta_{K,i})$$

be the vector of factor loadings for the i-th asset,

$$\boldsymbol{F}_t' = (F_{1,t}, F_{2,t}, \ldots, F_{K,t})$$

the vector of K factors at time t and suppose that $\boldsymbol{\Sigma_F}$ is the $K \times K$ covariance matrix of $\boldsymbol{F}_t$. The variance for the i-th asset returns is

$$Var(R_{i,t}) = \boldsymbol{\beta}_i'\boldsymbol{\Sigma_F}\boldsymbol{\beta}_i + \sigma_{\epsilon,i}^2.$$

And, for any $i \neq j$, the covariance between two assets is

$$Cov(R_{i,t}, R_{j,t}) = \boldsymbol{\beta}_i'\boldsymbol{\Sigma_F}\boldsymbol{\beta}_j.$$

More generally, let

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{1,1} & \dots & \beta_{1,j} & \dots & \beta_{1,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_{K,1} & \dots & \beta_{K,j} & \dots & \beta_{K,N} \end{pmatrix}$$

be the matrix of loadings and

$$\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \dots, \alpha_N)$$

the vector of intercepts.
Equation 3.8 can be rewritten, using matrix notation, as

$$\boldsymbol{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}' \boldsymbol{F}_t + \boldsymbol{\epsilon_t},$$

where

$$\boldsymbol{\epsilon_t}' = (\epsilon_1, \epsilon_2, \dots, \epsilon_N).$$

The vector of expected returns is

$$E(\boldsymbol{R}_t) = \boldsymbol{\alpha} + \boldsymbol{\beta}' E(\boldsymbol{F}_t)$$

and the $N \times N$ covariance matrix of $\boldsymbol{R}_t$ is given by

$$\boldsymbol{\Sigma_R} = \boldsymbol{\beta}' \boldsymbol{\Sigma_F} \boldsymbol{\beta} + \boldsymbol{\Sigma_\epsilon},$$

where $\boldsymbol{\Sigma_\epsilon}$ is the $N \times N$ diagonal covariance matrix of $\boldsymbol{\epsilon_t}$

$$\boldsymbol{\Sigma_\epsilon} = \begin{pmatrix} \sigma^2_{\epsilon,1} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2_{\epsilon,j} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \sigma^2_{\epsilon,N} \end{pmatrix}.$$

The $\boldsymbol{\Sigma_\epsilon}$ matrix is diagonal because of the assumption that specific risk factors are not correlated across assets.

Estimates for $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ are obtained from the time-series regression of the realized excess returns on the observable risk factors. The covariance matrix of the factor realizations may be estimated using the sample covariance matrix

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{F}} = \frac{1}{T-1}\sum_{t=1}^{T}(\boldsymbol{F}_t - \overline{\boldsymbol{F}})(\boldsymbol{F}_t - \overline{\boldsymbol{F}})', \quad \overline{\boldsymbol{F}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{F}_t.$$

Compared to the sample covariance matrix described in the previous section, when N is large and K is relatively small, the factor model approach requires estimating fewer parameters thus determining an increase in overall precision. More specifically, whereas the sample covariance matrix contains $N(N+1)/2$ estimates, the factor model requires estimating $N \times K$ parameters in $\boldsymbol{\beta}$, K^2 in $\boldsymbol{\Sigma}_{\boldsymbol{F}}$ and N in $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, for a total of $NK + N + K^2$. As an example, assuming the number of assets is 500 and the number of factors is 4, $N(N+1)/2 = 125,250$ while $NK + N + K^2 = 2,516$.

A downside of the factor model is that, in case of model misspecification, the estimate for $\boldsymbol{\Sigma}_{\boldsymbol{R}}$ will be biased. This is especially the case when $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is not diagonal as per assumption.

Equilibrium Returns

The CAPM and the Black-Litterman model for asset allocation (Black & Litterman, 1991, 1992) are based on the assumption that financial markets, in the long run, tend to be in a state of equilibrium. Markets are in equilibrium when the supply of financial activities matches the demand.

If assets returns are believed to be generated by a factor model such as the one in 3.8, the equilibrium assumption implies that there should not be abnormal returns, that is, the intercept in the regression must be equal to zero.

Under the equilibrium assumption, equation 3.9 for computing expected returns becomes

$$E(R_{i,t}) = \beta_{1,i}E(F_{1,t}) + \beta_{2,i}E(F_{2,t}) + \cdots + \beta_{K,i}E(F_{K,t}) \; .$$

Equilibrium returns are determined only by the assets exposure to the risk factors.

3.3 Testing Mean-Variance Efficiency

An important question, in the subject of portfolio analysis, is whether expanding the investment universe, by including an additional set of risky assets, can lead to a significant improvement in terms of mean-variance opportunities. Numerous tests have been proposed trying to address this type of question; most notably the mean-variance efficiency test proposed by Jobson (1982) and Gibbons et al. (1989) that will be here described following the approach used by Pastorello

(2001).

3.3.1 Testing Efficiency in the Presence of a Risk-Free Asset

As described in section 3.1.2, when investors are faced with the possibility of allocating part of their wealth on a risk-free asset, the set of efficient portfolios is represented, in the (σ, μ) plane, by the Capital Market Line. The slope of the CML is given by the Sharpe ratio of the MS portfolio attainable by investing in risky assets only.

Let now consider two sets of risky assets: one of K benchmark assets and an augmented one that, besides the benchmark assets, includes N additional test assets. If the CML relative to the first set does not significantly differ from the one relative to the larger set, then, expanding the investment universe (with the test assets) does not determine an improvement in terms of mean-variance opportunities.

Exploiting the properties of the CML, the problem it's equivalent to verify the following set of hypothesis:

$$\begin{cases} H_0 : Sh - Sh_0 = 0 \\ H_1 : Sh - Sh_0 > 0 \end{cases}, \tag{3.10}$$

where Sh_0 is the maximum Sharpe ratio achievable with the benchmark assets only and Sh the one attainable with the $K + N$ assets.

Using the subscripts 0 and 1 to refer to the K benchmark assets and N test assets, respectively, the vector of expected excess returns and the covariance matrix of the $N + K$ assets can be defined as

$$\mu = E(R_t) = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \quad \Sigma = Var(R_t) = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix}.$$

Following Gibbons et al. (1989), the hypotheses in 3.10 can be tested using the regression

$$R_{1,t} = \alpha + \beta R_{0,t} + \epsilon_t, \tag{3.11}$$

with $E(\epsilon_t) = 0_N$ and $E(\epsilon_t R'_{0,t}) = 0_{N\times K}$, where 0_N is a N-vector of zeros and $0_{N\times K}$ a $N \times k$ matrix of zeros.

The set of hypotheses in 3.10 can now be reformulated in terms of the intercepts in the regression as

$$\begin{cases} H_0 : \alpha = 0 \\ H_1 : \alpha \neq 0 \end{cases}.$$

The null hypothesis implies that the test assets expected returns can be obtained as a linear combination of the expected returns of the benchmark assets. Assuming that the excess returns are i.i.d., the parameters of the multivariate model in 3.11 can be estimated using OLS for each individual equation.

Gibbons et al. (1989) showed that the null hypothesis H_0 can be verified via the following test statistic:

$$\xi_{GRS1} = \frac{T-K-N}{N}\left(1+\widehat{\mu}_0'\widehat{\Sigma}_{00}^{-1}\widehat{\mu}_0\right)^{-1}\widehat{\alpha}'\widehat{\Sigma}_{\epsilon}^{-1}\widehat{\alpha}, \tag{3.12}$$

where $\widehat{\Sigma}_\epsilon$ is the consistent estimator of Σ_ϵ given by

$$\widehat{\Sigma}_\epsilon = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_t'.$$

Under H_0 and the additional assumption that the excess returns are jointly normally distributed, the test statistic in 3.12 has exact distribution $F_{N,T-K-N}$.

If excess returns are i.i.d. but not normal, H_0 can be verified using an asymptotic test statistic given by

$$\xi_{GRS2} = T\left(1+\mu_0'\Sigma_{00}^{-1}\mu_0\right)^{-1}\widehat{\alpha}'\Sigma_{\epsilon}^{-1}\widehat{\alpha}, \tag{3.13}$$

which, under H_0, is asymptotically distributed as χ^2_N. In order to compute ξ_{GRS2} the unknown quantities μ_0, Σ_{00} and Σ_ϵ can be replaced by their estimators $\overline{R}_0$, $\widehat{\Sigma}_{00}$ and $\widehat{\Sigma}_\epsilon$.

Gibbons et al. (1989) also showed that the test statistics in 3.12 and 3.13 can be interpreted in terms of Sharpe ratios:

$$\xi_{GRS1} = \frac{T-K-N}{N}\frac{\widehat{Sh}^2-\widehat{Sh}_0^2}{1+\widehat{Sh}_0^2} \tag{3.14}$$

and

$$\xi_{GRS2} = T\frac{\widehat{Sh}^2-\widehat{Sh}_0^2}{1+\widehat{Sh}_0^2}. \tag{3.15}$$

Using equation 3.5, the difference of the squared Sharpe ratios can be written as

$$\widehat{Sh}^2-\widehat{Sh}_0^2 = \widehat{\mu}'\widehat{\Sigma}^{-1}\widehat{\mu}-\widehat{\mu}_0'\widehat{\Sigma}_{00}^{-1}\widehat{\mu}_0$$

Exploiting the inverse partitioned formula for $\widehat{\Sigma}^{-1}$, the last equation is equal to

$$\widehat{Sh}^2-\widehat{Sh}_0^2 = \widehat{\mu}_{0|1}'\widehat{\Sigma}_{11|0}^{-1}\widehat{\mu}_{0|1}$$

where

$$\begin{aligned}\widehat{\mu}_{0|1} &= \widehat{\mu}_1-\widehat{\Sigma}_{10}\widehat{\Sigma}_{00}^{-1}\widehat{\mu}_0\\ \widehat{\Sigma}_{11|0} &= \widehat{\Sigma}_{11}-\widehat{\Sigma}_{10}\widehat{\Sigma}_{00}^{-1}\widehat{\Sigma}_{01}.\end{aligned} \tag{3.16}$$

Since the values defined in 3.16 are also equal to the OLS estimates of α and Σ_ϵ, respectively, it is verified that

$$\widehat{Sh}^2 - \widehat{Sh}_0^2 = \widehat{\alpha}'\widehat{\Sigma}_\epsilon^{-1}\widehat{\alpha}.$$

It is interesting to highlight that, for a known value of the risk-free rate, the intersection test procedures by Huberman and Kandel (1987) and De Roon and Nijman (2001) are equivalent to the efficiency tests here described.

Short-Selling Restrictions

The efficiency test so far reviewed relies on the crucial assumption consisting in the absence of short-selling restrictions. To overcome this limitation, De Roon et al. (2001) implemented a variation of the intersection (and spanning) test that can be applied also in the presence of two types of market frictions: short-selling restrictions and transaction costs. In particular, their procedure involves testing for inequality constraints on the parameters of a regression equivalent to the one defined in 3.11[3].

Alternatively, Glen and Jorion (1993) proposed an approach that exploits the already mentioned possibility of formulating the efficiency hypothesis test statistics in terms of Sharpe ratios. When short selling is not allowed, the distributions of the statistics in 3.14 and 3.15 are unknown but can be approximated through simulation. This methodology can be described as follows:

1. Expected excess returns and covariance matrix of the test and benchmark assets are computed from observed data.

2. Excess returns of the N test assets are modified in order to satisfy the null hypothesis implying the efficiency of the subset of K assets. This is achieved by first computing the MS portfolio of the K assets imposing positivity constraints. Then, the returns of the N test assets are regressed against the returns of the *MS* portfolio just computed:

$$R_{i,t} = \alpha_i + \beta_{MS,i}R_{MS,t} + \epsilon_i \quad i = 1, \ldots, N.$$

According to Glen and Jorion (1993), the null hypothesis implies that $\alpha_i = 0 \; \forall \; i = 1, \ldots, N$.[4] Therefore, null-restricted excess returns for the test assets can be computed as follows:

$$R^0_{i,t} = \widehat{\beta}_{MS,i}R_{MS,t} + \widehat{\epsilon}_i \quad i = 1, \ldots, N.$$

[3] With the difference that returns, and not excess returns, are used.

[4] De Roon et al. (2001) proved that, in case of short-selling restrictions, the null hypothesis is satisfied even with $\alpha_i \leq 0 \; \forall \; i = 1, \ldots, N$.

This ensures that the MS portfolio (with positivity constraints) of the $N + K$ assets will coincides with the one of the K benchmark assets, satisfying the null hypothesis.

3. A random sample of returns, of size $T \times (N + K)$, is drawn from a multivariate normal distribution using as parameters those computed on the null-restricted data. Expected returns and covariance matrix are then estimated from the simulated sample.

4. MS portfolios (with positivity constraint) are computed for the K assets and for the $N+K$ assets, using the parameters estimated in the previous step, and the test statistic in 3.15 is recorded[5].

5. Steps 3 and 4 are repeated a sufficiently large number of times.

6. Denoting by ξ_{GRS2} the original test statistic (computed on observed data), the p-value can be computed as

$$PV = \frac{\{\xi^{*,m}_{GRS2} \geq \xi_{GRS2}\} + 1}{M + 1},$$

where $\xi^{*,m}_{GRS2}$ is the test statistic computed from the m-th simulated sample and M is the number of performed simulations.

Alternatively to what described in step 3, random samples can be generated via block bootstrap. This involves resampling, with replacement, blocks of returns from the null-restricted data obtained in step 2. In particular, either the stationary bootstrap or the circular block bootstrap from Politis and Romano (1992, 1994) can be used. These methodologies have the advantage of not relying on any assumption regarding the underlying distribution and also allows preserving autocorrelation in stocks' returns.

The MATLAB code, implementing the two mean-variance efficiency tests (with and without short-selling), is included in Appendix B.

3.4 Performance Testing with the Sharpe Ratio

The Sharpe ratio is widely used by analysts and investors to compare the performances of two alternative investment strategies (such as stocks, portfolios, hedge funds, mutual funds etc.). Even though it may not be the most adequate measure when dealing with non normally distributed returns (since it relies only on the first two moments of the returns' distribution), Eling

[5] In this case, since their distribution is approximated through simulation, using test statistic 3.14 or 3.15 is equivalent.

and Schuhmacher (2007) showed that the Sharpe ratio can result in almost identical performance ranking when compared to alternative measures[6].

The Sharpe ratio computed from historical return data does represents an estimation and, therefore, is affected by estimation error. Consequently, the comparison between two Sharpe ratios, relative to two different investment strategies, must be based on statistical inference.

Jobson and Korkie (1981) proposed a test statistic that, under the assumption that returns are i.i.d. and normally distributed, allows verifying the hypothesis that the difference between two Sharpe ratios is equal to zero.

More recently, Ledoit and Wolf (2008) developed two inference methods that are valid in more general conditions: returns' distribution with heavier tail than the normal and autocorrelation. Starting from the difference of the estimated Sharpe ratios, the first test involves computing a HAC (Heteroskedasticity and autocorrelation consistent) standard error, while the second, which exhibits improved finite sample performance, involves building a studentized bootstrap confidence interval.

This section will be dedicated at briefly describing these two methods using a notation as faithful as possible to the original paper.

3.4.1 Framework

Suppose there are two investment strategies, i and n, whose excess returns[7] at time t are r_{ti} and r_{tn}, respectively. T return pairs $(r_{1i}, r_{1n})', \ldots, (r_{Ti}, r_{Tn})'$ are observed.

The returns' distribution has mean vector and covariance matrix given by

$$\mu = \begin{pmatrix} \mu_i \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_i^2 & \sigma_{in} \\ \sigma_{in} & \sigma_n^2 \end{pmatrix}.$$

Sample means and sample variances are denoted by $\widehat{\mu_i}$, $\widehat{\mu_n}$ and $\widehat{\sigma_i}$, $\widehat{\sigma_i}$.

The difference between the Sharpe ratios is given by

$$\Delta = Sh_i - Sh_n = \frac{\mu_i}{\sigma_i} - \frac{\mu_n}{\sigma_n}$$

[6]In particular, they tested 13 different performance measures: Sharpe's, Treynor's, and Jensen's measures, Omega and Sortino ratio, Kappa 3, the upside potential ratio, the Calmar ratio, the Sterling ratio, the Burke ratio, the excess return on value at risk, the conditional Sharpe ratio, and the modified Sharpe ratio.

[7]To be consistent with what done so far, excess returns are computed over the risk-free rate. Alternatively they can be computed over a given benchmark, such as a stock index, but in that case, as pointed ou in Ledoit and Wolf (2008), the performance measure used is the Information ratio (rather than the Sharpe ratio).

and its estimator is

$$\widehat{\Delta} = \widehat{Sh}_i - \widehat{Sh}_n = \frac{\widehat{\mu}_i}{\widehat{\sigma}_i} - \frac{\widehat{\mu}_n}{\widehat{\sigma}_n}.$$

In implementing their methodology, the authors preferred dealing with uncentered second moments, that is $\gamma_i = E(r_{1i}^2)$ and $\gamma_n = E(r_{1n}^2)$, with their sample counterparts $\widehat{\gamma}_i$ and $\widehat{\gamma}_n$. Defining $v = (\mu_i, \mu_n, \gamma_i, \gamma_n)'$ and $\widehat{v} = (\widehat{\mu}_i, \widehat{\mu}_n, \widehat{\gamma}_i, \widehat{\gamma}_n)'$, Δ and $\widehat{\Delta}$ can be rewritten as

$$\Delta = f(v), \quad \widehat{\Delta} = f(\widehat{v}),$$

where

$$f(a, b, c, d) = \frac{a}{\sqrt{c - a^2}} - \frac{b}{\sqrt{c - b^2}}. \tag{3.17}$$

Assuming that

$$\sqrt{T}(\widehat{v} - v) \xrightarrow{d} \mathcal{N}(0, \Psi),$$

where Ψ is an unknown symmetric positive semi-definite matrix, the delta method implies that

$$\sqrt{T}(\widehat{\Delta} - \Delta) \xrightarrow{d} \mathcal{N}(0, \nabla' f(v) \Psi \nabla f(v)), \tag{3.18}$$

where

$$\nabla' f(a, b, c, d) = \left(\frac{c}{(c - a^2)^{1.5}}, -\frac{d}{(d - b^2)^{1.5}}, -\frac{1}{2}\frac{a}{(c - a^2)^{1.5}}, \frac{1}{2}\frac{b}{(d - b^2)^{1.5}} \right).$$

Given a consistent estimator $\widehat{\Psi}$ for Ψ, the standard error for $\widehat{\Delta}$ can be computed as

$$s(\widehat{\Delta}) = \sqrt{\frac{\nabla' f(\widehat{v}) \widehat{\Psi} \nabla f(\widehat{v})}{T}}. \tag{3.19}$$

3.4.2 HAC Inference

A consistent estimator $\widehat{\Psi}$ can be obtained by using HAC kernel estimation. Once chosen the kernel function $k(\cdot)$ and a bandwith S_T, the kernel estimate for Ψ is given by

$$\widehat{\Psi} = \frac{T}{T-4} \sum_{j=-T+1}^{T-1} k(\frac{j}{S_T}) \widehat{\Gamma}_T(j),$$

where

$$\widehat{\Gamma}_T(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} \widehat{y}_t \widehat{y}_{t-j}' & \text{for } j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} \widehat{y}_{t+j} \widehat{y}_t' & \text{for } j < 0 \end{cases}, \quad \widehat{y}_t' = (r_{ti} - \widehat{\mu}_i, r_{tn} - \widehat{\mu}_n, r_{ti}^2 - \widehat{\gamma}_i, r_{tn}^2 - \widehat{\gamma}_n).$$

In their application, Ledoit and Wolf (2008) used both a Parzen and a Quadratic-Spectral (QS) kernel function[8], while the bandwidth was computed using an automatic method described by Andrews (1991).

Given the kernel estimator $\widehat{\Psi}$, the standard error $s(\widehat{\Delta})$ is given by equation 3.19. The asymptotic normality implied in 3.18 allows computing a two-sided p-value for the null hypothesis $H_0 : \Delta = 0$

$$PV = 2\Phi\left(-\frac{|\widehat{\Delta}|}{s(\widehat{\Delta})}\right),$$

where $\Phi(\cdot)$ is the cumulative density function (c.d.f.) of the normal distribution.

Alternatively, a $1-\alpha$ confidence interval for Δ is given by

$$\widehat{\Delta} \pm z_{1-\frac{\alpha}{2}} s(\widehat{\Delta}), \tag{3.20}$$

where z_λ is the λ quantile of the standard normal distribution.

3.4.3 Bootstrap Inference

The second approach involves constructing a two-sided bootstrap confidence interval for Δ with a nominal level of $1-\alpha$. If the interval does not contain 0, the null hypothesis $H_0 : \Delta = 0$ is rejected with significance level equal to α.

More precisely, the authors proposed constructing a symmetric studentized bootstrap confidence interval by approximating the distribution function of the studentized statistic as follows:

$$\mathcal{L}\left(\frac{|\widehat{\Delta}-\Delta|}{s(\widehat{\Delta})}\right) \approx \mathcal{L}\left(\frac{|\widehat{\Delta}^*-\widehat{\Delta}|}{s(\widehat{\Delta}^*)}\right),$$

where Δ is the unknown true difference between the Sharpe ratios, $\widehat{\Delta}$ is the estimated difference computed from observed data, $s(\widehat{\Delta})$ is the standard error for Δ, $\widehat{\Delta}^*$ is is the estimated difference computed from bootstrap data, and $s(\widehat{\Delta}^*)$ is the standard error for $\widehat{\Delta}^*$. $\mathcal{L}(X)$ denotes the distribution of the random variable X.

The bootstrap $1-\alpha$ confidence interval for Δ is computed as

$$\widehat{\Delta} \pm z^*_{|\cdot|,1-\alpha} s(\widehat{\Delta}), \tag{3.21}$$

where $z^*_{|\cdot|,1-\alpha}$ is the $1-\alpha$ quantile of $\mathcal{L}\left(|\widehat{\Delta}^*-\widehat{\Delta}|/s(\widehat{\Delta}^*)\right)$.

To generate bootstrap data, Ledoit and Wolf (2008) employed the circular block bootstrap from Politis and Romano (1992). This involves resampling, with replacement, blocks of return

[8] In their simulation study the two kernel functions resulted in virtually identical rejection probabilities.

pairs from the observed $(r_{1i}, r_{1n})', t = 1, \ldots, T$, where the block has a fixed size $b \geq 1$.[9]

The standard error $s(\widehat{\Delta})$ is computed as in equation 3.19, estimating Ψ via kernel estimation, while the bootstrap standard error for $\widehat{\Delta}^*$ is given by

$$s(\widehat{\Delta}^*) = \sqrt{\frac{\nabla' f(\widehat{v}^*)\widehat{\Psi}^*\nabla f(\widehat{v}^*)}{T}},$$

where $\widehat{v}^* = (\widehat{\mu}_i^*, \widehat{\mu}_n^*, \widehat{\gamma}_i^*, \widehat{\gamma}_n^*)$ is the estimator of v from bootstrap data.

Additionally, defining $l = \lfloor T/b \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, the equation used by Ledoit and Wolf (2008) to compute $\widehat{\Psi}^*$ is as follows

$$\widehat{\Psi}^* = \frac{1}{l}\sum_{j=1}^{l} \zeta_j \zeta_j',$$

with

$$\zeta_j = \frac{1}{\sqrt{b}}\sum_{t=1}^{b} y^*_{(j-1)b+t} \quad t = 1, \ldots, l.$$

Crucial in the described bootstrap methodology is the choice of an appropriate block size b. For this purpose, the authors chose to use a calibration method which, given a calibration function $g : b \to 1-\lambda$ and a desired confidence level $1-\alpha$, involves finding the value of b that minimizes $|g(b) - (1-\alpha)|$.[10]

Finally, a two-sided test for the null hypothesis $H_0 : \Delta = 0$ at significance level α can be performed by constructing a confidence interval with confidence level $1-\alpha$ for Δ as in equation 3.21; if zero is not included in the interval then the null hypothesis is rejected. Alternatively, a p-value can be obtained as

$$PV = \frac{\left\{\tilde{d}^{*,m} \geq d\right\} + 1}{M+1}, \tag{3.22}$$

where d is the “original” studentized test statistic

$$d = \frac{|\widehat{\Delta}|}{s(\widehat{\Delta})},$$

and $\tilde{d}^{*,m}$ is the centred studentized statistic computed from the m-th bootstrap sample

$$\tilde{d}^{*,m} = \frac{|\widehat{\Delta}^{*,m} - \widehat{\Delta}|}{s(\widehat{\Delta}^{*,m})} \quad m = 1, \ldots, M.$$

Ledoit and Wolf (2008) suggest setting M, the number of bootstrap resamples, equal to 4999.

[9] Ledoit and Wolf (2008) proposed two different bootstrap methodologies depending whether the observed return data is i.i.d. or of time series nature. The authors recommend always using the methodology for time series data, therefore the one for i.i.d. data will not be reviewed.

[10] For a detailed description of how to compute the optimal block size see algorithm 3.1 in the original paper.

3.5 Performance Testing with the Variance

When following an investment strategy aimed at minimizing returns variance, like investing in the GMV portfolio, it is more meaningful to compare alternative strategies' performance by analysing their variances, rather than their Sharpe ratios. Ledoit and Wolf (2011) showed that the robust inference methods described in the previous section can also be used, after appropriate modifications, to test for the difference in the variance of two investment strategies.

Since the methodology strongly resembles the one already reviewed, only the relevant differences will be here reported.

3.5.1 Framework

Using the same notation as before, the hypotheses to test are now

$$H_0 : \Delta = 0 \text{ vs. } H1 : \Delta \neq 0,$$

with

$$\Delta = log(\sigma_i^2) - log(\sigma_n^2)$$

and its sample counterpart

$$\widehat{\Delta} = log(\widehat{\sigma}_i^2) - log(\widehat{\sigma}_n^2).$$

The authors justify the use of the log-transformation of the variances stating that it conduces to better finite-sample properties of the inference methods.

The function defined in 3.17 needs now to be modified in

$$f(a, b, c, d) = log(c - a^2) - log(d - b^2)$$

and its gradient becomes

$$\nabla' f(a, b, c, d) = \left(-\frac{2a}{c - a^2}, \frac{2b}{d - b^2}, \frac{1}{c - a^2}, -\frac{1}{d - b^2} \right).$$

From this point forward, both the HAC and the bootstrap inference methods are exactly the same as those used for testing the difference in Sharpe ratios.

Chapter 4

Empirical Research

This chapter reviews all the procedures used to investigate the impact of SRI strategies on efficiency and portfolio performance. Starting by illustrating the process of data selection and management, it then proceeds describing the application of the methodology seen in the previous chapter. Results are then displayed and discussed with the aid of plots and tables.

4.1 Data

4.1.1 Reference Index

The first step of this research consisted in choosing the market region to analyse. In order to differentiate this thesis from existing studies (in particular Galema et al., 2009 and Herzel et al., 2012), which mainly focused on the North American market, the effects of investing according to socially responsible criteria were tested using data relative to European region.

The investment universe, from which select the companies to include in the portfolios, was chosen to coincide with a broad European equity index, the STOXX Europe 600. This is a capitalization-weighted index with a fixed number of 600 components including large, mid and small capitalization companies from 17 countries of the European region[1]. The index is reviewed on a quarterly basis and its constituents are updated at the end of March, June, September and December.

Constituents lists of the index, from December 2001 to March 2016, were obtained from Datastream and merged into a single list[2]. Multiple entries were then removed by controlling

[1]Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom.

[2]Datastream provides historical constituent lists for each month of the index life period. Since the STOXX Europe 600 is reviewed quarterly, only the constituents lists relative to March, June, September and December of each year were downloaded. The lists relative to these months are those already updated and valid for the three subsequent months.

both ISIN codes and Datastream identifiers. The result of this process was a "master" list consisting of the 1151 different companies which, at some point in time, since 2002, have been included in the STOXX Europe 600.

For each one of the 1151 companies the following data was downloaded from Datastream (datatypes in parentheses): Industry (ICBIN)[3], country of domiciliation (WC06026), daily Total Return Index (RI)[4] and Market Capitalization (MV) time series expressed in euro. Monthly Returns were then computed on the Total Return Index value of the last trading day of each month.

It is important here to clarify that the investment universe was not fixed and did not coincide with all the 1151 companies. Instead, its composition varied, matching, each trimester, that of the STOXX Europe 600.

4.1.2 ESG Ratings

In order to implement the SRI strategies, the next step consisted in obtaining ESG ratings for the companies included in the investment universe.

Through Datastream, Thompson Reuters provides access to the Asset4 ESG database. Asset4 is one of the most comprehensive ESG databases, providing, since 2003, yearly data for more than 4000 companies worldwide.

The data is structured in 4 pillars representing the four areas of company performance: economic, environment, social and corporate governance. More than 250 key performance indicators, computed from 750 data points, covering every aspect of sustainability reporting, are combined to provide an overall score for each one of the 4 pillars (see figure 4.1). For each company, Asset4 computes the scores, ranging from 0% to 100%, by equally weighting and z-scoring[5] all the underlying data points and comparing them against all other companies in their database. The resulting score represents, therefore, a relative measurement of performance.[6] The social and environment performance pillars also includes data points indicating whether a company generates revenue from controversial business activities and products[7].

[3]Using the Industry Classification Benchmark (ICB) taxonomy.

[4]The Total Return Index, opposed to price value, takes also into account dividends, assuming that these are always reinvested.

[5]Starting from a raw value, a z-score, or standard score, is computed by subtracting the population mean value and dividing the result by the population standard deviation. This allows to more easily distinguish between values that, otherwise, would be very close to each other

[6]The scores calculation methodology is described in an Excel file retrievable at the following link: `https://uvalibraryfeb.files.wordpress.com/2013/09/asset4_esg_data_glossary_april2013.xlsx` (accessed on March 22, 2016).

[7]These include: alcohol, gambling, tobacco, armaments, pornography, contraceptives, abortifacients, embryonic stem cell research, cluster bombs, anti-personal landmines, agrochemicals, animal testing, nuclear energy production, genetically modified organisms.

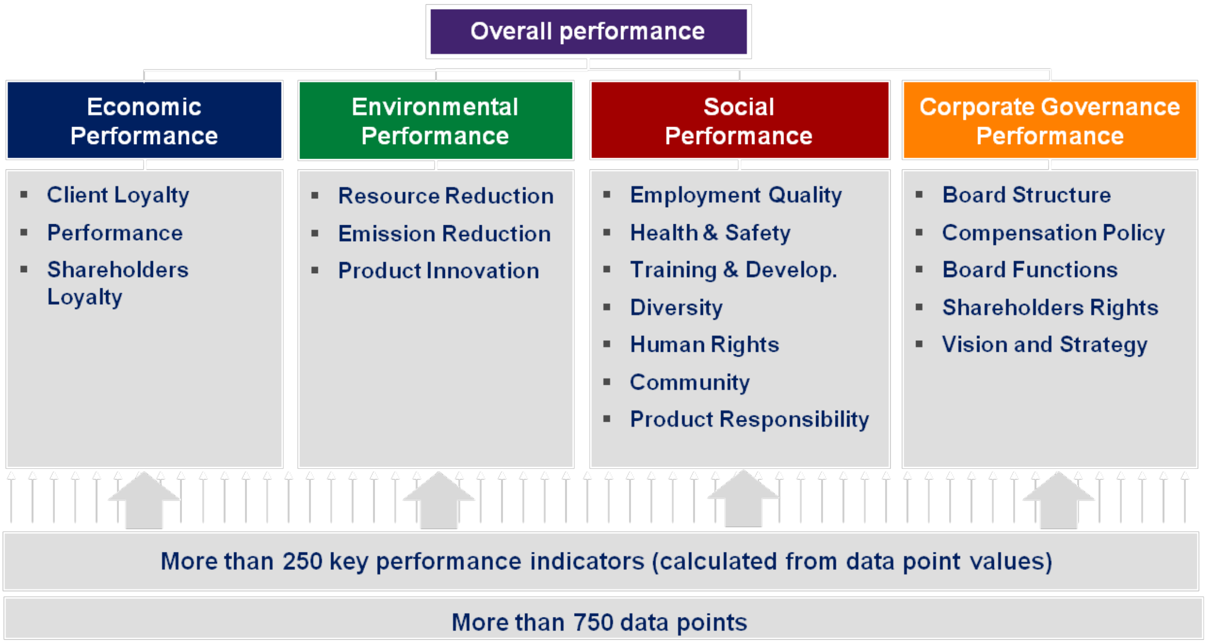


Figure 4.1: Overview of the Asset4 data framework showing how the data points and key performance indicators are structured into the four pillars. Image source: Thompson Reuters.

This research involved downloading the scores representing environment (ENV), social (SOC) and corporate governance (GOV) performance, starting from fiscal year 2002 up to 2015, for each one of the 1151 companies (when available). By equally-weighting the three distinct scores, an average ESG score was then computed. The economic performance score, concerning aspects not relevant for an ESG analysis, was not considered. Additionally, data points indicating company involvement in "sin" activities were also downloaded[8]. Following the approach used by STOXX Ltd. for their ESG and Sustainability indices[9], the controversial activities considered as "sin" were: production of alcoholic beverages, gambling activities, production of tobacco, production of vehicles, planes, armaments, or any combat materials used by the military and production or distribution of pornography.

4.1.3 Rating Matching and Yearly Coverage Ratio

Obtained the ESG data, it was necessary to match, for each trimester, the companies included in the investment universe with their ratings, valid for that particular year. Since Asset4 does not seem to release the data for all the companies in their database at the same time[10], it was necessary to assume that the ratings, relative to a particular fiscal year, were released at the end

[8]These data points, merely consisting in a Yes or No, indicate whether part of a company's revenue come from a particular controversial activity.

[9]In particular the STOXX Europe Sustainability Index ex AGTAFA (alcohol, gambling, tobacco, armaments, firearms and adult entertainment), see `https://www.stoxx.com/index-details?symbol=SUYP`.

[10]As of May 2016 almost only half of the companies rated for fiscal year 2014 had ratings available (at least on Datastream) relative to 2015.

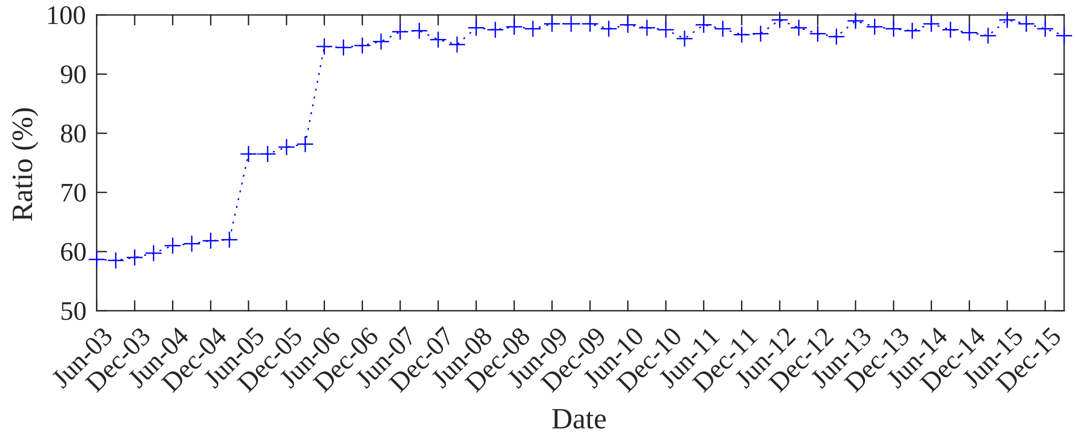


Figure 4.2: Percentage of companies included in the STOXX Europe 600 for which Asset4 provided ESG data.

of June of the subsequent year. Following this logic, the constituent lists of March of each year were matched with the ratings relative to two years prior, while those of June, September and December with those of one year prior. For example, an investor, willing to implement an SRI strategy at the end of March 2016, would have been bound to use ratings relative to the year 2014, given that no data would have been available for 2015. Asset4 provides ESG data starting from fiscal year 2002, therefore, the first constituent list with matching ratings was that of June 2003.

Completed the matching process, it was possible to compute, for each trimester (starting from June 2003 up to March 2016), a rating coverage ratio expressing the percentage of companies, constituents of the STOXX Europe 600, covered by Asset4. As shown in figure 4.2, the first two years were characterized by a coverage ratio of around 60% meaning that, out of the 600 companies included in the index, about 240 were not rated by Asset4. Starting from June 2005 the coverage ratio rose to almost 80% and since June 2006 it never fell below 94%.

In the paper from Herzel et al. (2012), the authors decided to exclude from the analysed investment universe all those companies with no available ESG data, applying the screening processes only on companies covered by KLD's ratings. In this analysis, instead, the exclusion of non rated assets was considered to be part of the screening process. This meant that, especially for the first three years, the effects of implementing an SRI strategy were also determined by the scarcity of ESG data. The intention was to replicate a more realistic scenario in which socially responsible investors are also faced with the difficulties of obtaining data for companies in their investment universe. This may have been true, in particular, in less recent years, when the SRI phenomenon was not as popular as today and the demand for ESG rating services was not as strong.

4.2 SRI Strategies Implementation

4.2.1 Preliminary Restriction on the Investment Universe

Before implementing any kind of SRI strategy, a necessary restriction was applied to the investment universe. In order to have, each trimester, sufficient historical data to perform the required estimations (expected returns and covariance matrix), companies with less than 5 years of past returns data were excluded. As a result, the actual investment universe, used in this empirical research, did not perfectly coincide with the STOXX Europe 600 index. Instead, it included a smaller and not fixed number of companies.

4.2.2 Screening Strategies

The typology of ESG ratings provided by Asset4 allowed to implement a combination of two SRI strategies: a negative, or ethical, screening strategy, consisting in excluding companies involved in "sin" activities, and a positive screening strategy, consisting in selecting only the best performing companies in environment, social and corporate governance aspects.

For the positive screening strategy, a Best-in-Universe approach was adopted. This consists in selecting the best performing companies over the entire investment universe, as opposed to a Best-in-Class approach in which the selection process is applied industry by industry. The Asset4 rating methodology makes this kind of approach feasible; as already seen, companies' ESG performances are evaluated in such a way that the resulting scores are also comparable across different industires.

Starting from June 2003, and for each trimester up to March 2016, the SRI strategies were implemented in the following order:

1. Exclusion of companies not rated by Asset4 (negative screening).

2. Exclusion of rated companies involved in "sin"activities (negative screening).

3. Selection of a predetermined percentage of best performing companies (among those not already excluded) based on the average ESG score[11] (positive screening).

As in Herzel et al. (2012), to study the impact of different degrees of social responsibility, three distinct percentages were used for the positive screening process: 90%, 70% and 50%.[12]

Additionally, in order to compare the impact on portfolio performance of the three different aspects of social responsibility, the environment, social and corporate governance scores were

[11] From this point forward the average ESG score will simply be referred to as ESG score.

[12] For example, with the 90% positive screening the worst 10% is excluded.

also separately taken into account. In this case, the SRI strategy consisted in applying only the positive screening process based on the four types of score: ESG, ENV, SOC, GOV.

Throughout most of this chapter, the analysis focuses on the combined negative and positive screening strategies based on ESG score, for which all results are displayed. The results for the remaining SRI strategies are reported in the last section.

4.2.3 Descriptive Analysis

The implementation of a SRI strategy inevitably affects the composition of the investment universe, both in qualitative and quantitative terms. By taking as reference the investment universe of a traditional investor (in this case, companies included in the STOXX Europe 600 and with at least five years of returns data), it is possible to investigate the effect of implementing a SRI strategy by examining three characteristics:

1. Actual number of companies included in the investment universe.

2. Industry composition.

3. Total market capitalization.

The plot in figure 4.3 shows, for each trimester, the number of companies remaining after applying the different levels of screening. Consistent with what already seen in figure 4.2 for the rating coverage ratio, up until June 2006, excluding companies with no ESG rating meant reducing the investment universe of a considerable number of components. In the first two years, almost 200 companies were lacking rating from Asset4 and almost 100 the following year. Instead, since June 2006 nearly all companies included in the reference investment universe were covered. Companies involved in "sin" activities represented a significant and slightly growing fraction of the rated ones, starting from 7% in 2003 up to around 11% in recent years. By looking at the plot, it is reasonable to expect that the effects of implementing a SRI strategy in terms of optimal portfolio performance, if any, would be particularly strong in the first three years, due to the large number of assets excluded because of scarcity of ESG data.

Implementing a Best-in-Universe strategy means that the screened investment universe may differ from the reference one in terms of industries composition. This can impact SRI portfolios performances because of the different degree of exposure to industry specific risks compared to a conventional portfolio. Figure 4.4 shows how the composition, in terms of industries, changed after applying the different levels of screening in the last trimester of the sample (March 2016). Both from the bar plot and the actual percentages reported in table 4.1 it appears that the composition did not drastically change, at least up to the 70% screening level.

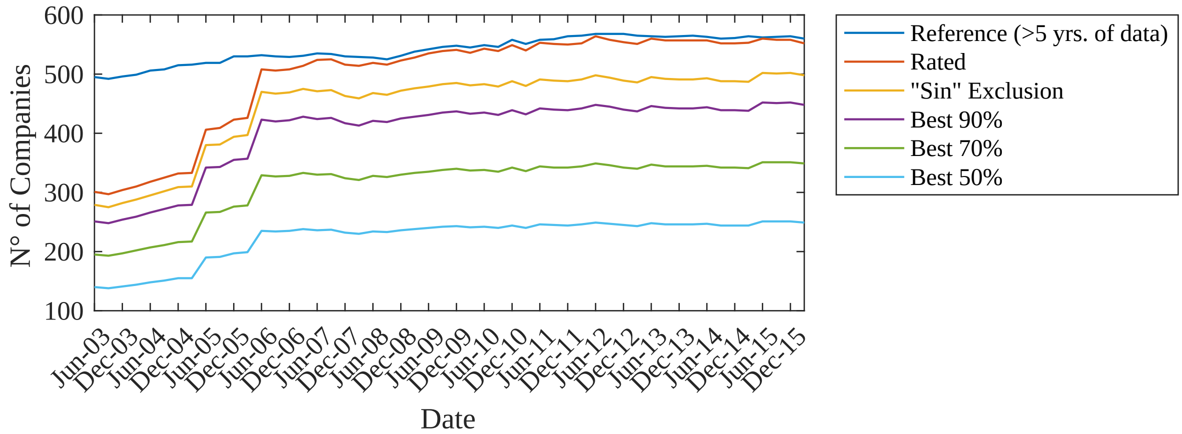


Figure 4.3: Number of companies for the different levels of screening.

It may seem odd, however, that the Oil & and Gas industry, commonly associated with practices and products that have a negative impact on the environment (e.g. fracking, CO_2 emissions, oil spills etc.), was progressively more represented as the level of SRI screening increased. Moreover, this was not an isolated case since both in June 2003 and December 2009 (figures and tables included in Appendix C), start and midpoint of the analysed time frame, the percentage of companies belonging to this sector increased after the screening processes.

This type of result can be explained by the following arguments:

1. The Oil & Gas industry, according to the ICB taxonomy, contains three sectors: 1.oil & gas producers, 2.oil equipment, services & distribution and 3.alternative energy. Companies belonging to the alternative energy sector, which are expected to have high ENV score, may be able to compensate for the possible bad performance of the other two and positively affect the overall industry score. In March 2016 the average ENV score of the 18 companies belonging to the oil & gas industry, excluding the alternative energy sector, was 74%, almost coinciding with that of the reference investment universe. Two companies belonged to the alternative energy sector (in particular to the renewable energy equipment subsector) with an average ENV score of 94%.

2. The ESG score is obtained by equally weighting the ENV, SOC and GOV ones. Oil & Gas industry companies may have simply performed above average on the Social and Corporate Governance aspects. This was the case in March 2016.

3. The Assets4 methodology, with regard to environment performance, consists in evaluating companies in terms of emission reduction, product innovation and resource reduction. Besides actual achievements in these fields, effort and commitment is also rewarded. This evaluation method may allow Oil & Gas companies to be assigned a relatively high ENV score.

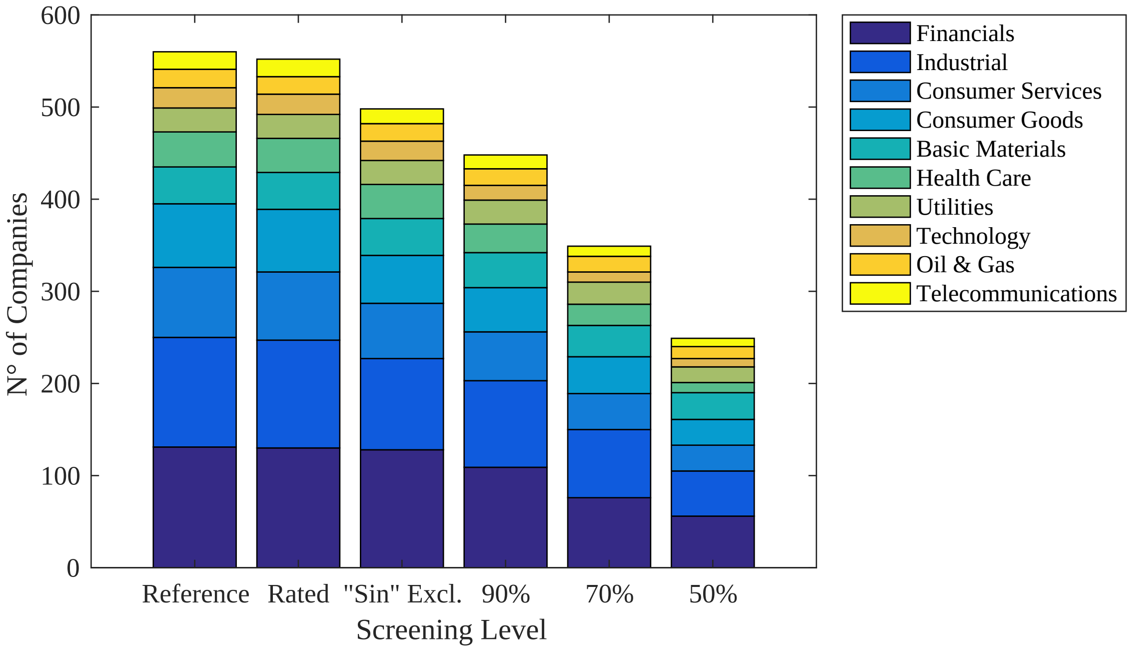


Figure 4.4: Number of companies per industry for the different levels of screening (as of March 2016).

Lastly, it was possible to investigate how the exclusion of non-SR companies translated in terms of total market capitalization loss. Figure 4.5 shows the loss, in percentage terms, with respect to the unrestricted investment universe. As expected, in the first three years, losses were strongly affected by the exclusion of a large number of non rated companies. In the subsequent years, even after eliminating more than half of the companies with the 50% screening (see figure 4.3), losses never exceeded the 40% level. This could mean that, on average, larger companies tend to perform better (in ESG terms) than smaller ones and that SR investors will prefer to invest in large and mid capitalization stocks.

Table 4.1: Industry composition for the different levels of screening (as of March 2016).

Industry Composition						
Industry	Reference	Rated	”Sin” Excl.	Best 90%	Best 70%	Best 50%
Financials	23.4%	23.6%	25.7%	24.3%	21.8%	22.5%
Industrial	21.3%	21.2%	19.9%	21.0%	21.2%	19.7%
Consumer Services	13.6%	13.4%	12.0%	11.8%	11.2%	11.2%
Consumer Goods	12.3%	12.3%	10.4%	10.7%	11.5%	11.2%
Basic Materials	7.1%	7.2%	8.0%	8.5%	9.7%	11.6%
Health Care	6.8%	6.7%	7.4%	6.9%	6.6%	4.4%
Utilities	4.6%	4.7%	5.2%	5.8%	6.9%	6.8%
Technology	3.9%	4.0%	4.2%	3.6%	3.2%	3.6%
Oil & Gas	3.6%	3.4%	3.8%	4.0%	4.9%	5.2%
Telecommunications	3.4%	3.4%	3.2%	3.3%	3.2%	3.6%

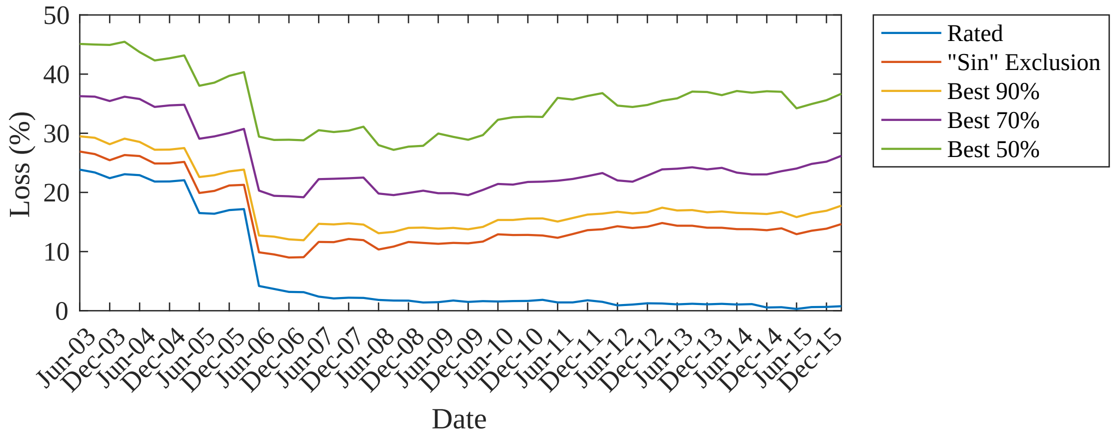


Figure 4.5: Market capitalization loss due to the implementation of the different levels of screening.

4.3 Testing SRI Portfolios Performance

So far, the effects of implementing a SRI strategy have been analysed only in terms of the investment universe composition. One of the main purposes of this research, though, was to verify whether the exclusion of non-SR assets determined a significant loss in terms of efficiency and performance.

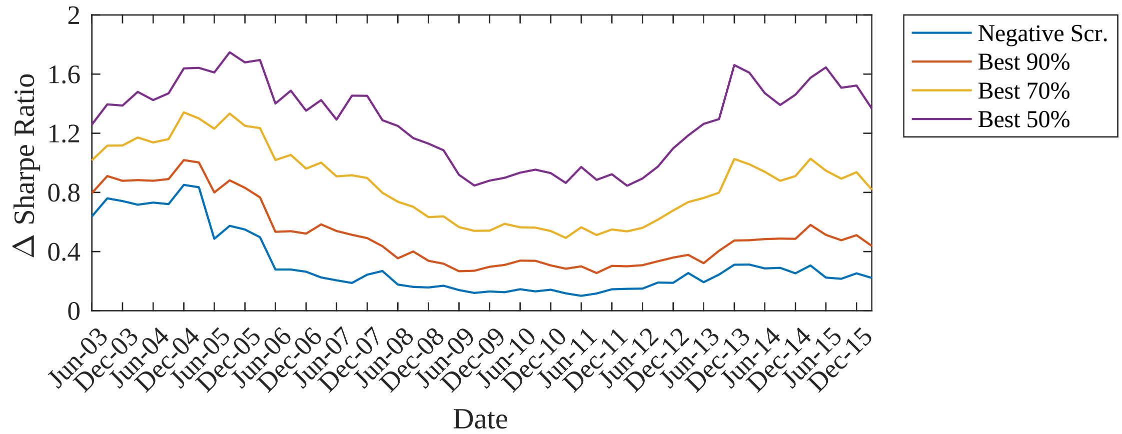
Given a set of risky asset and assuming the existence of a risk-free rate, the maximum achievable performance can be measured, ex-ante, by estimating the Sharpe ratio of the tangency portfolio. This value can be computed as seen in section 3.1, having previously estimated the assets expected returns and covariance matrix.

Figures 4.6a and 4.6b show the loss, in terms of maximum achievable Sharpe ratio, for the negative[13] and the three different levels of positive screening strategies based on the ESG score, with and without short-selling, respectively. For this purpose, expected returns and covariance matrix were estimated on a rolling basis, starting from June 2003 up to March 2016, using, each trimester, the last five years of monthly returns data. As for the risk-free rate, the U.S. one month t-bill rate was used.

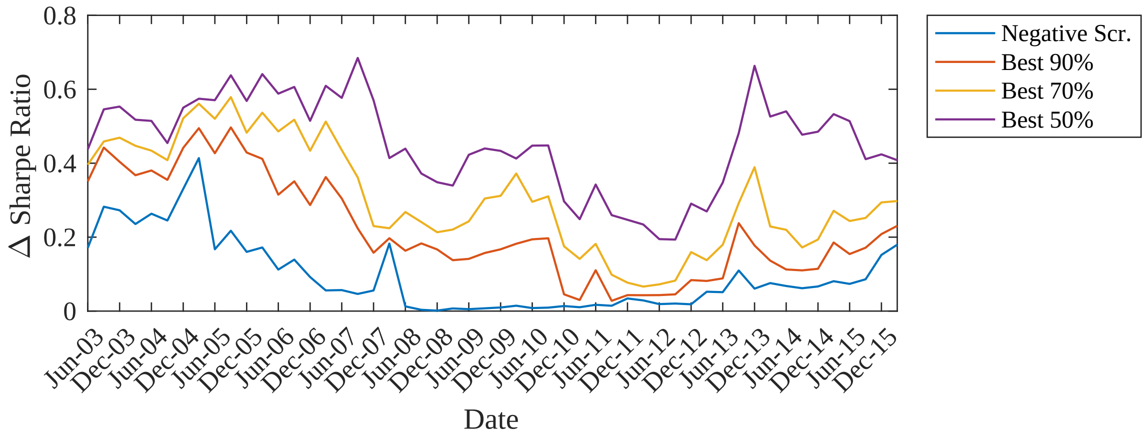
Expected returns were estimated using sample means (as in equation 3.7, p. 31), while the covariance matrix, due to the large number of assets included in the reference investment universe, needed to be estimated employing a factor model as described in section 3.2.2. The chosen risk factors were the three from Fama and French (excess market return, Small Minus Big and High Minus Low) and Carhart's Momentum, all relative to the European region.[14]

[13] From this point forward, the negative screening strategy is characterized by the exclusion of both non rated and "sin" companies.

[14] The time series relative to the risk factors and of the U.S. one month t-bill rate were downloaded from K.R. French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The factors returns available from the website are in U.S. dollars so, prior to be employed in the factor model estimation, they were converted in EUR using monthly exchange rates time series from Banca d'Italia database.



(a) Short-selling



(b) No short-Selling

Figure 4.6: Sharpe ratio loss, relative to the tangency portfolios, due to the implementation of the different levels of screening, with and without short-selling restrictions.

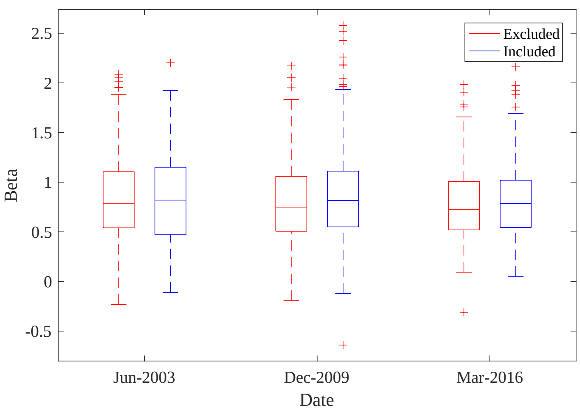
In accordance with MV theory, losses in terms of Sharpe ratio increased with the level of screening; assets exclusion determined a reduction in mean-variance opportunities. This effect was considerably smaller when imposing short-selling restriction and, for both the unrestricted and restricted cases, during the period from June 2008 to June 2013.

While the two plots show the performance loss an SRI investor would have expected to experience (ex-ante), investing in a particular point in time, they do not tell whether that loss was statistically significant or merely due to estimation error. This type of information can be obtained by implementing the efficiency test seen in section 3.3.1.

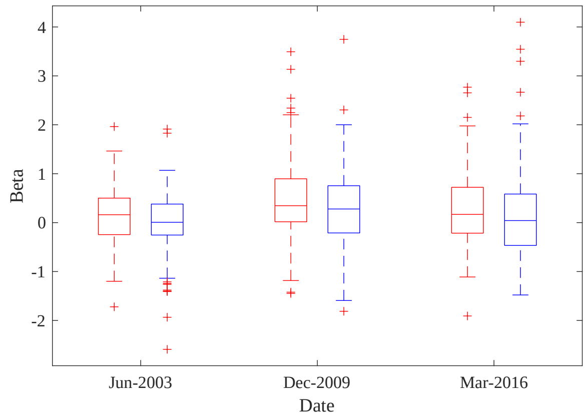
Risk Factors Exposure Comparison

The betas obtained from the four factor model implementation, employed for estimating the covariance matrices, also allowed verifying whether SR and non-SR companies exhibited differences in terms of risk factors exposure.[15]

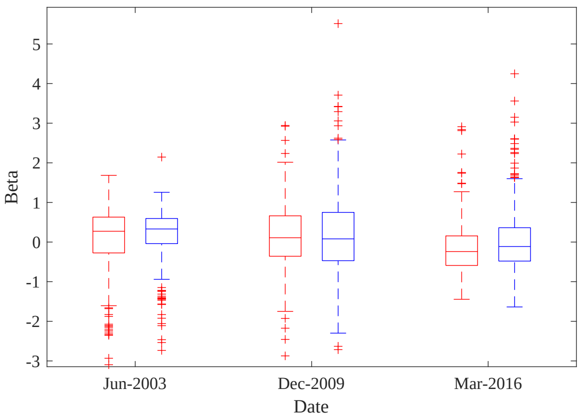
[15]For this type of analysis, the two asset classes were determined by applying up to the 70% screening strategy based on ESG score.



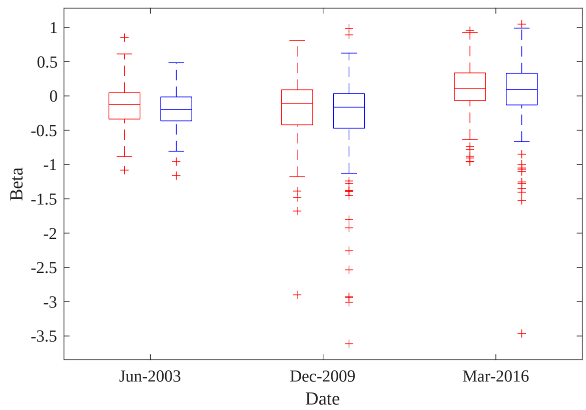
(a) Excess Market Return



(b) Small Minus Big



(c) High Minus Low



(d) Momentum

Figure 4.7: Risk factors exposure comparison between SR (blue) and non-SR (red) companies at three different points in time.

The box-plots[16] in figure 4.7 allow performing a visual comparison between the factor loading distributions, relative to the two asset classes, for the four risk factors at three different points in time: June 2003, December 2009 and March 2016. While no major differences are visible for any of the four risk factors, it seems that, overall, SR companies tend to exhibit a smaller exposure to the Small Minus Big factor compared to non-SR companies; boxes' edges (first and third quartiles) and medians are shifted downwards in all three analysed time frames. This result, although not characterized by statistical significance, is related to what already seen when analysing the effects of the screening processes in terms of market capitalisation: excluded companies were, on average, smaller than the remaining SR ones.

[16]On each box, the central mark indicates the median, while the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

Table 4.2: Assets grouping scheme.

Assets Grouping	
Portfolio N°	Assets Included
1	Non rated companies
2	Companies involved in sin activities
3	Companies with ESG score in first decile (worst 10%)
4	Companies with ESG score in second decile
5	Companies with ESG score in third decile
6	Companies with ESG score in fourth decile
7	Companies with ESG score in fifth decile
8	Companies with ESG score in sixth decile
9	Companies with ESG score in seventh decile
10	Companies with ESG score in eighth decile
11	Companies with ESG score in ninth decile
12	Companies with ESG score in last decile (best 10%)

4.3.1 Testing Efficiency

A major problem encountered in applying the efficiency test to this particular setup was related to the large number of assets involved; in order to estimate the regression parameters in equation 3.11 (p. 35) it is required that the total number of assets is lower than the number of observations or, using the usual notation, $N + K < T$, where K and N are the umber of benchmark and test assets, respectively.

In this case, the test assets should have coincided with the excluded companies (according to the different screening levels), while the benchmark assets with the remaining ones. As already shown in figure 4.3, the total number of assets ($N + K$) almost never fell below 500, way above 60, the number of historical monthly returns taken into account (T).

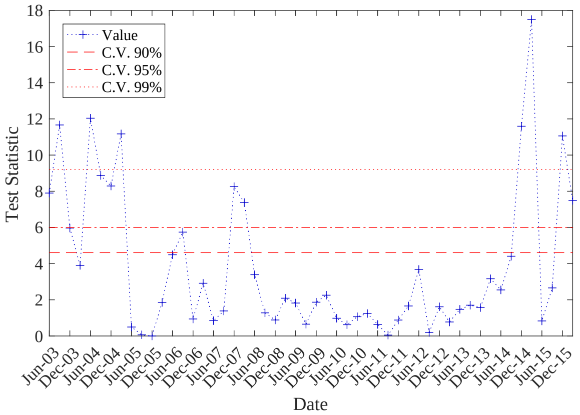
The only viable solution was to reduce dimensionality by grouping assets in a reasonable number of equally-weighted portfolios[17]. The criterion used to create the different portfolios consisted in sorting the assets depending on their SR "status", as described in table 4.2.[18]

Each of the 4 different levels of screening corresponded to the exclusion of a specific number of portfolios:

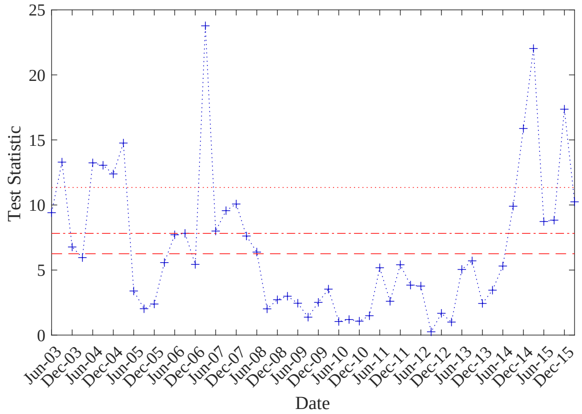
1. Negative screening: portfolios n° 1 and 2.
2. Neg. scr. + selection of best 90%: portfolios n° 1,2 and 3.
3. Neg. scr. + selection of best 70%: portfolios n° 1,2,3,4 and 5.

[17] As a robustness check, the test was also performed using value-weighted portfolios.

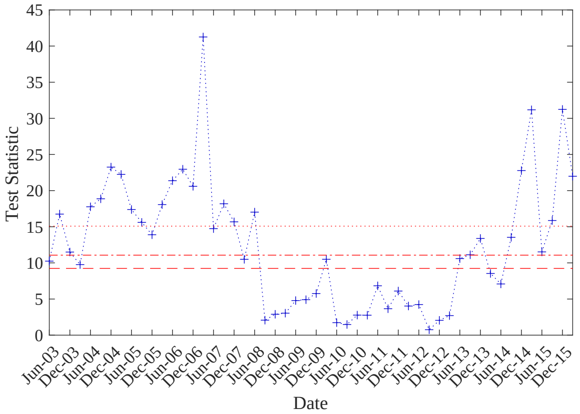
[18] This process was repeated for the other strategies, consisting in the positive screening only, and different score type (ESG, ENV, SOC, GOV). In those cases the number of portfolios was 11 since the distinction in "sin" companies was not needed.



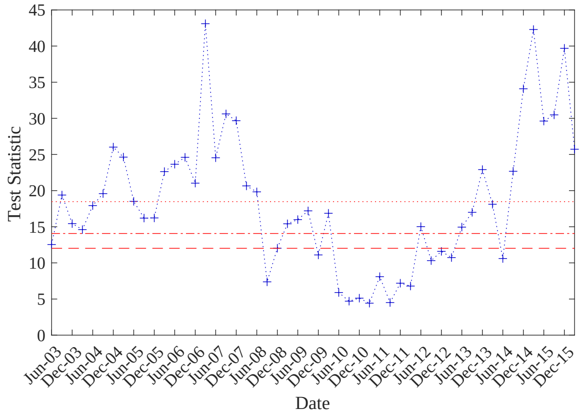
(a) Negative screening



(b) Best 90% selection (ESG)



(c) Best 70% selection (ESG)



(d) Best 50% selection (ESG)

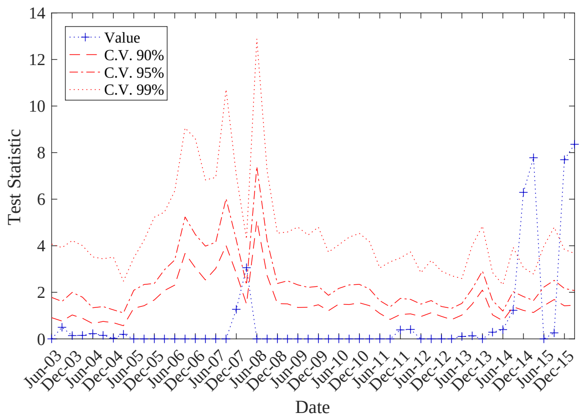
Figure 4.8: Results of the efficiency test for the four different levels of screening. When the test value lies above the critical value the null hypothesis is rejected.

4. Neg. scr. + selection of best 50%: portfolios n° 1,2,3,4,5,6 and 7.

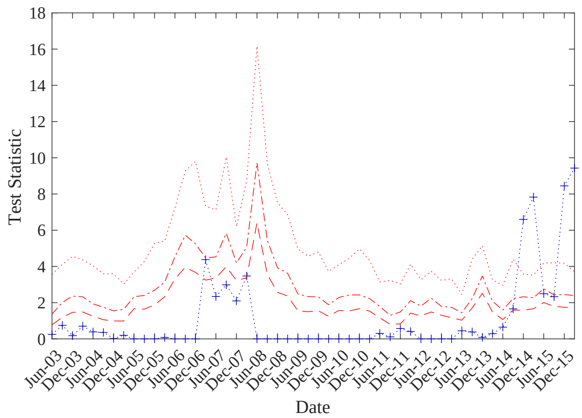
The test was then performed on a quarterly rolling basis, starting from June 2003 up to March 2016, using, for each point in time, the last five years of monthly excess returns.

Figure 4.8 shows, for the four levels of screening, the test statistic and the critical values corresponding to three different significance levels (10%, 5% and 1%).[19] As expected, the frequency of rejection of the null hypothesis, consisting in the efficiency of the subset of SR assets, increased with the level of screening. For the negative screening only, the efficiency hypothesis was rejected, with a significance level of 5%, 11 times out of 52, mostly concentrated at the beginning and at the end of the sample. For the highest degree of screening, instead, rejection cases rose to 36 out of 52. During the period from September 2008 up to June 2014, the exclusion of non-SR assets affected efficiency to a noticeable lesser degree than in the rest of the time frame, for all screening levels.

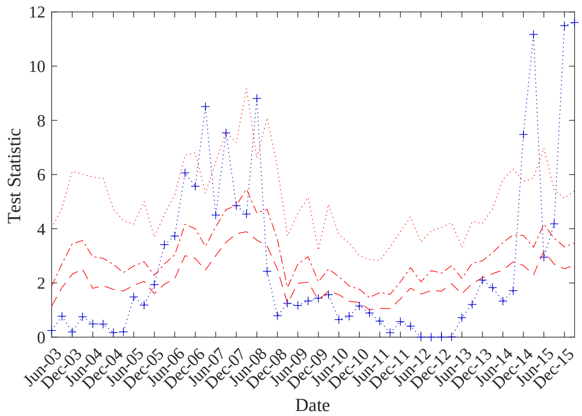
[19] Portfolio returns were assumed to be i.i.d. but non-normally distributed, therefore, the test used was the asymptotically valid one (equation 3.13, p. 36) and critical values were computed as the 90th, 95th and 99th percentiles of a chi-square distribution with N degrees of freedom.



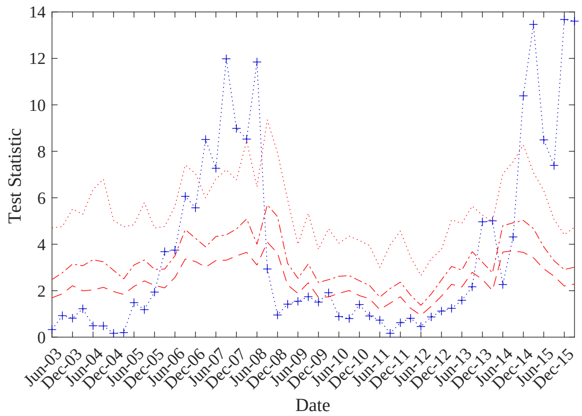
(a) Negative screening



(b) Best 90% selection (ESG)



(c) Best 70% selection (ESG)



(d) Best 50% selection (ESG)

Figure 4.9: Results of the efficiency test in the case of short-selling restrictions. Critical values are not constant since the test statistic distribution is approximated via simulation.

Next, it was important to verify the effects of short-selling restrictions, imposed both on test (non-SR) and benchmark assets (SR). Figure 4.9 shows the test results obtained by applying the simulation based test procedure described in section 3.3.1[20] to the same strategies analysed for the unrestricted case. For all levels of screening, the frequency of rejection of the null hypothesis was greatly reduced compared to the unrestricted case; negative screening alone almost never determined a significant efficiency loss. Again, rejection cases were concentrated in specific time periods: from the beginning of 2006 up to June 2008 and from September 2014 up to March 2016. Short-selling restrictions had a great impact especially during the period from June 2003 up to March 2005; while for the unrestricted case those two years were characterised by the inefficiency of SRI portfolios, this was not true any more when imposing positivity constraints.

It is interesting to highlight that, based on these results, the years between December 2008 and June 2014 represented a favourable period for SR investors; investing in low rated companies did not offer additional mean-variance opportunities. Given the time frame, it is not too

[20]Simulations were performed using the stationary bootstrap procedure with average block size = 3 and 2000 resamples

far-fetched to assume that both the global financial crisis and the European sovereign debt crisis may have contributed in determining the necessary conditions for this to occur and that the effects may have lasted for several years.[21] This result can also help explain what Becchetti et al. (2015) observed in their research. As already mentioned in chapter 2, they found that, in the years following the global financial crisis, SRI funds outperformed conventional ones.

As a robustness check, efficiency tests, for all levels of screening and for both the restricted and unrestricted case, were also performed on value-weighted portfolios[22]. These tests yielded results (included in Appendix A) consistent with those obtained with equally-weighted portfolios and, therefore, lead to the same considerations.

4.3.2 Dynamic Allocation Strategies and Out-of-Sample Performance

Sharpe-Ratio Comparison

The analyses described so far were based on the in-sample performance of screened and non-screened optimal portfolios, computed using historical returns. This research would not have been complete without also comparing the out-of-sample performance (relative to realized returns) of these portfolios.

In particular, the process involved simulating, for all levels of screening, a dynamic allocation strategy with quarterly portfolio rebalancing. This consisted in repeating, for each trimester (from June to 2003 up to December 2015), the following steps:

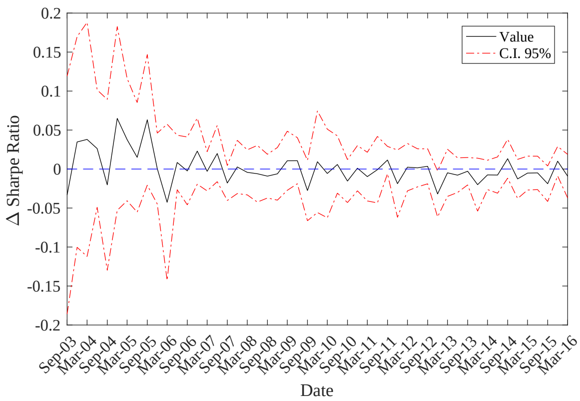
1. Estimation of the Maximum Sharpe portfolio, with and without short-selling restrictions. The parameters of the optimization problems, expected returns and covariance matrix, were estimated, as always, using the last five years of monthly returns. Expected returns were estimated by the sample means while the covariance matrix was obtained by employing the four-factor model.

2. Computation of both daily and monthly realized returns, relative to the subsequent three months, for the Maximum Sharpe portfolio estimated in the previous step.

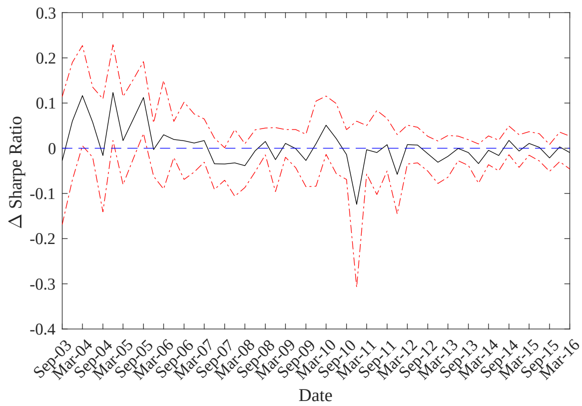
Comparisons were then performed computing the difference in Sharpe ratios between non-SRI and SRI portfolios (for all screening levels) and testing its significance using the methodologies from Ledoit and Wolf (2008), described in section 3.4.

[21] Recall that the test was performed on a rolling basis, using, each time, 5 years of historical returns. This means that on September 2014 test results were still affected by returns dating back to 2009.

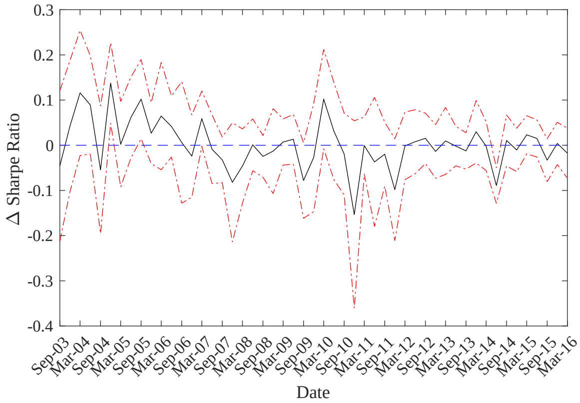
[22] Efficiency tests using value-weighted portfolios were performed starting from June 2006 because of a lack of market capitalisation data prior to 2001.



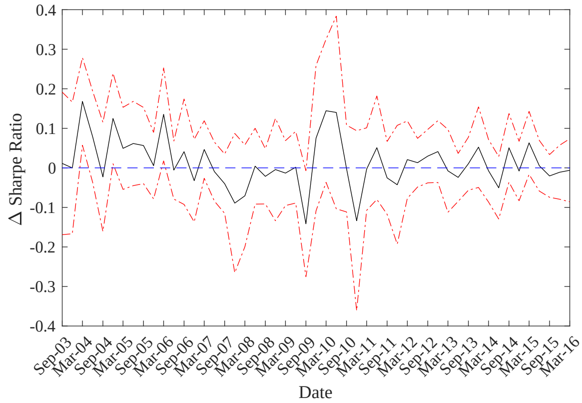
(a) Negative screening



(b) Best 90% selection (ESG)



(c) Best 70% selection (ESG)



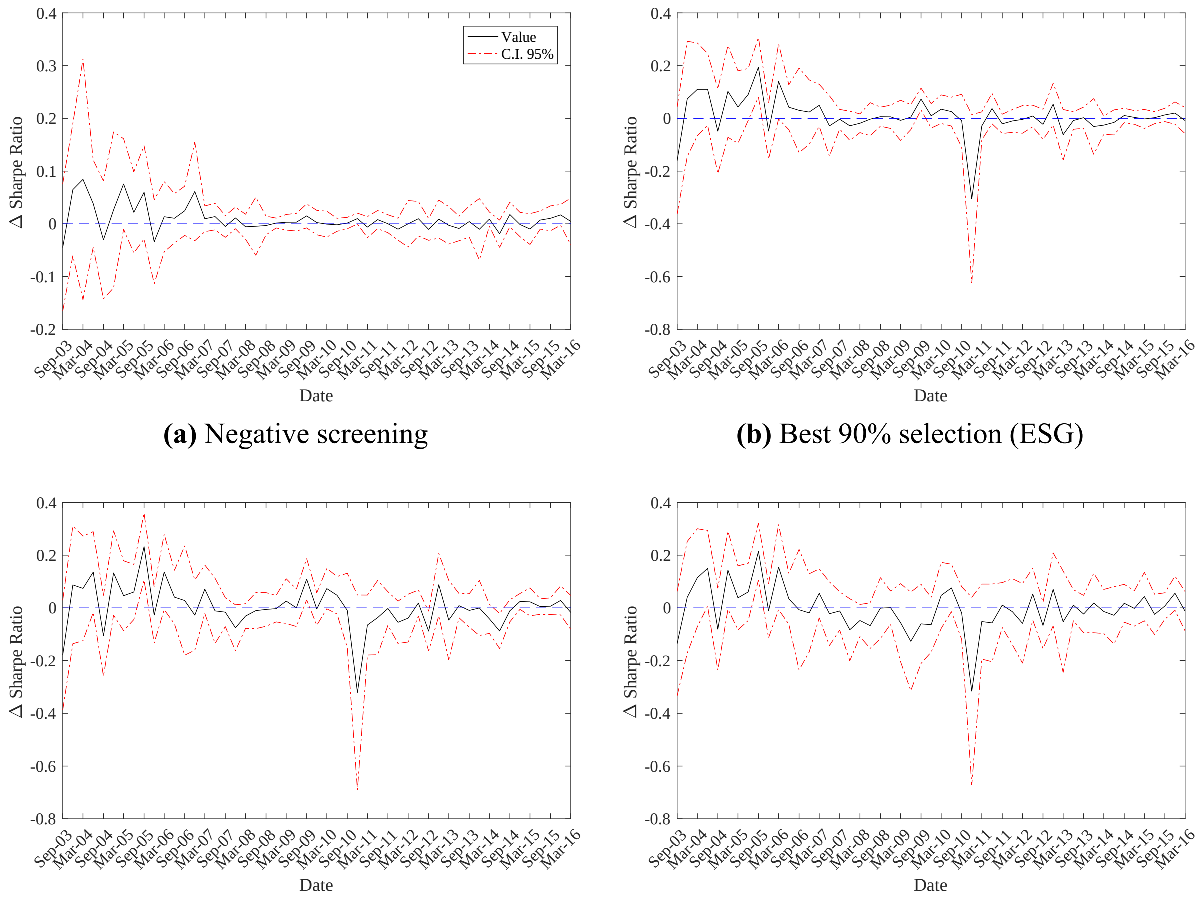
(d) Best 50% selection (ESG)

Figure 4.10: Results of the Sharpe ratio test performed, each trimester, on the last three months of realized daily returns. A negative value means that the screened portfolio outperformed the non-screened one.

Figure 4.10 shows the evolution, through time, of the difference in terms of Sharpe ratio, for the four levels of screening, computed on realized daily returns. For each point in time, the difference is relative to the previous 3 months of returns and the 95% confidence interval was obtained as in equation 3.20 (p. 41), using the HAC inference method. If the confidence interval contains zero, then the difference is not statistically significant.

In all four plots it is visible that the differences mostly hovered around the zero value, not showing a clear dominance of one strategy over the other, and almost never being statistically significant (at the 5% level). Only during the first two years, the non-screened portfolio occasionally outperformed the screened ones with significant gains in terms of Sharpe ratio. It is safe to assume that this result is related to the loss of mean-variance opportunities determined by the exclusion of a large number of non rated assets. Results for portfolios with short-selling restrictions are displayed in figure 4.11. No major differences with respect to the unrestricted case are observable.

Realized monthly returns were then employed to compare the performance of non-screened and screened portfolios over the entire time frame, from July 2003 up to March 2016. The sta-



(a) Negative screening **(b)** Best 90% selection (ESG)

(c) Best 70% selection (ESG) **(d)** Best 50% selection (ESG)

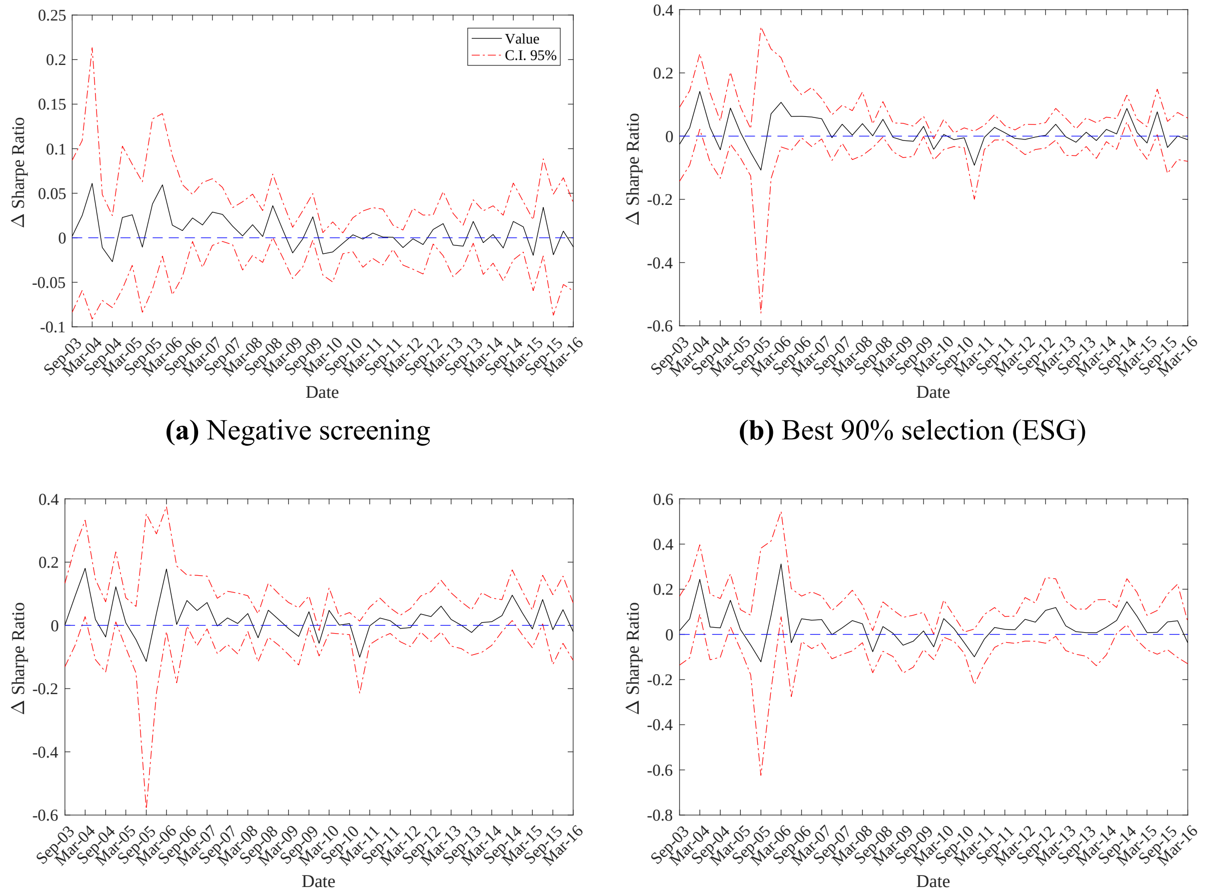
Figure 4.11: Results of the Sharpe ratio test performed on portfolios built imposing short-selling restrictions.

tistical tests, in this case, were performed adopting the bootstrap inference method described in section 3.4.3, which, compared to the HAC methodology, should lead to more robust results. The observed differences in Sharpe ratios, with the relative p-values, obtained as in equation 3.22 (p. 42), are displayed in table 4.3. Consistent with previous results, almost no screened portfolios performed significantly different from the non-screened one. The only exception

Table 4.3: Results of the Sharpe ratio test performed on realized monthly returns relative to the entire period (June 2003-March 2016).

Sharpe Ratio Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Sharpe ratio	p-value	Δ Sharpe ratio	p-value
Negative screening	−0.027**	0.041	0.016	0.171
Best 90% (ESG)	0.035	0.313	−0.025	0.552
Best 90% (ESG)	0.004	0.881	−0.047	0.349
Best 90% (ESG)	0.002	0.963	−0.075	0.193

Notes: A negative value means that the screened portfolio outperformed the non-screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.



(a) Negative screening

(b) Best 90% selection (ESG)

(c) Best 70% selection (ESG)

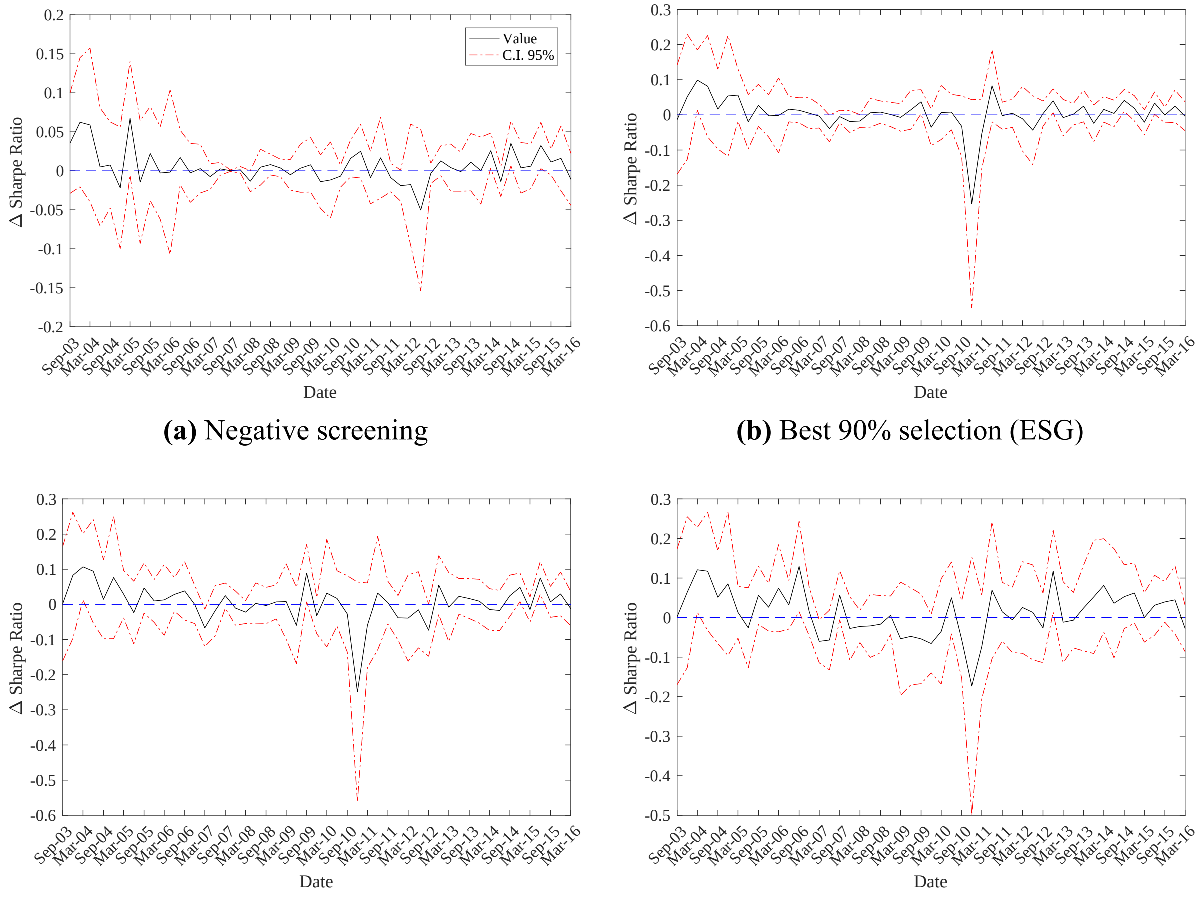
(d) Best 50% selection (ESG)

Figure 4.12: Results of the Sharpe ratio test performed on portfolios built employing equilibrium returns.

regarded the portfolio obtained by the exclusion of non rated and "sin" companies which performed significantly better at the 5% level.

It is known, however, that the mean-variance optimization process is strongly affected by estimation error. Michaud (1989) defined it as an "estimation-error maximizers" that significantly overweights (underweights) securities that have large (small) estimated returns, negative (positive) correlations and small (large) variance and that, especially when not imposing any form of constraints, may lead to extreme and unfeasible asset weights. Additionally, mean-variance optimizations are highly unstable; small changes in the inputs often translate into large changes in the solutions. Combined, these flaws may lead to allocations which are not truly optimal and, consequently, to poor out-of-sample performance. In some cases, optimal portfolios can be outperformed by more naive allocation approaches, such as equally-weighting (Jobson & Korkie, 1980). Michaud (1989) also noted that one of the major contributors to the error-maximizing character of MV optimization is the usage of sample means (computed on historical data) as estimators of expected returns.

It is possible, then, that the results seen so far could have been distorted by the overall poor out-of-sample performance of the estimated optimal portfolios, especially in the unrestricted



(a) Negative screening **(b)** Best 90% selection (ESG)

(c) Best 70% selection (ESG) **(d)** Best 50% selection (ESG)

Figure 4.13: Results of the Sharpe ratio test performed on portfolios built employing equilibrium returns and imposing short-selling restrictions.

case (when allowing for negative positions). As a form of robustness check, sample means were then replaced, in the mean-variance optimization process, by equilibrium returns, computed as described in section 3.2.2 employing the usual four-factor model. These, generally, should lead to more stable and less extreme allocations.

Tests were performed as above and results for the unrestricted and restricted cases are dis-

Table 4.4: Results of the Sharpe ratio test performed on realized monthly returns (portfolios built employing eq. returns) relative to the entire period (June 2003-March 2016).

Sharpe Ratio Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Sharpe ratio	p-value	Δ Sharpe ratio	p-value
Negative screening	−0.004	0.913	0.017	0.222
Best 90% (ESG)	−0.045	0.590	−0.009	0.766
Best 70% (ESG)	−0.024	0.803	0.011	0.766
Best 50% (ESG)	0.043	0.636	0.045	0.314

Notes: A negative value means that the screened portfolio outperformed the non-screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.

played in figure 4.12 and 4.13, respectively.

Again, there was not a clear dominance of one strategy over the other. Non-screened portfolio occasionally performed significantly better during the first two years and, differently from previous results, also towards the end of the sample (for some screening level).

Table 4.4 shows results relative to the entire time period obtained using monthly returns and the bootstrap inference methodology. No significant differences in performance are observable between non-screened and screened portfolios.

Overall, it looks like the exclusion of non-SR assets, at all screening levels, does not determine a significant realized performance loss.

Variance Comparison

An additional test consisted in verifying whether the loss of diversification opportunities, due to the exclusion of non-SR assets, determined a significant increase in portfolio risk, measured by its realized variance.

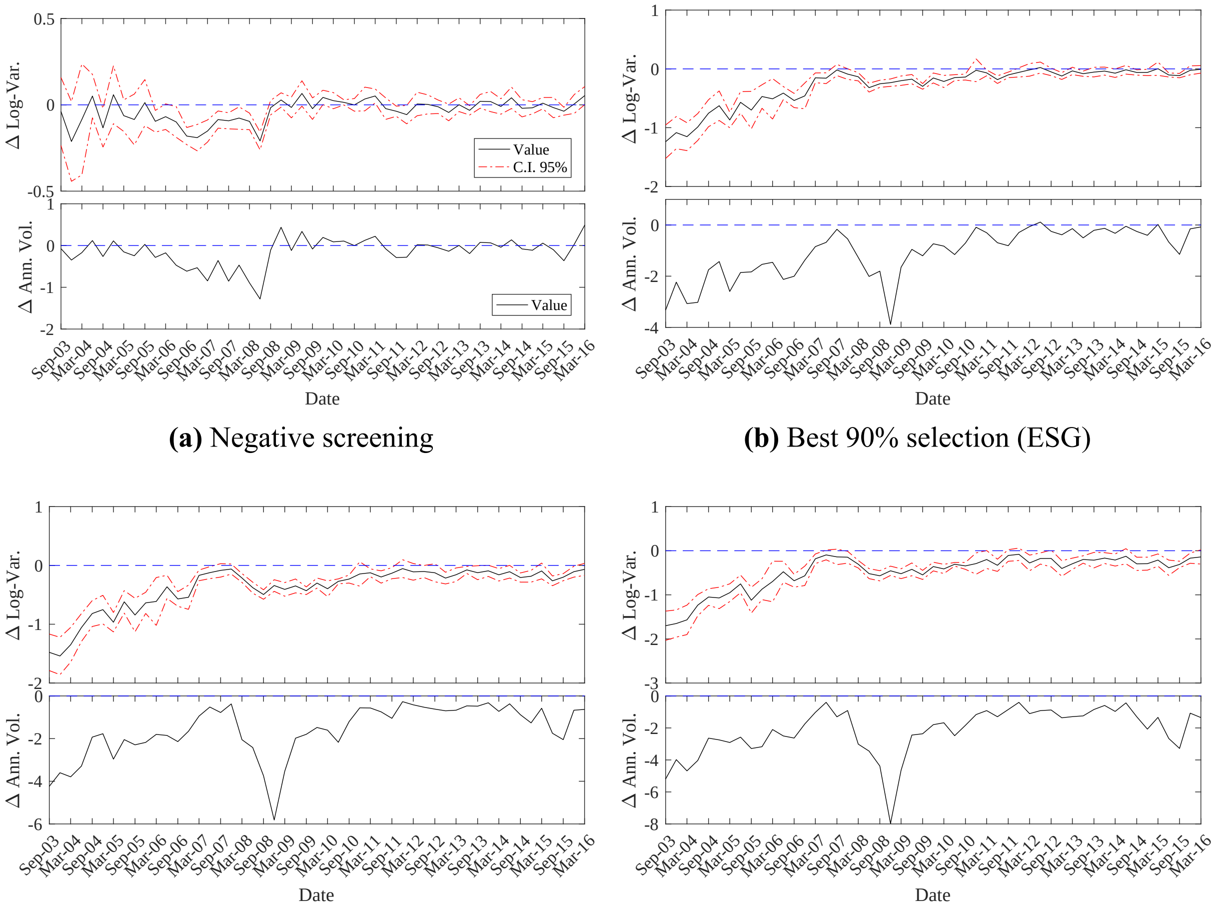
The procedure strongly resembled the one used for the Sharpe ratio comparison, with the main difference being that the target of the allocation strategies was the Global Minimum Variance portfolio, rather than the Maximum Sharpe. Differences in variance were then tested using the methodology from Ledoit and Wolf (2011), described in section 3.5.

Figure 4.14 contains the four plots showing the difference in performance between non-screened and screened portfolios, without short-selling restrictions. Tests were computed, like for the Sharpe ratio, each trimester, using the last three months of realized daily returns. The 95% confidence interval was obtained using the HAC inference method.

Besides statistical significance, it was also important being able to assign an economic significance to the performance differential. Log-variance, in this regard, is not the most appropriate dimension to use since it is not commonly employed as a risk measure and it is, therefore, difficult to interpret. Consequently, in the lower portion of each plot, the difference is also displayed in terms of annualized volatility, making the analysis more straightforward.

Results show that, besides for the first level of screening, the exclusion of non-SR assets almost always determined a significant increase in volatility, with the effect being stronger as the number of excluded assets increased. Other than the first two years, where results were strongly influenced by the exclusion of a large number of non rated companies, the most affected period was from mid 2007 up to the end of 2010 (coinciding with the global financial crisis) with a difference in annualized volatility reaching up to 8% for the highest level of screening.

When introducing short-selling restrictions, as shown in figure 4.15, the increase in volatility seems to have followed the same trend as for the unrestricted case but, overall, the effect was



(a) Negative screening **(b)** Best 90% selection (ESG)

(c) Best 70% selection (ESG) **(d)** Best 50% selection (ESG)

Figure 4.14: Results of the variance test performed on the last three months of realized daily returns. A negative value means that the non-screened portfolio outperformed the screened one. In the lower portion of each plot the difference is displayed in terms of annualized volatility.

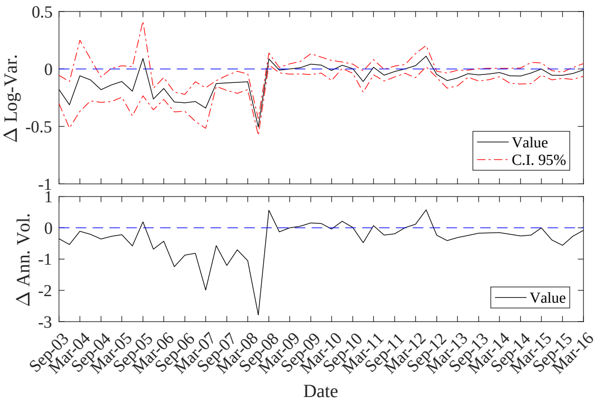
stronger.

Like for the Sharpe ratio, the differences in returns' variance was also tested over the entire period and computed on monthly returns. The results, displayed in table 4.5, confirm what seen so far; exception made for the exclusion of non rated and "sin" companies, both with and

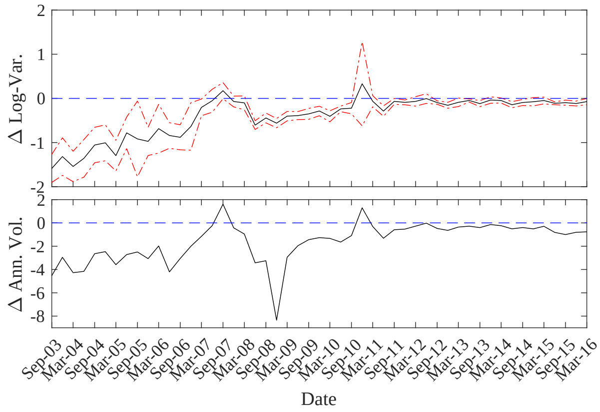
Table 4.5: Results of the variance test performed on realized monthly returns relative to the entire period (June 2003-March 2016).

Variance Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Log-variance	p-value	Δ Log-variance	p-value
Negative screening	−0.011	0.733	−0.068*	0.079
Best 90% (ESG)	−0.123***	0.007	−0.197**	0.028
Best 70% (ESG)	−0.180***	0.005	−0.315***	0.007
Best 50% (ESG)	−0.223**	0.018	−0.340***	0.008

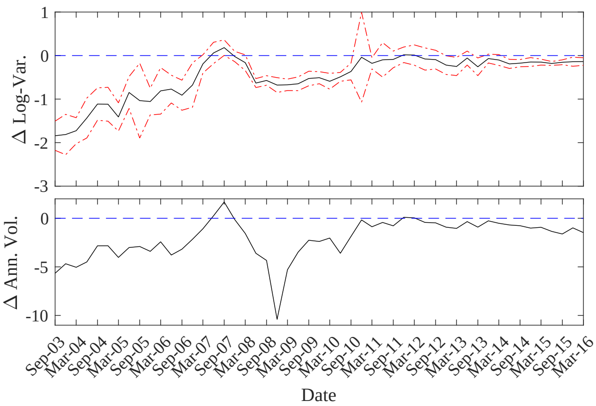
Notes: A negative value means that the non-screened portfolio outperformed the screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.



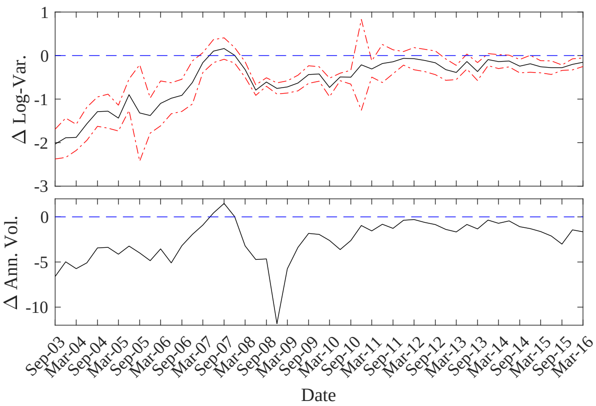
(a) Negative screening



(b) Best 90% selection (ESG)



(c) Best 70% selection (ESG)



(d) Best 50% selection (ESG)

Figure 4.15: Results of the variance test performed on portfolios built imposing short-selling restrictions.

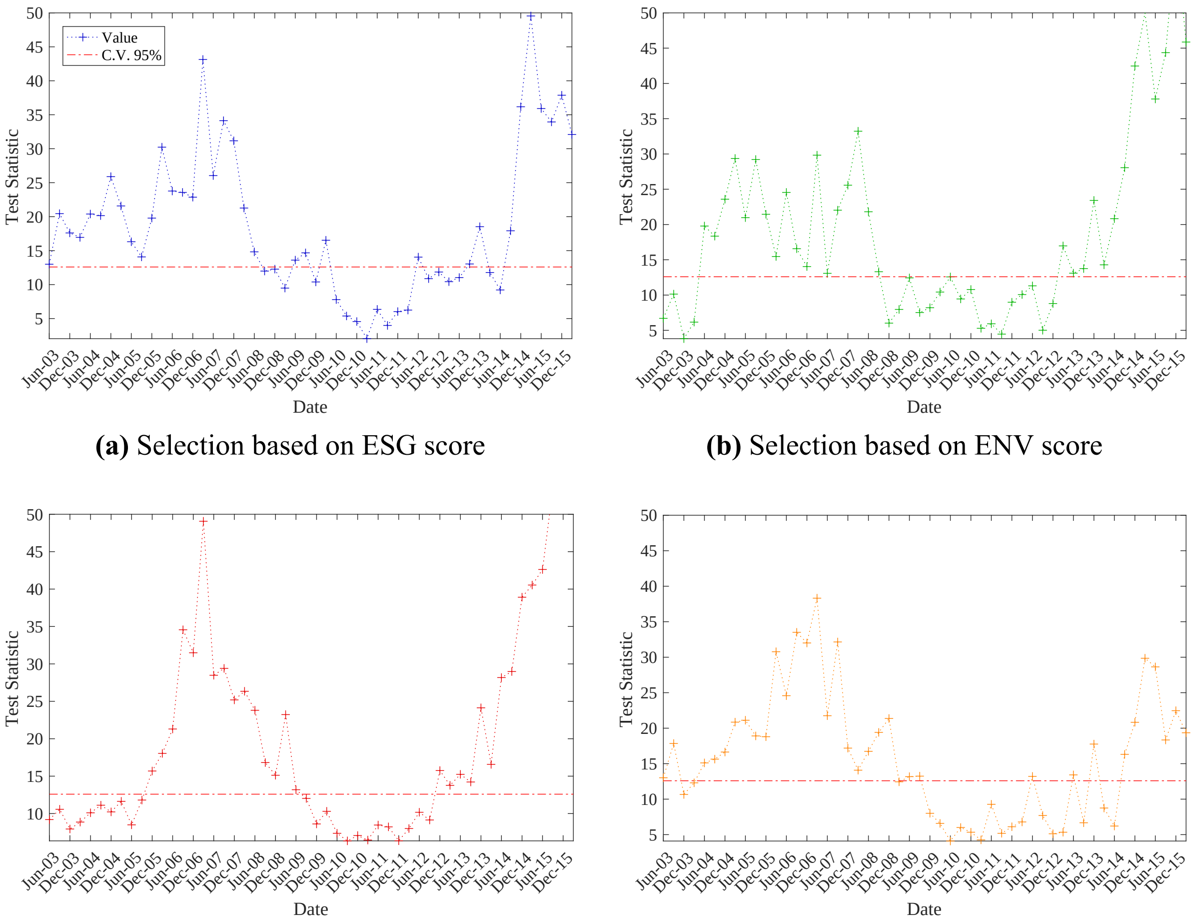
without short-selling restrictions, all the screened portfolios performed significantly worse than the non-screened one at the 5% level or lower.

These results state that mean-variance investors, following a variance-minimizing strategy, are better off by not applying any form of SRI screening (possibly with the exception of negative screening only), especially during high volatility periods. It is up to investors decide whether the gains in terms of ESG performance are sufficient to offset the negative effects (increase in realized variance) of the screening strategies.

4.4 Environment, Social and Governance Screening

The last part of this research was aimed at verifying whether the three different dimensions of social responsibility—environment, social and corporate governance—have a different impact on efficiency and portfolio performance.

As already mentioned, the methodology involved implementing only the positive screening strategy for each one of the four types of score (ENV,SOC,GOV and ESG) and with the usual three different levels of screening (90%, 70% and 50%). Not implementing the negative



(a) Selection based on ESG score

(b) Selection based on ENV score

(c) Selection based on SOC score

(d) Selection based on GOV score

Figure 4.16: Results of the efficiency test for the highest level of screening (Best 50%) based on the four distinct scores (ESG,ENV,SOC,GOV). When the test value lies above the critical value the null hypothesis is rejected.

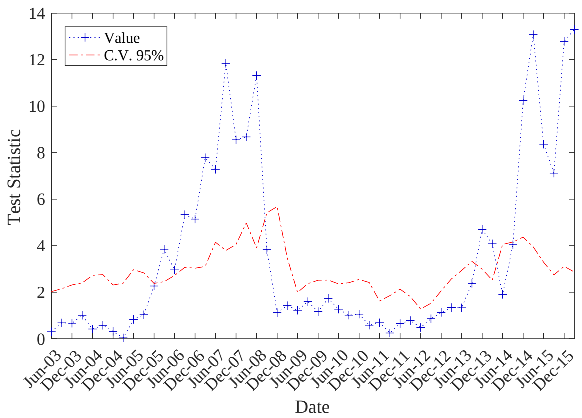
screening meant that companies were not excluded based on their involvement in controversial activities but exclusively on their rating. The purpose was to better isolate the effects of investing following a specific ESG thematic.

4.4.1 Testing Efficiency

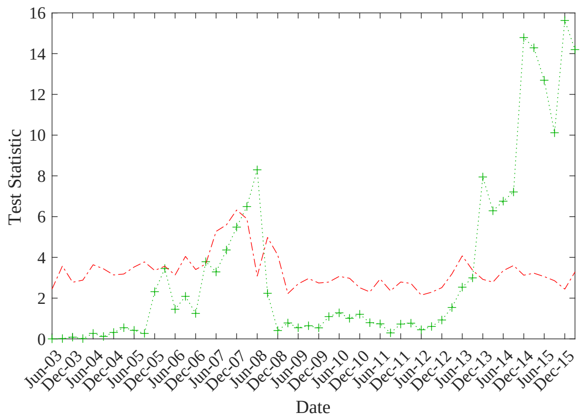
Results of the efficiency tests for the highest level of screening (best 50%), with and without allowing for short-selling, are shown in figures 4.16 and 4.17, respectively[23].

The four strategies, overall, lead to comparable results: cases of rejection of the null hypothesis (at the 5% level) were concentrated at the beginning and the end of the sample, while the period from December 2008 up to the second half of 2013 was characterized by more favourable results for SRI portfolios. The introduction of short-selling restrictions resulted, as usual, in a reduction of rejection cases, especially during the first two years.

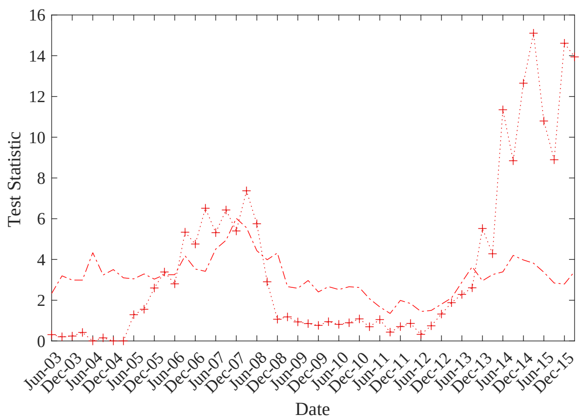
[23] These results are relative to the tests performed on equally-weighted portfolios. For those performed on value-weighted ones see appendix A



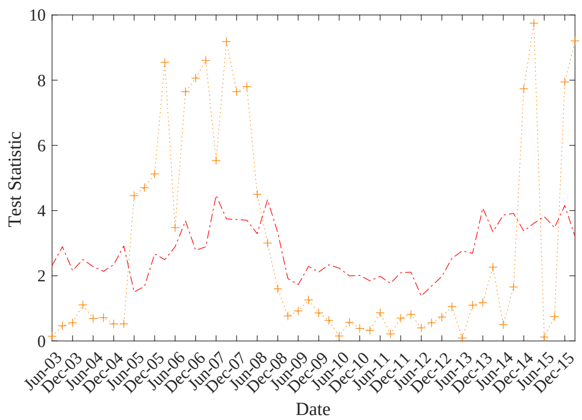
(a) Selection based on ESG score



(b) Selection based on ENV score



(c) Selection based on SOC score



(d) Selection based on GOV score

Figure 4.17: Results of the efficiency test for the highest level of screening (Best 50%) based on the four distinct scores (ESG,ENV,SOC,GOV) in the case of short-selling restrictions. Critical values are not constant since the test statistic distribution is approximated via simulation.

From this analysis it seems that the four selection criteria, based on the different dimensions of ESG performance, all have a very similar impact on efficiency. However, it is still possible to identify at least two cases in which one strategy differed from the others:

1. The social score based strategy was the only one that, when allowing for short-selling, yielded slightly different results; the efficiency hypothesis was never rejected during the first two years of analysis (from June 2003 up to December 2005).

2. When imposing short-selling restrictions, the environment score based strategy almost never had a significant impact during the period from late 2005 to late 2008.

Results for the two remaining degrees of screening, displaying similar levels of homogeneity among the four strategies, are included in Appendix C (figures C.3 and C.4).

4.4.2 Out-of-Sample Performance Comparison

Whether the four screening criteria had different impacts on realized portfolio performances was verified employing the same methodology as in section 4.3.2. This consisted in testing the difference in Sharpe ratio and variance between screened and not screened portfolios, for each screening strategy.

The tests were performed for the period from June 2003 to March 2016 using monthly realized returns of portfolios built following two different dynamic allocation strategies: one aimed at maximizing the Sharpe ratio (for the Sharpe ratio test) and the other at minimizing returns variance (for the variance test).

Results for the Sharpe ratio test, reported in tables 4.6 and 4.7, show that none of the screening strategies significantly underperformed nor outperformed the reference one. The variance test, instead, yielded completely different results (table 4.8), with almost all screened portfolios being significantly (at the 5% level) riskier than the non-screened one. More importantly, in both tests, there are no evident differences between the four screening criteria.

Table 4.6: Results of the Sharpe ratio test performed on realized monthly returns relative to the entire period (June 2003-March 2016).

Sharpe Ratio Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Sharpe ratio	p-value	Δ Sharpe ratio	p-value
Best 90% (ESG)	0.061	0.203	−0.026	0.553
Best 70% (ESG)	0.028	0.369	−0.043	0.403
Best 50% (ESG)	0.013	0.801	−0.066	0.403
Best 90% (ENV)	0.064	0.175	0.011	0.810
Best 70% (ENV)	0.023	0.706	−0.036	0.510
Best 50% (ENV)	−0.015	0.536	−0.051	0.312
Best 90% (SOC)	0.051	0.171	−0.033	0.465
Best 70% (SOC)	0.004	0.792	−0.044	0.388
Best 50% (SOC)	−0.030	0.322	−0.051	0.335
Best 90% (GOV)	0.060	0.238	−0.027	0.605
Best 70% (GOV)	0.065	0.357	−0.035	0.508
Best 50% (GOV)	0.097	0.164	−0.005	0.930

Notes: A negative value means that the screened portfolio outperformed the non-screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.

Table 4.7: Results of the Sharpe ratio test performed on realized monthly returns (portfolios built employing eq. returns) relative to the entire period (June 2003-March 2016).

Sharpe Ratio Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Sharpe ratio	p-value	Δ Sharpe ratio	p-value
Best 90% (ESG)	−0.042	0.550	−0.012	0.642
Best 70% (ESG)	−0.046	0.589	0.003	0.922
Best 50% (ESG)	0.024	0.790	0.016	0.675
Best 90% (ENV)	−0.020	0.653	0.016	0.541
Best 70% (ENV)	−0.060	0.485	−0.032	0.379
Best 50% (ENV)	−0.008	0.935	−0.031	0.430
Best 90% (SOC)	−0.050	0.497	−0.034	0.280
Best 70% (SOC)	−0.058	0.483	0.012	0.730
Best 50% (SOC)	−0.025	0.791	0.021	0.646
Best 90% (GOV)	−0.051	0.434	−0.020	0.475
Best 70% (GOV)	−0.022	0.815	0.013	0.731
Best 50% (GOV)	0.014	0.883	0.056	0.311

Notes: A negative value means that the screened portfolio outperformed the non-screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.

Table 4.8: Results of the variance test performed on realized monthly returns relative to the entire period (June 2003-March 2016).

Variance Test				
	Short-Selling		No Short-Selling	
Screening Level	Δ Log-variance	p-value	Δ Log-variance	p-value
Best 90% (ESG)	−0.126***	0.002	−0.187**	0.023
Best 70% (ESG)	−0.181***	0.001	−0.298***	0.003
Best 50% (ESG)	−0.217***	0.001	−0.322***	0.003
Best 90% (ENV)	−0.095***	0.009	−0.149*	0.057
Best 70% (ENV)	−0.158***	0.001	−0.263***	0.015
Best 50% (ENV)	−0.200***	0.004	−0.339***	0.005
Best 90% (SOC)	−0.126***	0.007	−0.198**	0.023
Best 70% (SOC)	−0.177***	0.007	−0.258**	0.011
Best 50% (SOC)	−0.204***	0.000	−0.291***	0.007
Best 90% (GOV)	−0.114**	0.017	−0.164*	0.051
Best 70% (GOV)	−0.184***	0.000	−0.267***	0.007
Best 50% (GOV)	−0.276***	0.002	−0.400***	0.008

Notes: A negative value means that the non-screened portfolio outperformed the screened one.
*Rejection of the null hypothesis at the 10% level.
**Rejection of the null hypothesis at the 5% level.
***Rejection of the null hypothesis at the 1% level.

Chapter 5

Conclusions

The main purpose of this master thesis was to verify whether the introduction, in the investment process, of ethical and moral principles, commonly achieved through the implementation of one or more of the SRI strategies seen in chapter 1, comes at a cost for investors. In particular, the focus was on the effects that choosing not to invest in certain companies because of their (relatively) bad ESG performance or involvement in controversial activities has on mean-variance efficiency and portfolio performances.

As reported in Chapter 1, the SRI phenomenon is constantly gaining in popularity and, since its beginning, whether SR investors have to sacrifice financial performance in order to pursuit their social goals has been a highly debated topic.

In the last two decades, many studies, some of which are reviewed in chapter 2, managed to shed some light on this matter, making use of different methodologies (Jensen's alpha comparison, multi-factor analysis, spanning test, etc.) applied to a variety of datasets and forms of investment.

Personally, I would divide these studies into two groups:

1. Those aimed at verifying whether stocks of socially responsible companies underperform those of conventional ones.

2. Those aimed at investigating the effects, in terms of mean-variance opportunities, of the exclusion of non-SR companies from the investment universe.

For the first group, findings were, with some exceptions, consistent with the "no effect" hypothesis, stating that there are no significant differences between (risk-adjusted) expected returns of socially responsible and conventional stocks. Regarding the second group, researchers found that, when not imposing constraint to the mean-variance optimization problem, optimal portfolios consisting exclusively of SR companies' stocks perform significantly worse than their non-screened counterparts. Additional results showed that losses in terms of mean-variance opportunities are mainly due to foregone risk reduction and not in terms of foregone returns. Moreover, when imposing short-selling restrictions these losses tend to disappear.

This research, clearly belonging to the second group, introduced some novelties with respect to previous comparable work, in particular those by Galema et al. (2009), and Herzel et al. (2012). First, it focused on the European equity market, taking as reference index the STOXX Europe 600 and using ESG ratings from Asset4. Second, the effects of excluding non-SR companies were tested, using the efficiency test from Gibbons et al. (1989), on a quarterly rolling basis starting from June 2003 and ending on March 2016. This allowed verifying whether these effects are constant through time or vary depending on market conditions. Lastly, an important part of the research was dedicated, employing the methodology from Ledoit and Wolf (2008, 2011), at comparing the realized performances of screened and non-screened portfolios.

5.1 Main Findings

The effects of the exclusion of non-SR companies, implemented via a combination of a negative screening and three levels of positive screening strategies, were first analysed in terms of investment universe composition. Results showed that companies involved in “sin” activities, excluded by the negative screening strategy, represented a significant and slightly growing fraction of the rated ones, starting from 7% in 2003 up to around 11% in more recent years. The screening process, even though adopting a Best-in-Universe approach, did not drastically affect the investment universe in terms of industry composition. This means that, at least using Asset4’s ratings, screened and non-screened investment universes does not differ in terms of exposure to industry specific risks. The total market capitalisation analysis showed, instead, that large companies, on average, tend to have higher ESG ratings than smaller ones, result possibly confirmed by the risk factor loadings comparison, where SR companies showed a smaller exposure to the Small Minus Big factor.

The (ex-ante) efficiency tests, performed on a rolling basis, showed that the effects of excluding from the investment universe non-SR companies are not constant through time. For the negative screening strategy, efficiency losses were almost never statistically significant, with some exception at the beginning and at the end of the analysed time frame. For the remaining screening strategies, the frequency of statistically significant losses increased with the level of screening.

An interesting result was that, for all level of screening, the cases of efficiency of screened portfolios were concentrated between September 2008 and June 2014. The tests performed during this time frame were based on returns relative to a period which included both the global financial crisis and the European sovereign debt crisis. This could mean that, during market crisis, investing in non-SR companies does not offer significant additional mean-variance opportunities.

The introduction of short-selling restrictions, which, for many investors, represents a more realistic scenario, drastically increased the number of cases of efficiency of screened portfolios, for all screening level. This is consistent with the results obtained by Galema et al. (2009) and Herzel et al. (2012).

Tests for differences in realized performance of screened and non-screened portfolios showed that socially responsible investors, following a Sharpe ratio maximizing strategy, are not worse off compared to their conventional counterparts, even at the highest levels of screening. Opposite results for investors following variance minimizing strategies, for which the exclusion of non-SR companies determines significant losses in risk reduction opportunities. This was noticeable especially during the period following the global financial crisis. In both cases, the introduction of short-selling restrictions did not drastically affect the results.

Lastly, additional tests showed that socially responsible investors, following specific ESG thematics (environment, social, governance), all sustain similar effects, both in terms of ex-ante efficiency and realized portfolio performances, to those who adopt a less specific screening criteria (average ESG score).

5.2 Recommendations for Future Research

During the carrying out of this master thesis I realized that both the research area and the methodology are well suitable for future developments and improvements. What follows is a list of suggestions which would contribute to further expand the research and obtain more robust results:

Different market regions: The first thing I have noticed, which, as already mentioned, lead me to chose the European equity market as the object of this research, is that the majority of the existing literature focuses on the North American region. It would be interesting, data availability permitting, applying the same methodology to other, less investigated, market regions, like, for example, Asia. The major problem to overcome would be the lack of SRI data relative to Asian firms. As seen in Chapter 1, in this region, SRI strategies are still scarcely adopted with SRI assets representing, in late 2014, less than 1% of the total managed assets. Further research, relative to this market region, would probably also help increase the adoption rate (as long as the findings are positive for socially responsible investors).

Different rating providers: What companies to exclude from the investment universe is, ultimately, determined by the methodology adopted by the ESG rating provider; different methodologies may result in the exclusion of different companies. To test whether the impact on portfolio performance, determined by application of the SRI strategies, is affected by the rating methodology it would be sufficient to retrieve the ESG data from a different provider, keeping everything else constant (investment universe, screening criteria, testing methodology etc.).

Related to this matter is the fact that, with the methodology here adopted, it is not have been possible to determine whether the effects of the SRI strategies application were driven by the characteristics of the excluded companies (relatively low ESG ratings or involvement in controversial activities) or merely by their number. This information could be obtained through simulation, randomly assigning ESG ratings to the constituent companies of the investment universe. Repeating this process multiple times and, again, keeping constant everything else, it would be possible to compute a confidence interval for the statistics of interest (ex-ante Sharpe ratio loss, market capitalization loss, etc.). If the statistic relative to the original case falls in the confidence interval it would mean that the result is mostly determined by the number of assets excluded and not by their characteristics. Herzel et al. (2012) implemented a similar procedure defining it as "Random screening".

Investment universe size: The exclusion of non-SR assets may have different effects depending on the initial size of the investment universe. For this research, a relatively large index, the STOXX Europe 600, was used as reference. Further research could be based on even larger sized indices, such as the STOXX Europe Total Market (with over 1000 constituents), or smaller ones, maybe focusing on specific European countries or company characteristics (capitalization, industry, etc.). This would also allow obtaining results useful to socially responsible investors following more specific screening criteria.

Introduction of additional constraints: Short-selling restriction was the only type of constraint taken into account in this research. While for the efficiency tests it is not possible to investigate the effects of the introduction of additional constraints (the methodology does not allow it), it would be possible to apply the Sharpe ratio and variance tests to the realized returns of portfolios built imposing, for example, size, group and turnover constraints. Unfortunately, the computational resources required to solve the mean-variance optimization process increase with the number of constraints and assets involved. With more than 500 assets, like in this case, the time required to perform all the necessary computation would have represented a serious issue.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817. doi:10.2307/2938229

Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, *27*(11), 1101–1122. doi:10.1002/smj.557

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance*, *29*(7), 1751–1767. doi:10.1016/j.jbankfin.2004.06.035

Becchetti, L., Ciciretti, R., Dalò, A., & Herzel, S. (2015). Socially responsible and conventional investment funds: Performance comparison and the global financial crisis. *Applied Economics*, *47*(25), 2541–2562. doi:10.1080/00036846.2014.1000517

Black, F., & Litterman, R. (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income*, *1*(2), 7–18. doi:10.3905/jfi.1991.408013

Black, F., & Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, *48*(5), 28–43. doi:10.2469/faj.v48.n5.28

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57. doi:10.2307/2329556

De Roon, F. A., & Nijman, T. E. (2001). Testing for mean-variance spanning: A survey. *Journal of Empirical Finance*, *8*(2), 111–155. doi:10.1016/S0927-5398(01)00022-6

De Roon, F. A., Nijman, T. E., & Werker, B. J. M. (2001). Testing for mean-variance spanning with short sales constraints and transaction costs: The case of emerging markets. *The Journal of Finance*, *56*(2), 721–742. doi:10.1111/0022-1082.00343

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analysts Journal*, *61*(2), 51–63. doi:10.2469/faj.v61.n2.2716

Diltz, J. D. (1995). The private cost of socially responsible investing. *Applied Financial Economics*, *5*(2), 69–77. doi:10.1080/758529174

Domini, A. L. (2001). Past is prologue: How the SRI movement started, and how it continuously evolves. In *Socially responsible investing. making a difference and making money* (1st ed., pp. 28–48). Dearborn Trade.

Eling, M., & Schuhmacher, F. (2007). Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking & Finance*, *31*(9), 2632–2647. doi:10.1016/j.jbankfin.2006.09.015

European Fund and Asset Management Association. (2014). *EFAMA Report on Responsible Investment.*

Eurosif. (2013). *Shareholder Stewardship: European ESG Engagement Practices 2013.*

Eurosif. (2016). *European SRI Study.*

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. doi:10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, *51*(1), 55–84. doi:10.1111/j.1540-6261.1996.tb05202.x

Fan, J., Fan, Y., & Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, *147*(1), 186–197. doi:10.1016/j.jeconom.2008.09.017

Galema, R., Scholtens, B., & Plantinga, A. (2009). The cost of socially responsible portfolios: Testing for mean-variance spanning. *SSRN Electronic Journal*, 1–28. doi:10.2139/ssrn.1086560

Geczy, C. C., Stambaugh, R. F., & Levin, D. (2003). Investing in socially responsible mutual funds. *SSRN Electronic Journal*, 1–55. doi:10.2139/ssrn.416380

Gibbons, M. R., Ross, S. a., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121. doi:10.2307/1913625

Glen, J., & Jorion, P. (1993). Currency Hedging for International Portfolios. *The Journal of Finance*, *48*(5), 1865. doi:10.2307/2329070

Global Sustainable Investment Alliance. (2014). *Global Sustainable Investment review.*

Hamilton, S., Jo, H., & Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, *49*(6), 62–66. doi:10.2469/faj.v49.n6.62

Herzel, S., Nicolosi, M., & Stărică, C. (2012). The cost of sustainability in optimal portfolio decisions. *The European Journal of Finance*, *18*(3-4), 333–349. doi:10.1080/1351847X.2011.587521

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15–36. doi:10.1016/j.jfineco.2008.09.001

Huberman, G., & Kandel, S. (1987). Mean-variance spanning. *The Journal of Finance*, *42*(4), 873–888. doi:10.2307/2328296

Jensen, M. C. (1968). The performance of mutual Funds in the period 1945-1964. *The Journal of Finance*, *23*(2), 389–416. doi:10.1111/j.1540-6261.1968.tb00815.x

Jobson, J. D. (1982). Potential performance and tests of portfolio efficiency. *Journal of Financial Economics*, *10*(4), 433–466. doi:10.1016/0304-405X(82)90019-8

Jobson, J. D., & Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, *75*(371), 544–554. doi:10.1080/01621459.1980.10477507

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, *36*(4), 889–908. doi:10.2307/2327554

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*(5), 908–922. doi:10.1111/j.1468-036X.2007.00402.x

Kinder, P. D., & Domini, A. L. (1997). Social screening: Paradigms old and new. *The Journal of Investing*, *6*(4), 12–19. doi:10.3905/joi.1997.408443

Ledoit, O., & Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*(4), 110–119. doi:10.3905/jpm.2004.110

Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. doi:10.1016/S0047-259X(03)00096-4

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859. doi:10.1016/j.jempfin.2008.03.002

Ledoit, O., & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *2011*(55), 86–89. doi:10.1002/wilm.10036

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13. doi:10.2307/1924119

Louche, C., Arenas, D., & van Cranenburgh, K. C. (2012). From preaching to investing: Attitudes of religious organisations towards responsible investment. *Journal of Business Ethics*, *110*(3), 301–320. doi:10.1007/s10551-011-1155-8

Louche, C., & Lydenberg, S. (2006). Socially Responsible Investment: Differences between Europe and United States. *Vlerick Leuven Gent Working Paper Series*, *22*, 1–37.

Louche, C., & Lydenberg, S. (2010). Responsible Investing. In *Finance ethics* (pp. 393–417). Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9781118266298.ch21

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. doi:10.1111/j.1540-6261.1952.tb01525.x

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*(1), 31–42. doi:10.2469/faj.v45.n1.31

Novethic. (2014). *Overview of ESG Rating Agencies.*

Paredes-Gazquez, J. D., Lazcano Benito, L., & De la Cuesta Gonzalez, M. (2014). Drivers and barriers of Environmental, Social and Governance information in investment decision-making: The Spanish case. *International Journal of Business and Management*, *9*(9), 16–28. doi:10.5539/ijbm.v9n9p16

Pastorello, S. (2001). La frontiera efficiente. In *Rischio e rendimento. teoria finanziaria e applicazioni econometriche* (pp. 51–76). Il Mulino.

Politis, D. N., & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the limits of bootstrap* (pp. 263–270). New York, NY: John Wiley.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313. doi:10.1080/01621459.1994.10476870

Porter, M. E., & van der Linde, C. (1995). Green and competitive: Ending the stalemate. *Harvard Business Review*, *73*(5), 120–134.

Renneboog, L., ter Horst, J. R., & Zhang, C. (2007). The price of ethics: Evidence from socially responsible mutual funds. *SSRN Electronic Journal*. doi:10.2139/ssrn.985265

Ruppert, D. (2011). Factor models and principal components. In *Statistics and data analysis for financial engineering* (pp. 443–476). New York, NY: Springer New York. doi:10.1007/978-1-4419-7787-8_17

Sandberg, J., Juravle, C., Hedesström, T. M., & Hamilton, I. (2009). The heterogeneity of socially responsible investment. *Journal of Business Ethics*, *87*(4), 519–533. doi:10.1007/s10551-008-9956-0

Schröder, M. (2007). Is there a difference? The performance characteristics of SRI equity indices. *Journal of Business Finance & Accounting*, *34*(1-2), 331–348. doi:10.1111/j.1468-5957.2006.00647.x

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425. doi:10.2307/2977928

Sparkes, R. (2001). Ethical investment: Whose ethics, which investment? *Business Ethics: A European Review*, *10*(3), 194–205. doi:10.1111/1467-8608.00233

Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, *65*(4), 33–46. doi:10.2469/faj.v65.n4.5

Stone, B. K., Guerard, J. B. J., Gultekin, M. N., & Adams, G. (2001). Socially responsible investment screening : Strong evidence of no significant cost for actively managed portfolios. *Dept. of Finance. Provo, Brigham Young University*.

Walley, N., & Whitehead, B. (1994). It's not easy being green. *Harvard Business Review*, *72*(3), 46–51.

Zhou, G., & Kan, R. (2012). Tests of mean-variance spanning. *Annals of Economics and Finance*, *13*(1), 145–193.

Appendix A

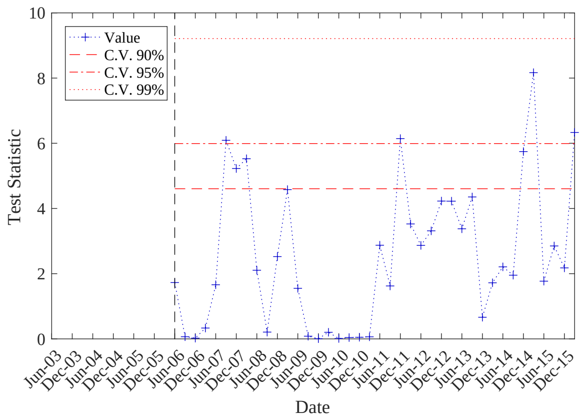
Efficiency Test Using Value-Weighted Portfolios

As described in Chapter 4, a major issue encountered when implementing the mean-variance efficiency test was related to the number of assets included in the analysis; the test is only valid when this is lower than the number of observations (T). The adopted solution involved grouping assets in 12 equally-weighted portfolios based on their rating status, as described in table 4.2. As already mentioned in the chapter, the tests were also performed, as a robustness check, on value-weighted portfolios, results of which are reported in this appendix. Unfortunately, since it is not have been possible to retrieve market capitalization data prior to May 2001, the tests were performed starting from June 2006.

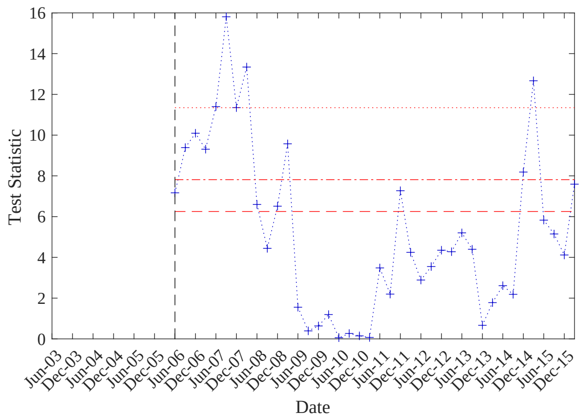
Results of the efficiency tests for the usual four levels of screening are displayed in figure A.1. It is visible that these results strongly resemble those obtained using equally-weighted portfolios, with cases of acceptance of the null hypothesis concentrated in the period from late 2008 up to mid 2014. A minor difference regards the last year of analysis, which is not characterized by the inefficiency of SRI portfolios as clearly as for the equally-weighted case. This is true especially when introducing short-selling restrictions.

Figure A.2 displays the results of the efficiency tests, relative to the highest level of screening, aimed at verifying whether the three different dimensions of social responsibility had a different impact on portfolio efficiency. In this case too, test results using value-weighted portfolios are consistent with those obtained with equally-weighted ones.

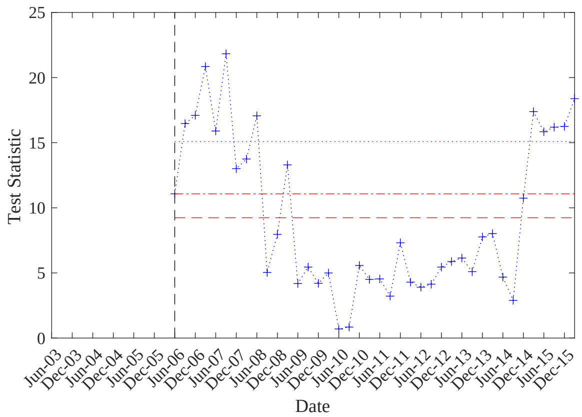
Overall, the results displayed in this appendix confirm what already stated in chapter 4.



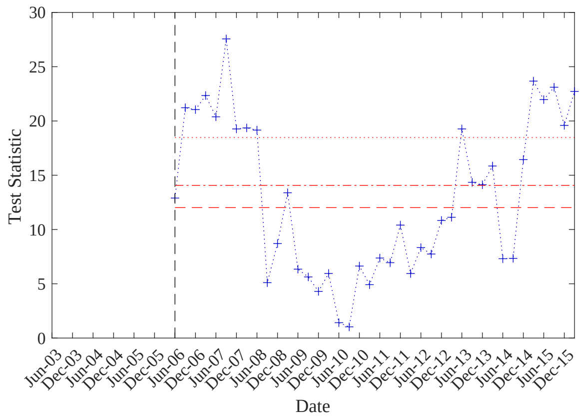
(a) Negative screening



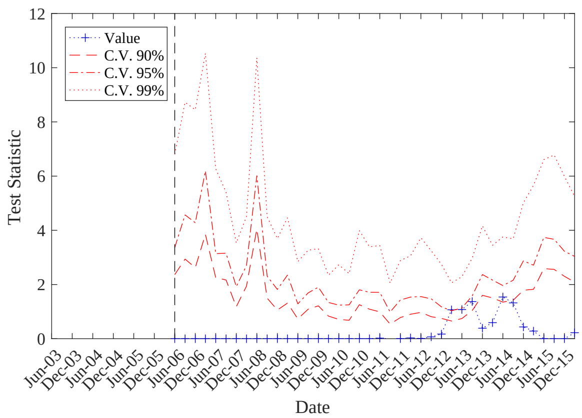
(b) Best 90% selection



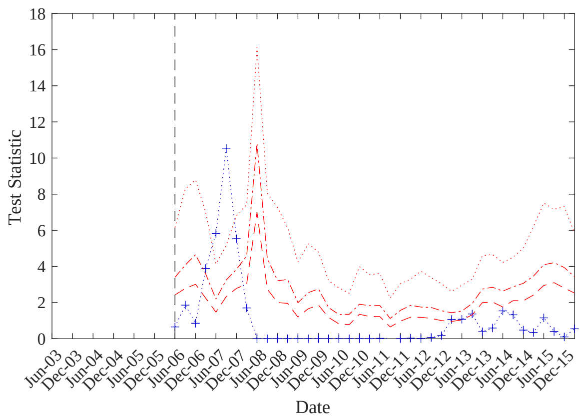
(c) Best 70% selection



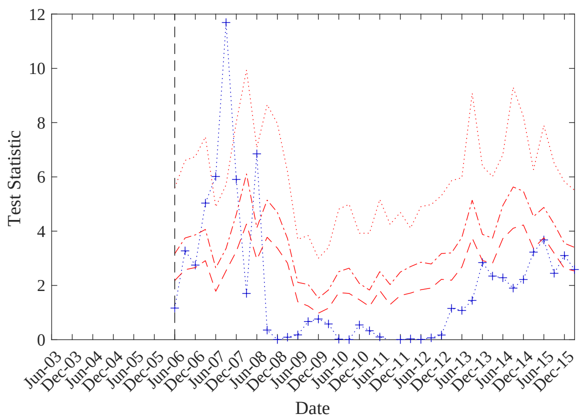
(d) Best 50% selection



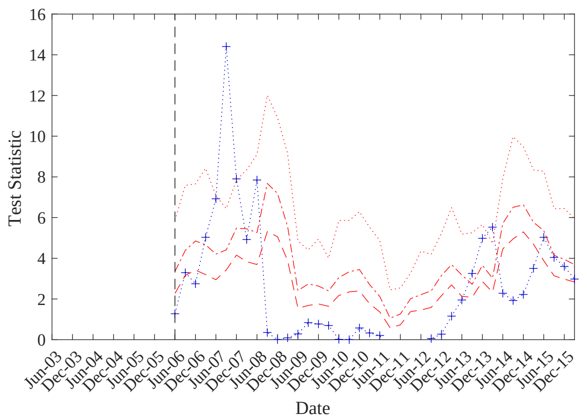
(e) Negative screening (NSS)



(f) Best 90% selection (NSS)

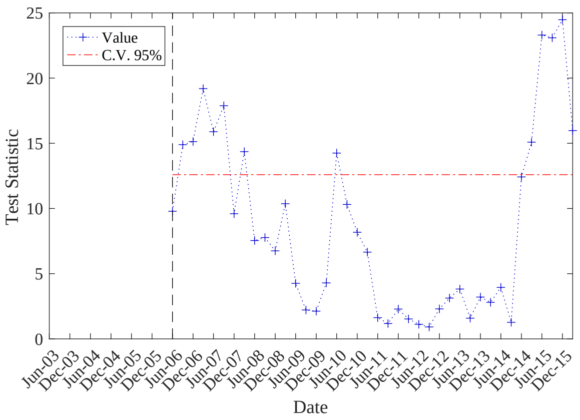


(g) Best 70% selection (NSS)

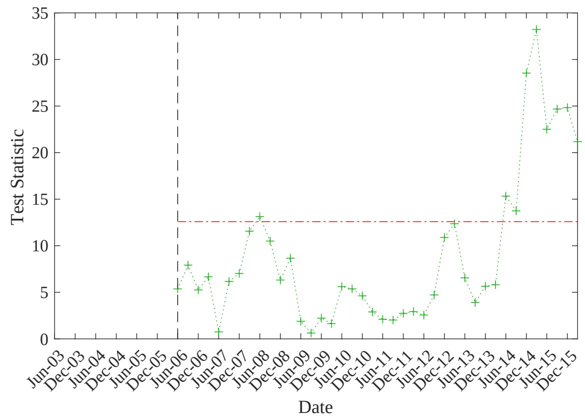


(h) Best 50% selection (NSS)

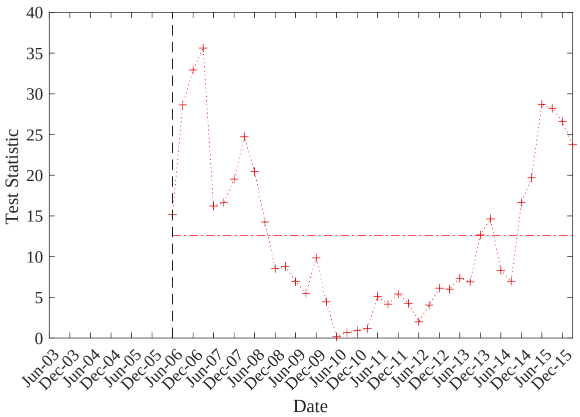
Figure A.1: Results of the efficiency test for the four different levels of screening, with and without short-selling restrictions. When the test value lies above the critical value the null hypothesis is rejected.



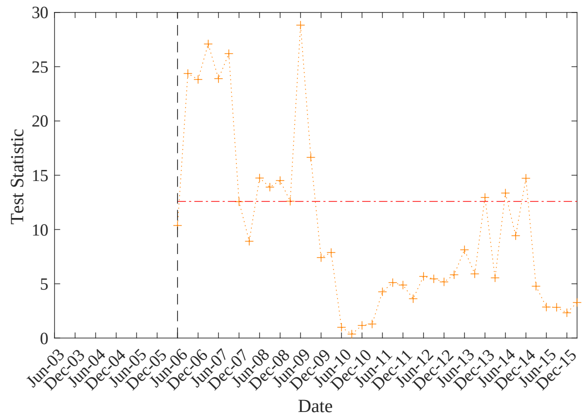
(a) Best 50% selection (ESG)



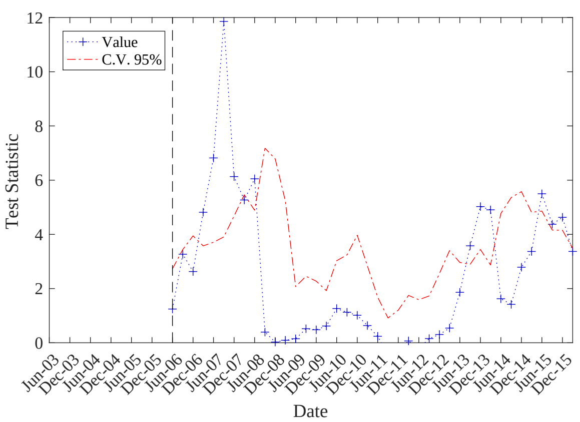
(b) Best 50% selection (ENV)



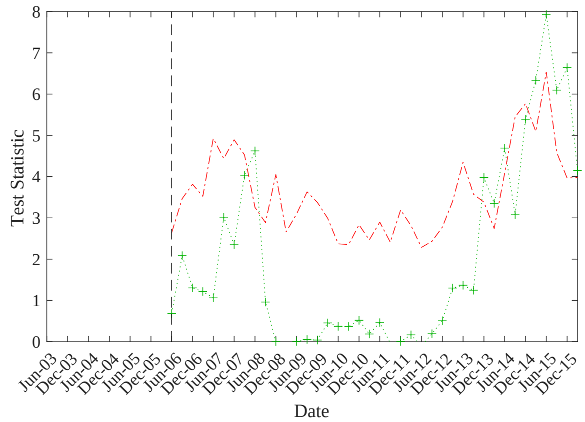
(c) Best 50% selection (SOC)



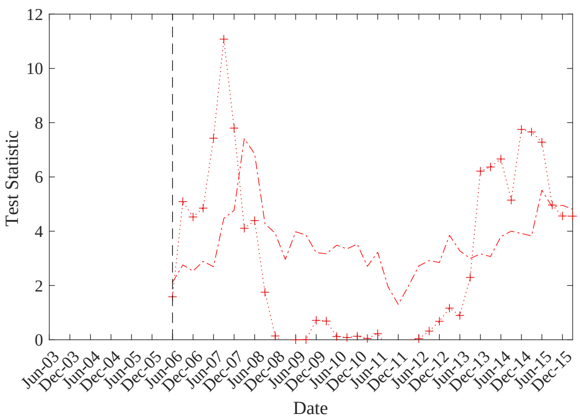
(d) Best 50% selection (GOV)



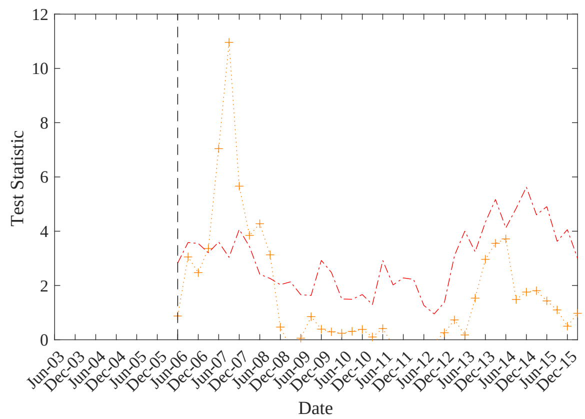
(e) Best 50% selection (ESG,NSS)



(f) Best 50% selection (ENV,NSS)



(g) Best 50% selection (SOC,NSS)



(h) Best 50% selection (GOV,NSS)

Figure A.2: Results of the efficiency test for the highest level of screening (Best 50%) based on the four distinct scores (ESG,ENV,SOC,GOV), with and without short-selling restrictions. When the test value lies above the critical value the null hypothesis is rejected.

Appendix B

MATLAB Code

B.1 Mean-Variance Efficiency Test

```
function [xiGRS1,xiGRS2,p1,p2] = efficiency(R1,R2)
%
%   This Matlab function performs the mean-variance efficiency test
%   as proposed by Jobson(1982) and Gibbons, Ross & Shanken(1989).

%   Input:
%   RB: TxK matrix of excess returns of benchmark assets
%   RN: TxN matrix of excess returns of test assets
%
%   Output:
%   xiGRS1: test statistic when assuming normally distributed returns
%   xiGRS2: asynthotic test statistic
%   p1: p-value for xiGRS1
%   p2: p-value for xiGRS2
%

[T,K] = size(R1);
N = size(R2,2);

X = [ones(T,1) R1];


B = (X'*X)\(X'*R2);
e = (R2-X*B);
omega = (1/T)*((R2-X*B)'*(R2-X*B));
alpha = B(1,:)';
CovB = kron(inv(X'*X),omega);

mu0 = mean(R1)';
sigma0 = cov(R1,1);

% compute test statistics
xiGRS1 = ((T-(N+K))/N)*inv(1+mu0'*inv(sigma0)*mu0)...
    *alpha'*inv(omega)*alpha;
xiGRS2 = T*inv(1+mu0'*inv(sigma0)*mu0)*alpha'*inv(omega)*alpha;

% compute p-values
p1 = 1-fcdf(xiGRS1,N,T-N-K);
p2 = 1-chi2cdf(xiGRS2,N);
```

B.2 Mean-Variance Efficiency Test with Short-Selling Restrictions

```
function [xiNSS,pval,percentiles] = efficiencyNSS(RB,RT,blocksize,Nsim)
%
%   This Matlab function performs the mean-variance efficiency test
%   with restrictions to short sales. Since the test statistic has
%   an unknown distribution, p-value and critical values are computed
%   via simulation adopting the approach from Glen and
%   Jorion(1993).
%   Simulations are pefromed using the stationary bootsrap procedure from
%   Politis and Romano(1994)
%
%   Input:
%   RB: TxK matrix of excess returns of benchmark assets
%   RT: TxN matrix of excess returns of test assets
%   Nsim: number of desired simulations (default = 1000)
%   blocksize: desired block size for the block bootsrap (default = 1)
%
%   Output:
%   xiNSS: test statistic
%   pval: p-value approximated via simulation
%   percentiles: critical values (90%, 95%, 99%) approximated via
%   simulation

%% Initialization %%

% set default values
if nargin<4;
    Nsim = 1000;
    if nargin<3;
        blocksize = 1;
    end;
end;

% set the options for the optimization function
options = optimoptions('fmincon','Algorithm','sqp','Display','none',...
    'MaxIterations',10000,'MaxFunEvals',100000,'OptimalityTolerance',...
    0.0000001,'FunctionTolerance',0.0000001,'StepTolerance',...
    1.0e-7*0.00001);

%% Compute orginal test statistic %%

[T,K] = size(RB);
N = size(RT,2);

R = [RB RT];

% estimate MS portfolios using the Portfolio Object functions
p = Portfolio;
p = setAssetMoments(p,mean(R),cov(R));
p = setDefaultConstraints(p);
pwMS = estimateMaxSharpeRatio(p);
[psMS, prMS] = estimatePortMoments(p,pwMS);
Sh = prMS/psMS; %Maximum Sharpe ratio using test plus benchmark assets
rMS = R*pwMS;

p1 = Portfolio;
p1 = setAssetMoments(p1,mean(RB),cov(RB));
p1 = setDefaultConstraints(p1);
pwMSB = estimateMaxSharpeRatio(p1);
```

```
[psMSB, prMSB] = estimatePortMoments(p1,pwMSB);
ShB = prMSB/psMSB; %Maximum Sharpe ratio using only benchmark assets
rMSB = RB*pwMSB;

% original test statistic
xiNSS = T*(Sh^2-ShB^2)/(1+ShB^2);

%% Compute simulated test statistics %%

% modify observed data in order to satisfy the null hypothesis
X = [ones(T,1) rMSB];
B = (X'*X)\(X'*RT);
alpha = B(1,:);
H0alpha = zeros(size(alpha));
H0B = [H0alpha;B(2:end,:)];
e = (RT-X*B);
H0RT = X*H0B+e;
H0R = [RB H0RT];

% set seed for the random number generator
rng('default');

% generate the Nsim sequences of bootstrapped indices
sequence = [];
for i = 1:Nsim;
    [tempsequence] = sbSequence(T,blocksize);
    sequence = [sequence tempsequence];
end;

% compute simulated test statistics
simxiNSS = zeros(Nsim,1);
parfor i = 1:Nsim;
% If the Parallel Computing Toolbox is installed, the parfor loop
% reduces computation time taking advantage of multi-core CPUs.
% Otherwise, it behaves as a regular for loop.

    % bootstrap returns from modified data
    simR = H0R(sequence(:,i),:);
    simRB = simR(:,1:size(RB,2));
    simmu = mean(simR)';
    simSigma = cov(simR);

    % function to minimize (negative of the Sharpe ratio)
    f = @(w) (-(w*simmu)/sqrt(w*simSigma*w'));
    % set constraints
    lb = zeros(1,size(simR,2));       % lower bound
    ub = ones(1,size(simR,2));        % upper bound
    Amat = ones(1,size(simR,2));      % linear equality contraint
    Bmat = 1;
    simw0 = ones(1,size(simR,2))/size(simR,2); % initial point
    % estimate the MS portfolio for all assets using fmincon
    simpwMS = fmincon(f,simw0,[],[],Amat,Bmat,lb,ub,[],options);
    simSh = (simpwMS*simmu)/sqrt(simpwMS*simSigma*simpwMS');

    simmu = mean(simRB)';
    simSigma = cov(simRB);
    f = @(w) (-(w*simmu)/sqrt(w*simSigma*w'));
    lb = zeros(1,size(simRB,2));       % lower bound
    ub = ones(1,size(simRB,2));        % upper bound
    Amat = ones(1,size(simRB,2));      % linear equality contraint
    Bmat = 1;
    simw0 = ones(1,size(simRB,2))/size(simRB,2); % initial point
    % estimate the MS portfolio for benchmark assets using fmincon
    simpwMSB = fmincon(f,simw0,[],[],Amat,Bmat,lb,ub,[],options);
    simShB = (simpwMSB*simmu)/sqrt(simpwMSB*simSigma*simpwMSB');

    % compute simulated test statistic
```

```
    simxiNSS(i,1) = T*(simSh^2-simShB^2)/(1+simShB^2);
end;

%% Compute outputs %%

% compute p-value and critical values
pval = (sum(simxiNSS>=xiNSS)+1)/(Nsim+1);
percentiles = quantile(simxiNSS,[0.90 0.95 0.99])';
```

Appendix C

Additional Tables and Figures

Table C.1: Industry composition for the different levels of screening (as of June 2003).

Industry Composition						
Industry	Reference	Rated	”Sin” Excl.	Best 90%	Best 70%	Best 50%
Financials	25.3%	22.6%	24.4%	21.9%	19.0%	17.1%
Industrial	19.8%	20.6%	19.7%	20.3%	21.0%	23.6%
Consumer Services	15.6%	15.9%	15.8%	16.3%	16.9%	16.4%
Consumer Goods	13.1%	13.6%	10.8%	10.8%	12.8%	10.0%
Basic Materials	6.7%	7.3%	7.9%	8.4%	8.7%	9.3%
Health Care	6.5%	6.0%	6.5%	6.0%	5.1%	6.4%
Utilities	4.0%	4.3%	4.7%	5.2%	5.6%	5.7%
Technology	3.4%	3.7%	3.9%	4.4%	2.6%	3.6%
Oil & Gas	2.8%	3.0%	3.2%	3.6%	4.1%	4.3%
Telecommunications	2.8%	3.0%	3.2%	3.2%	4.1%	3.6%

Table C.2: Industry composition for the different levels of screening (as of December 2009).

Industry Composition						
Industry	Reference	Rated	”Sin” Excl.	Best 90%	Best 70%	Best 50%
Financials	23.9%	23.5%	26.2%	22.0%	21.5%	21.8%
Industrial	19.9%	20.1%	18.6%	19.9%	20.3%	20.6%
Consumer Services	12.8%	12.9%	10.5%	11.4%	11.2%	10.7%
Consumer Goods	11.5%	11.5%	9.9%	10.1%	10.9%	10.3%
Basic Materials	7.3%	7.2%	8.0%	8.9%	9.4%	9.5%
Health Care	6.4%	6.5%	7.2%	7.3%	4.7%	4.5%
Utilities	4.7%	4.8%	5.4%	5.7%	7.1%	7.8%
Technology	4.4%	4.4%	4.7%	4.6%	4.1%	3.3%
Oil & Gas	5.8%	5.7%	6.4%	6.6%	6.8%	7.0%
Telecommunications	3.3%	3.3%	3.1%	3.4%	4.1%	4.5%

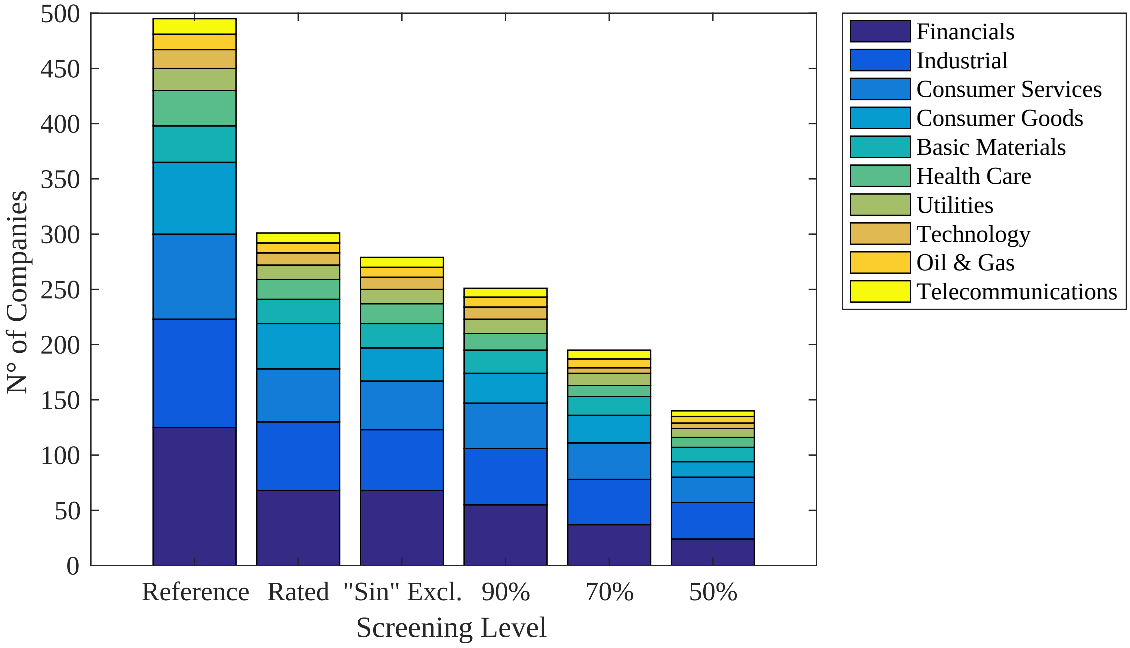


Figure C.1: Number of companies per industry for the different levels of screening (as of June 2003).

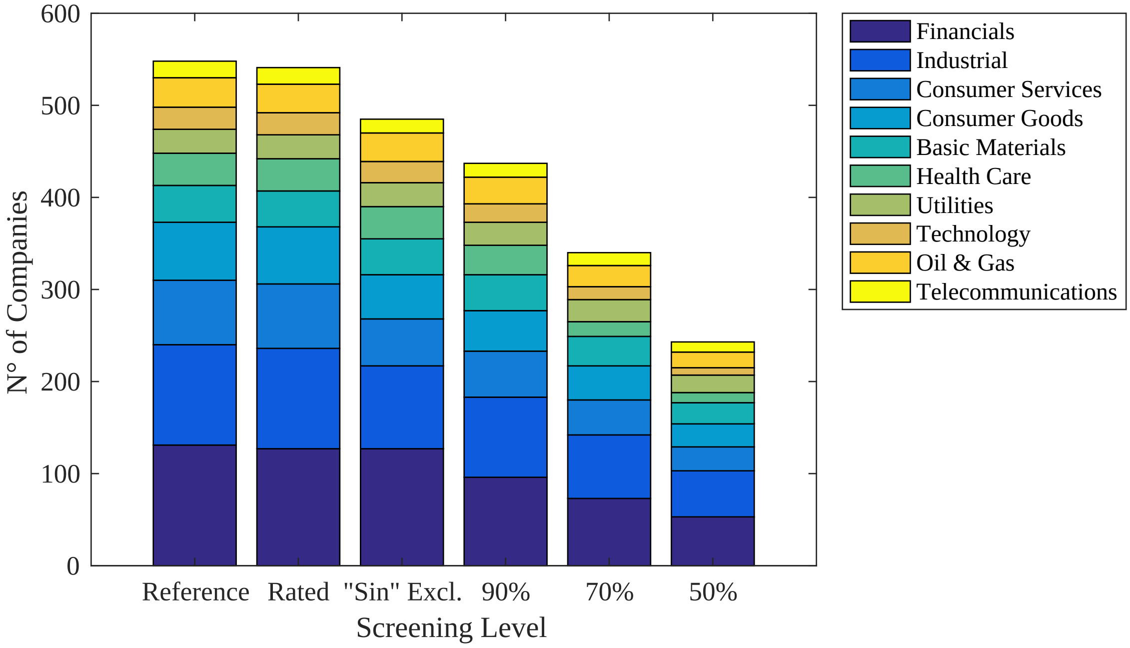
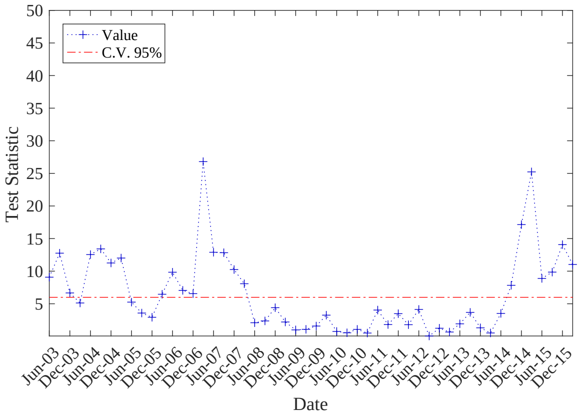
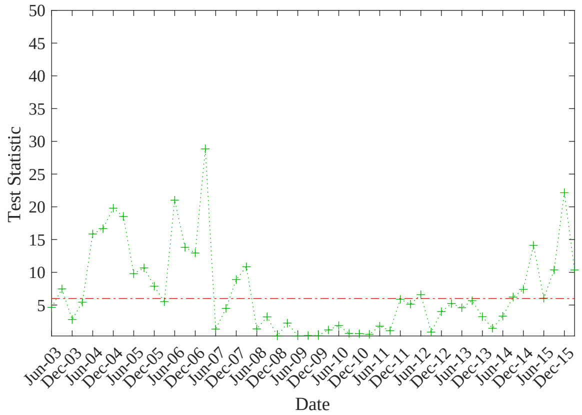


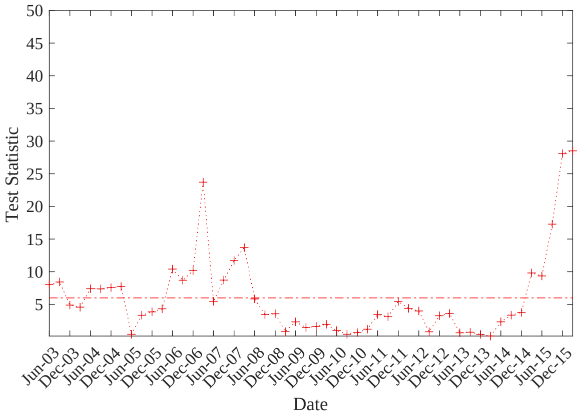
Figure C.2: Number of companies per industry for the different levels of screening (as of December 2009).



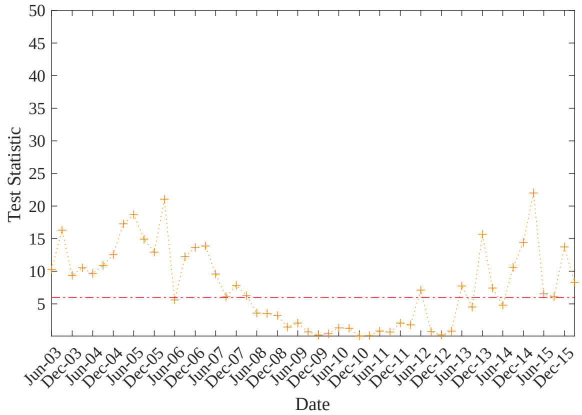
(a) Best 90% selection (ESG)



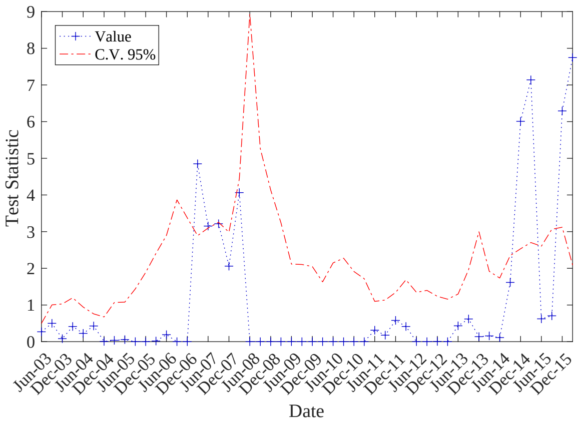
(b) Best 90% selection (ENV)



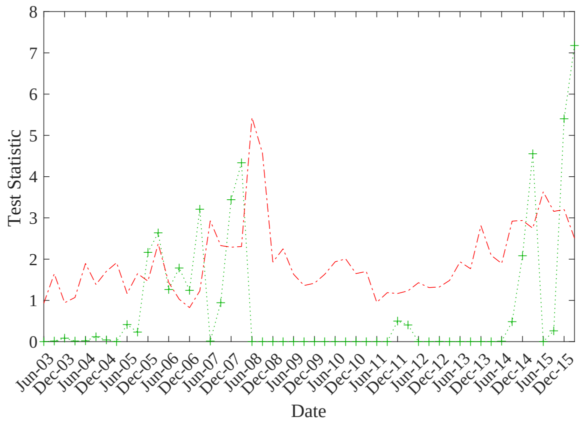
(c) Best 90% selection (SOC)



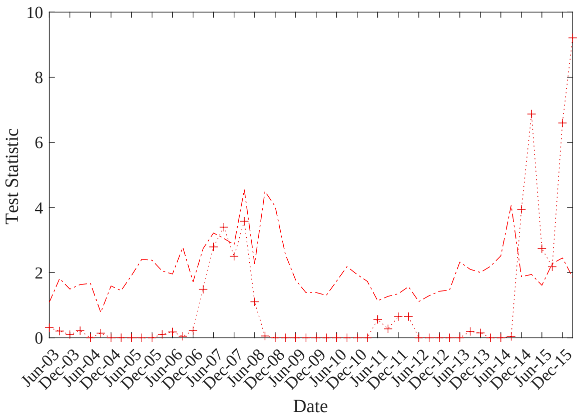
(d) Best 90% selection (GOV)



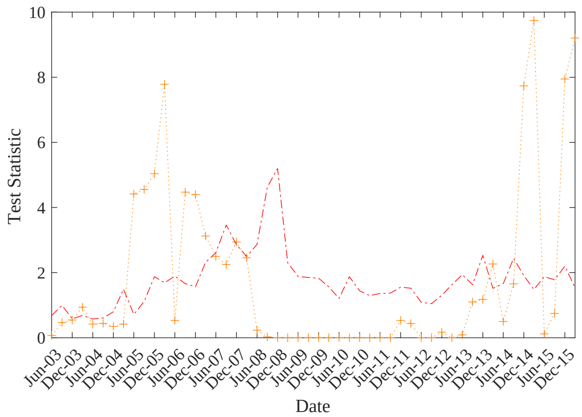
(e) Best 90% selection (ESG,NSS)



(f) Best 90% selection (ENV,NSS)

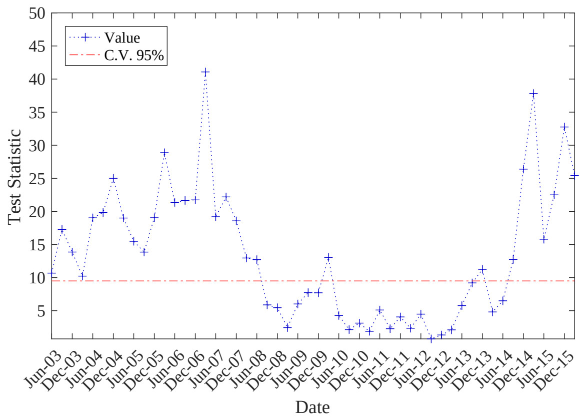


(g) Best 90% selection (SOC,NSS)

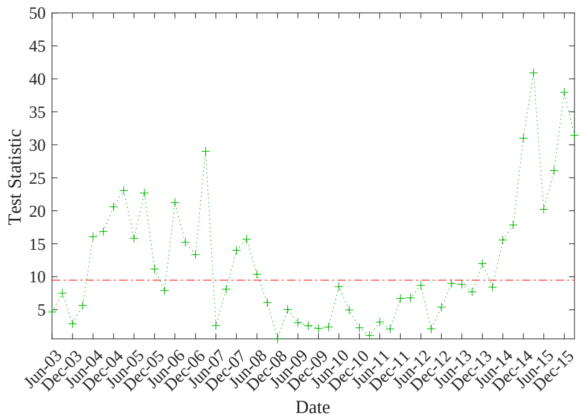


(h) Best 90% selection (GOV,NSS)

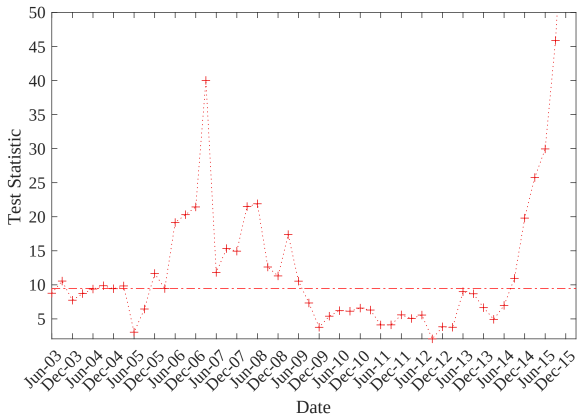
Figure C.3: Results of the efficiency test for the lowest level of screening (Best 90%) based on the four distinct scores (ESG,ENV,SOC,GOV), with and without short-selling restrictions. When the test value lies above the critical value the null hypothesis is rejected.



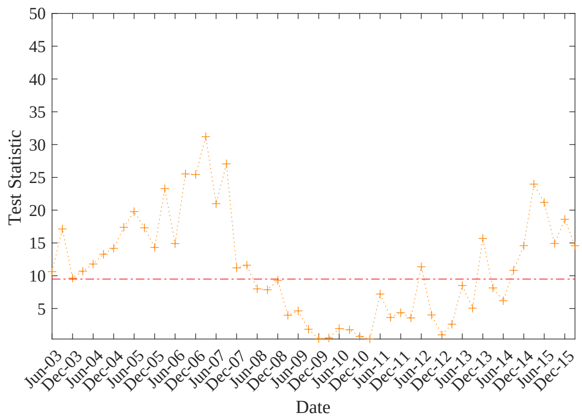
(a) Best 70% selection (ESG)



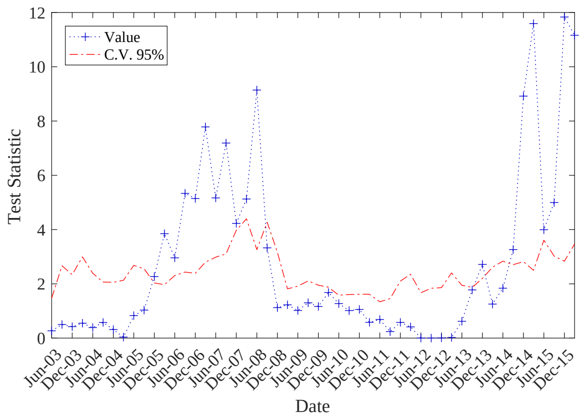
(b) Best 70% selection (ENV)



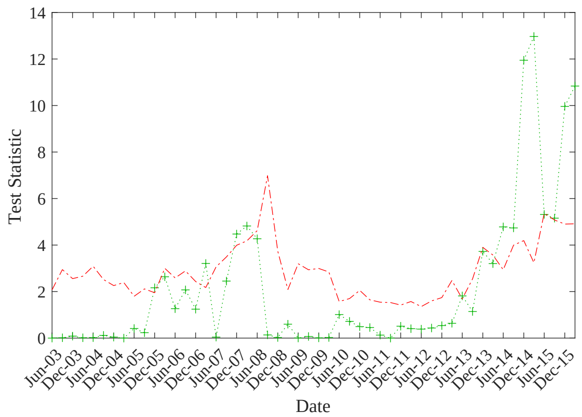
(c) Best 70% selection (SOC)



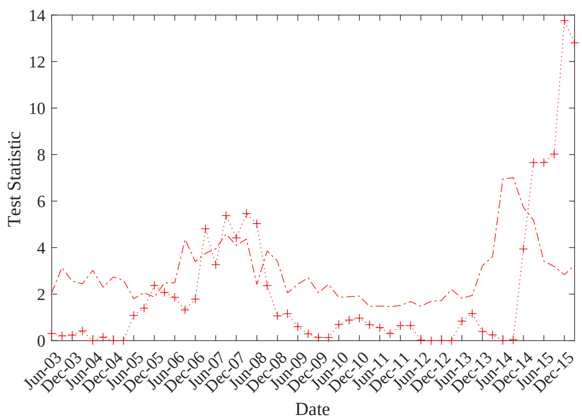
(d) Best 70% selection (GOV)



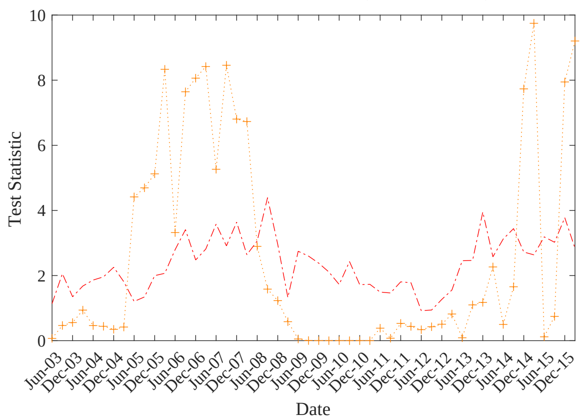
(e) Best 70% selection (ESG,NSS)



(f) Best 70% selection (ENV,NSS)



(g) Best 70% selection (SOC,NSS)



(h) Best 70% selection (GOV,NSS)

Figure C.4: Results of the efficiency test for the intermediate level of screening (Best 70%) based on the four distinct scores (ESG,ENV,SOC,GOV), with and without short-selling restrictions. When the test value lies above the critical value the null hypothesis is rejected.