Economics and Business Administration

MSc in Finance and Strategic Management

Master’s Thesis

The Performance of Socially Responsible Investments in Public Equity

Risk and Return of ethical funds and stocks

Copenhagen Business School, 2018

Authors: Marco Ciancabilla, Lukas Maximilian Sauckel

Supervisor: Francesco Di Lorenzo

Hand-in date: 11/09/2018

Pages: 102 / 158 incl. appendix

Characters (incl. spaces): 174,153

Abstract

Among the trends that constantly shape the behaviour of investors on the financial market, growing attention has been given to the duality of "doing well while doing good". With an overall market estimated to be worth $22.89 trillion, social investments caught the attention of academics and investors, who had been looking closely to the financial performance of SRI (Socially Responsible Investing) strategies for many decades. Notwithstanding the vast literature that has been created – mainly through empirical studies – there is neither a definite position on whether social investments perform differently than conventional investments, nor on what might cause potential differences.

Against this background, the thesis studies the performance of social stocks and funds and compares those to their non-social peers and to the market index, respectively. The analysis develops on two levels. On a firm level, social and non-social stocks are compared in terms of excess returns, through the Augmented Dickey-Fuller test, and Sharpe ratio, through a paired circular block bootstrap. On a fund level, the monthly returns of social funds are regressed over three different timeframes against the market index (identified in the MSCI world), and the presence of a positive or negative Jensen's alpha is investigated.

The results do not show a statistically significant difference between most social and non-social stocks in terms of both excess returns and Sharpe ratios – except for two groups of companies operating in North America where non-responsible stocks gained superior performances. Finally, no significant differences in performance between social funds and the market index are found, with the majority of the alphas not being statistically different from zero. Therefore, we conclude that investors do not need to discount financial performance when embracing SRI strategies.

1. Introduction

1.1 Relevance of SRI

Socially responsible investments have kept growing over the last decade. In 2016, 26% of all professionally managed assets were invested according to responsible investing principles (GSIA, 2017). The idea to consider the environmental and social impact of an investment has been present ever since the existence of investments, but it only played a marginal role until the 1960s – when American anti-war groups required portfolios free from manufactures related to the gun industry. In the following decades, major changes in the way and scope in which ethical considerations are applied occurred. By the beginning of the 21st century, socially responsible investing (SRI) had become substantially more common and the growing attention peeked in the formation of the United Nations' principles of responsible investing (PRI) (Vandekerckhove, et al., 2011, pp. 21-23). Since the PRI association first formulated its principles to incorporate environmental, social and governance (ESG) factors into investment analysis in 2006, the number of signatories has risen from 100 to over 1.800 (PRI Association, 2018a). Between 2014 and 2016, responsible invested assets grew by 25% globally and are now estimated to be worth $22.89 trillion (GSIA, 2017). This rise of SRI has taken place in developed and emerging markets alike and, in 2015, has significantly outpaced the growth of total assets under management (Bloomberg , 2017a).

The drivers of this growth are manifold. For once, investor behaviour had previously been assumed to be 'rational' and investing strategies were built strictly on maximizing risk-adjusted returns. This barrier to SRI has decreased as nowadays utility and profit maximization are not seen as identical, a difference that influences the investment decisions of asset managers – who now take their clients' ethical concerns into consideration (Vandekerckhove, et al., 2011). Other drivers include the emergence of new social and environmental challenges that are increasingly tackled through socially responsible investing, as well as the development of new investing frameworks and methods to measure the ESG performance of a company, organisation or project (O'Toole & Vogel, 2011; Barton, 2011). Nevertheless, the shift of SRI into mainstream investing, specifically the consideration of ESG criteria by more and more conventional investors, is also motivated by a growing body of literature suggesting that ESG engagement positively affects corporate financial performance (Friede, et al., 2015; Revelli & Viviani, 2015; O'Toole & Vogel, 2011).

The world of social investing is wide and branched. SRI stands for a range of investment strategies and principles which consider environmental, social and governance (ESG) factors. Usually terms like ‘responsible-’, ‘ethical-’ or ‘sustainable investing’ are used synonymous, while other terms, e.g. ‘Impact Investing’, often describe different types or subsets of SRI. In contrast to conventional investments, “SRI applies a set of investment screens to select or exclude assets based on ecological, social, corporate governance or ethical criteria, and often engages in the local communities and in shareholder activism to further corporate strategies towards the above aims” (Renneboog, et al., 2008a). This implies that investors not only demand a financial return, but also wish to derive social utility from their investments (Bollen, 2007), and that therefore their expectations differ from that of conventional investors – as is shown by Bialkowski and Starks (2016). Thus, socially responsible investors also pursue financial returns but additionally derive utility from the ESG dimension of the assets they hold. Nonetheless, “by investing in mutual funds rather than in non-profit organizations, SRI investors desire to enhance their financial utility as they expect positive risk-adjusted returns on their investments” (Renneboog, et al., 2008b). In other words, satisfying socially responsible investing principles alone is not a sufficient condition for most SRI investors.

Consequentially, the arising question is whether and how ESG factors are connected to corporate financial performance and whether investors have to sacrifice financial return to gain a positive ESG impact (Halbritter & Dorfleitner, 2015). Understanding this connection is highly relevant, not only for private investors but also for large institutional investors – from pension, endowment and mutual funds to insurance companies – which often feel reluctant to engage in SRI because of their fiduciary duty. Indeed, despite an interest in responsible investments, the principles made in protection of their clients only allow for ESG criteria consideration if it has no detrimental financial effect (Freshfields Bruckhaus Deringer, 2005). Even though a vast body of literature has been produced, studies concerned with the effect of ESG on corporate performance have yielded mixed results that vary across markets, sectors and study types (Friede, et al., 2015).

Therefore, this thesis seeks to contribute to the ongoing research in the field by studying the financial performance of socially responsible, public equity investments. Specifically, we take the side of a private investor willing to pursue social investments, as long as their risk-adjusted returns do not statistically differ from the ones of common stock. For this to be possible, two conditions need to be met. First, that the integration of ESG principles in the company’s culture does not impact its financial performance negatively. Second, that the integration of SRI strategies during the process of portfolio-creation does not penalize an investor with inferior returns. Based on this framework, two question are asked: does ESG influence corporate performance? And can an investor pursue SRI strategies without having to suffer a financial loss?

1.2 Thesis objectives and structure

For asset owners, the most common investment strategy based on ESG criteria is to screen a given investment universe to include or exclude certain stocks depending on their environmental, social or governance performance (US SIF, 2017a). The screens either aim to select investments in sectors, companies or projects for their positive ESG performance (positive screening) or to exclude certain sectors or companies engaged in activities deemed unacceptable or controversial (negative screening) (US SIF, 2017b). On the public equity market, screens can be applied by using ESG ratings on stocks, which are conducted by a variety of providers that also make them available on financial data bases (e.g. Bloomberg, Thomson Reuters, Morning Star).

Differently, among money managers ESG integration is the most common strategy, referring to "the systematic and explicit inclusion by investment managers of ESG factors into financial analysis" (US SIF, 2017a). This strategy does not aim to include or exclude stock per se, but evaluates how the social score might affect the value of the investment and whether it is possible to integrate SRI principles with other investment strategies. In this thesis, we will address both ESG screening and ESG integration.

When assessing the value of impact investing for private investors, it is not sufficient to prove that companies following SRI principles benefit financially. Indeed, as Friede, et al. (2015) point out, while the case for a single company to 'do well while doing good' seems strong, studies on socially responsible investment portfolios and SRI funds have shown that potential ESG benefits at the company level are difficult to be capture by investors. Based on these findings, we split up our analysis in two. The first part focuses on the financial performance of socially responsible stocks when compared to controversial stocks. This resembles a non-portfolio study[1] in the sense of Friede, et al. (2015) and allows us to analyse the impact of a firm's ESG engagement on its stock returns.

[1] SRI portfolio-studies analyse the performance of long-short ESG portfolios and mainly focus on fund performance (Friede, et al., 2015)

Since this approach only focuses on screening to select stocks, we conduct the second analysis on socially responsible funds – that integrate social criteria with other investing strategies. As those are actively managed and classified as either "social" or "ethical" funds, we assume ESG integration to be a substantial part of their portfolio creation strategy. The rational for the second analysis is based on the argument that even if ethical corporate behaviour benefits a firm's financial performance, investors might not be able to exploit this advantage due to lack of diversification, management fees and other trading costs (Friede, et al., 2015).

In the following section (2), we provide a review of the relevant literature and derive two research questions on which the afore mentioned research objectives are based. Next, the data we collected as well as the methodology of our analyses is presented in two parts, one for each analysis (section 3). We then study socially responsible investment performance on the stock level in section 4 and display our results. Subsequently, in our second analyses we study SRI funds and show the findings in addition. For both analyses, we provide a thorough description of the statistical tools we have used to test our results. Afterwards we discuss limitations to our findings before we finally derive the conclusion in section 6.

2. Literature Review

2.1 SRI and corporate financial performance

Why do companies invest in sustainability? One can distinguish between three opposing views on this issue. The first, suggested by Bénabou and Tirole (2010), argues that socially responsible companies serve as an efficient channel to express the personal values of their share- and stakeholders, who sponsor the ESG engagement of the company (‘delegated philanthropy’). Another point of view is that firms engage in sustainability based on the initiative of managers who desire to capture private benefits, employing resources that could be used in more profitable ways, resembling a type of ‘corporate philanthropy’ (Brammer & Millington, 2008; Bénabou & Tirole, 2010; Khan, et al., 2017). Support for this argument also comes from Jensen (2002), who points out that increasing stakeholder value in contrast to shareholder value is expanding the power of managers in an unproductive way according to principle-agent theory. What the two views mentioned have in common is that they do not require a positive influence of ESG performance on corporate financial performance to explain the existence of SRI, because investors and other stakeholders are ultimately willing to sacrifice risk-adjusted returns for their ESG cause. In fact, it seems plausible that firms using resources for ESG related issues have a competitive disadvantage, since they are restricted in their behaviour towards business opportunities or because some resources could be used in other, more lucrative ways (Mcwilliams & Siegel, 1997; Hillman & Keim, 2001; Brammer & Millington, 2008).

The third view on SRI is that ESG investments lead to measurable improvements in many areas of a company and thus have a positive impact on its stock returns. In this view, socially responsible behaviour leads to higher revenues and lower costs (Eccles, et al., 2014), as it happens with the Walmart case, where the multi-national retail company experienced substantial cost savings after they made of environmental sustainability a strategic goal (O'Toole & Vogel, 2011). Such an influence of ESG measures on operational metrics seems to be the norm rather than the exception. According to an equity research conducted by Nordea Bank, firms with top ESG ratings have, amongst others, a higher ROE and lower net-debt/EBITDA than the market – with more stable returns, margins and share prices (Porse, et al., 2017). Eccles et al. (2014) find that firms with high ESG ratings outperform low rated firms in terms of stock and accounting performance. Similarly, others provide evidence that successful implementation of ESG measures improves the return on assets, profit margin, asset turnover, and sales over employees' ratios significantly (Dimson, et al., 2015). There is also reason to believe that engaging in ESG activities enhances firm reputation (Fombrun & Shanley, 1990; Fombrun, 2005; Freeman, et al., 2007). Results by Turban and Greening (1997) indicate that – on the grounds of reputation – socially responsible firms gain a competitive advantage by attracting more and better employees.

Another common argument for the positive effect of SRI on financial performance is that investing in ESG issues reduces the probability of regulatory, legislative and fiscal sanctions (Coleman, 2011; Kumar, et al., 2016). However, the connection between ESG practices and regulatory risk remains highly controversial (Flood, 2018). The petrol company BP is an example that high ESG standards do not necessarily safeguard against environmental and social disasters. Despite being hailed for its commitment to climate change (including the change of its name from British Petrol to Beyond Petrol, investing in solar power and reducing greenhouse-gas emissions by 10% within 11 years), BP was struck by three immensely value-destroying incidents within just 5 years; inadequate maintenance resulted in 15 deaths in Texas and several oil spills in Alaska, only to be exceeded in gravity by the Deepwater Horizon drilling platform catastrophe in 2010 (O'Toole & Vogel, 2011). Although counterintuitive, a study conducted by the investment company BlackRock suggests that BP is not an exceptional case; results show that firms adhering to environmental, social and governance policies are more likely to face lawsuits and regulatory action (Garvey, et al., 2016). One possible reason is that ESG ratings rely heavily upon a company's ESG disclosure and historic ESG performance, which might dilute true corporate social performance (Lee & Moscardi, 2018). On the other hand, the Diesel gate affair at VW in 2016 is an example where ESG scores seem to reflect or even predict the risk of a company's business quite accurate, since VW ranked lower than 72% of the firms in MSCI's governance ratings prior to the software tempering revelation. According to Howard Sherman, head of corporate governance business development at MSCI, investors should have been more concerned about potential risks due to the low governance performance that was deteriorating since 2014 (Grene & Marriage, 2015).

Looking at stock returns, several studies conclude that the risk of high-ESG rated companies is lower than for low rated companies, usually measured by volatility or beta (Porse, et al., 2017; Kumar, et al., 2016; Verheyden, et al., 2016; Chaudhry, et al., 2016). Further, a recent report from the PRI association has shown a negative correlation between companies' ESG score and the volatility of their returns, conjecturing the lower risk of these companies with the higher quality and stability associated with their social commitment (PRI, 2016). Finally, another interesting result concerning the nature of SRI is given by Glode (Glode, 2011) and Kosowski (Kosowski, 2011) who studied the performances of SRI funds during crisis periods and concluded that those funds performed better during crisis than during times of market stability – and by Nofsinger and Varma (Nofsinger & Varma, 2014), that showed how social funds outperform conventional peers during market crisis.

Despite mixed results of studies investigating the relation between ESG engagement and firm performance, it is argued for the potential positive effects on materiality, that ESG data should be part of the financial analysis investors conduct. The PRI[2] association for instance believes that "ESG issues can affect the performance of investment portfolios (to varying degrees across companies, sectors, regions, asset classes and through time)" and that material ESG factors should be included in a risk-return analysis, even before or without considering the social and environmental impact of an investment (PRI Association, 2018b). Several researchers argue, that one needs to distinguish between ESG activities that influence materiality[3] or affect stakeholder value and immaterial ESG activities. A thoroughly conducted study on the effect of ESG factors on materiality is given by Khan et al. (2017), who examine the future performance effects of material versus immaterial sustainability investments. The authors find that the highest performing firms have both good ratings on material sustainability issues and low ratings on immaterial sustainability issues. They conclude that "collectively these results are consistent with materiality guidance being helpful in improving the informativeness of ESG data for investors". Similarly, Hillman and Keim (2001) also analyse different types of ESG engagement and distinguish between activities that directly benefit stakeholders and social investments which go beyond that. Their findings support that the former do lead to shareholder wealth creation while the latter do not.

Even if granting the positive effect on materiality and shareholder wealth, investors are sceptical of the financial performance of socially responsible investments for several other reasons. One of the greatest concerns is that SRI strategies limit the stock universe available to investors, thereby generating 'diversification costs' (Girard, et al., 2007). Based on the modern portfolio theory

[2] The PRI association promotes the United Nations Principles of Responsible Investing framework and supports its signatories regarding SRI engagement (PRI Association, 2018b)
[3] Material information is defined as presenting a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the total mix of information made available (U.S. Supreme Court 1976)

established by Markowitz (1952), the option to reduce portfolio risk by selecting investments whose returns have a low correlation is thereby diminished. The theory also suggests (meanwhile strengthened by empirical research) that the risk-adjusted returns of a portfolio are mainly determined by the diversification effect of asset allocation instead of superior stock selection (Roselle, 2016; Bollen, 2007). However, as Bollen (2007) points out, the portfolio constraints must prove to be binding on performance – otherwise they could also function as screens for management quality and generate superior risk-adjusted returns. Thus, despite the concern that social screens may constrain portfolio optimization, some argue that "investors frequently choose to limit their investment universes by favouring certain asset classes, industries or companies, and that this process is a fundamental part of investment practice" (Caplan, et al., 2013). Furthermore, the great liquidity, efficiency and size of stock markets might make the diversification effect negligible (Diltz, 1995). Under the assumption of incomplete information, it is also possible that socially responsible investors maximize returns with a portfolio that appears to be under-diversified in the sense of Markowitz (1952) (Revelli & Viviani, 2015). Assuming asymmetric information among investors, which implies that a fully diversified portfolio is not efficient anymore, it would be reasonable that some investors are better in gathering or analysing ESG data than others. In fact, investors and portfolio managers have experienced difficulties in doing so due to the lack of standardised, consistent and reliable frameworks to measure ESG performance (Bloomberg , 2017b).

This touches upon another cause for scepticism among investors, based on the efficient market hypothesis. It argues that all relevant, publicly-available information, including ESG data, is priced by investors and reflected in the share price of a company (Fama, 1970). ESG proponents argue, that analysing the ESG characteristics of a firm is not part of the mainstream investor considerations and therefore represents an opportunity to generate returns from mispriced stocks (Caplan, et al., 2013). Although this might have been an explanation in the past (Hamilton, et al., 1993), it is now understood that significantly more investors incorporate ESG information when evaluating investments (GSIA, 2017; US SIF, 2017a).

Additionally, various providers of company ratings, specialised on social and sustainability issues, have made it easier for investors to examine the ESG dimension of a firm. From early frameworks like RobecoSAM's ESG analysis, used for the DowJones Sustainability Index (DJSI), to more recent ones (e.g. Thomson Reuters/Asset4, Sustainalytics, MSCI, RepRisk), increasingly more institutional investors, asset managers, financial institutions and other stakeholders rely on these organizations to evaluate a firm's ESG involvement (Davis Polk & Wardwell LLP, 2017). Most of the ESG rating providers use public data as well as input from companies to create scores or similar measures (Davis Polk & Wardwell LLP, 2017). Numerous studies, as well as this one, use these scores as proxies to create groups of 'good' and 'bad' companies, which can then be compared to analyse the social responsibility/corporate performance relation. If ESG ratings are an adequate proxy of the ESG involvement of a firm, and ESG activity is positively correlated with corporate performance (but not priced correctly by the market), socially responsible investors might be able to gain abnormal returns by investing in these companies. For instance, Porse et al. (2017) conclude that "changes in MSCI ESG ratings matter and [we] argue that the underlying change in ESG performance drives returns. These benefits should yield a justifiable valuation premium, which we also see in the market". Turning to empirical evidence to strengthen this proposition, many studies promote the impact of ESG engagement on materiality and many promote SRI as an efficient strategy to create positive alphas. However, the causality is far from clear. Overall, the literature has produced mixed results about the connection between ESG and financial performance, especially from an investor perspective.

The most comprehensive overview of the research investigating the connection between ESG and corporate performance is given by Friede et al. (2015). The authors use a vote-count approach and meta-analysis on 2.200 studies to test whether there is a business case for SRI and come to the conclusion that on average empirical evidence does promote the argument for SRI. The analysis shows that 90% of the studies report a non-negative relation and most of them find ESG criteria positively correlated to corporate performance. Another crucial insight derives from the distinction between portfolio studies, i.e. studies analysing long-short ESG portfolios and focusing on mutual funds, and non-portfolio studies. While the literature on ESG provides a congruent and distinct positive picture for corporations, studies which analyse the relation from an investor perspective yield more mixed results, with a significantly higher number reporting a negative correlation between responsible investments and portfolio performance. According to the authors, the cause of this discrepancy lies in the difficulties investors face when trying to realize the ESG surplus in a portfolio of investments. Possible reasons are that overlapping factors and combinations of positive/negative screening lead to potential ESG-based effects cancelling each other out and that additional investment costs (management fees, trading costs) make the ESG surplus disappear. Although the authors conclude that an average-skilled investor is unlikely to exploit the positive ESG-corporate financial performance relation, they state that “at the worst case, investors in ESG mutual funds can expect to lose nothing compared to conventional fund investments” (Friede, et al., 2015).

2.2 Research questions

To sum up, there is a great amount of literature supporting the theory that a firm can do well by doing good, and scholars largely agree that socially responsible engagement can enhance corporate financial performance. Further, many academics sustain the argument that a company can improve operational metrics by investing in ESG and hence reduce costs or increase revenue. However, when we move the focus from a firm to a portfolio or fund level, the number of results showing a negative relation between ESG and financial performance increases significantly. Trading costs, diversification costs, and mixing up potentially positive with negative ESG effects inside one portfolio could pose a barrier for investors striving to gain superior returns.

We want to contribute to this field of research by studying the financial performance of ESG-screened investments. Specifically, we are interested in two issues connected to SRI: the impact of corporate social performance on stock returns and the risk-adjusted return of socially responsible funds relative to the market. We therefore derive two research questions, one for each academic topic. In the first step, we group stocks by their ESG score and compare the top-ESG firms with their bottom peers. The aim is to investigate whether firms with solid ESG principles and corporations that refrain in large parts from ESG related activities differ in the nature of their absolute and risk-adjusted returns:

RQ1) *How do socially responsible stocks perform when compared to non-socially responsible ones?*

However, based on previous research, even if companies with high ESG scores perform as good as companies with low ESG scores, this would not be sufficient to assert that the social market is as competitive as the pure financial market. Therefore, with the second research question we aim to understand whether the diversification issues arising from SRI strategies can be offset by a higher quality of the selected companies. If confirmed, this would indicate that actively managed funds are able to overcome diversification and information constraints and investors are able to get the benefits of social investments without losing financial return. To test this hypothesis, we will analyse the competitiveness of social funds by using a market index as benchmark. Finally, we will test whether specific conditions, as market recessions or geographic location, influence the financial performance of SRI funds. As different investors have different utility curves based on their preferred trade-off between social and financial return, we phrase the research question so that it focuses on the threshold relevant for most socially responsible investors; that social investments at least yield risk-adjusted returns similarly to the overall market:

RQ2) *Does an investor have to discount financial performance when investing in socially responsible funds?*

The way in which the analysis is constructed will give us further details concerning the quality and the nature of SRI performance. First of all, by comparing social funds to a market index (and not to their non-social peers), we are able to include trading costs in the equation. Indeed, an extended inefficiency of the actively-traded market (e.g. due to the expansion of online brokerage) could undermine the profitability of social funds but would not be shown in a study that compares different subtypes of actively traded funds.

Additionally, running the stock and fund analysis over the same time-frame allow us to compare the results of the two – and therefore to potentially derive some insights on performance differences based on management costs, portfolio selection constraints and on ESG screening vs. ESG integration. Even though both of those questions have been already studied (separately) in literature, the performance of SRI has rarely been analysed over a such long time-frame. We assume longer observation periods might provide further insight as SRI promotes a long-term investing approach. Moreover, since the bulk of the literature was produced, ESG data on public equity has not only become available for a longer period of time, but the number of rated companies has increased as well. To conclude, we will analyse the performance over 12 years of monthly returns on more than 5.000 stocks and 300 funds and we will analyse how both single firms and socially responsible investment portfolios performed over the same period of time.

3. Methodology and Data

3.1 Stock analysis

To answer the first research question, we group stocks of publicly traded companies by their social and sustainability impact, measured by ESG score. Instead of regressing the social companies on a market index, we decided to directly compare the absolute and risk-adjusted returns of social companies with those of non-social ones. In fact, in this first analysis we are not interested in understanding whether social investments are actually profitable on the financial market, but we aim to study whether there are statistical differences in the nature (e.g. average, volatility) of the returns of social and non-social companies. To approach this question we collected data entailing, among others, information about the financial performance and ESG scores of more than 5.000 publicly traded stocks.

3.1.1 *Data*

ESG score

We create the stock universe for the first analysis by obtaining monthly ESG scores from the Thomson Reuters database for the period of December 2006 to February 2018. The score can take a number between 0 and 100 (the higher, the better) and is provided both as a combined overall score or as single score in one of the 400 ESG metrics, each focused on a specific aspect of social impact, grouped in 10 categories (figure 3.1). The data is retrieved from annual reports, company websites, NGO websites, Stock exchange fillings, CSR reports and news sources (Thomson Reuters, 2018). The whole database consists of more than 6.500 globally distributed public companies, which form the basis for our stock universe.

Figure 3.1

The score is calculated using the following formula:

$$Score = \frac{N.companies\ with\ worst\ value + \frac{N.companies\ with\ same\ value}{2}}{N.companies\ with\ a\ value}$$

Being calculated in this way, score results are based on a percentile rank, and therefore are not very sensitive to outliers. Additionally, the normalized weights are calculated excluding quantitative indicators, for which no data is available in the public domain, as it would be highly inaccurate to assign a default value. Finally, different categories have different weights when combined to calculate the ESG overall score. The category weights are determined by the number of indicators that make up each category in comparison to all indicators used in the ESG Score framework. The following table (Table 3.1) summarizes the weight of each category.

Table 3.1

Pillar	*Category*	*Indicators in Rating*	*Weights*
Environmental	Resource Use	20	11%
	Emissions	22	12%
	Innovation	19	11%
Social	Workforce	29	16%
	Human Rights	8	4,50%
	Community	14	8%
	Product Responsability	12	7%
Governance	Management	34	19%
	Shareholders	12	7%
	CSR Strategy	8	4,50%
Total		178	100%

While the Thomson Eikon platform offers ESG ratings on a small number of stocks dating as far back as 2002, we choose a shorter period that we regard as sufficiently long for the analysis. This in order to limit the vast amount of data and to ensure the scope of this thesis is met. Additionally, for the years prior to 2006, the number of stocks considered in the ESG score rating by Thomson is comparatively low. This is a further reason for us to limit the time frame, as we want to avoid the sample size becoming too small. Thus, the time-frame we will use for our analysis reaches from December 2006 to February 2018. Since in the database the number of companies rated by ESG performance increases each month between 2006 and 2018, the number of firms in our population is significantly lower in the beginning (1.716) compared to the end of the period (6.047). Thus, the number of stocks composed in the groups that we create, as described in the methodology section, will also change when a new company is introduced in the universe.

Performance data

The return data is extracted from Thomson as well and is labelled ‘total return’ on the platform. It is calculated as the percentage increase in share price from the last trading day of the prior month to the last trading day of the given month, taking re-invested dividends into account. In order to calculate our portfolio performance measure, the Sharpe ratio, we have to transform the total returns to excess returns, by subtracting the risk-free rate from the stock returns.

We obtain the risk-free rate by downloading 1-month T-bill rates from the Kenneth French website (French, 2018). In addition to total returns, we also retrieve price-to-book ratios and the market capitalisation for each firm from the Thomson Eikon application, enabling us to check if the creation of groups by ESG scores is not merely reflecting other already existing investment screens. Since it is more practical to analyse, we calculate book-to-market ratios by dividing 1 by the price-to-book ratio we collected. To test for distinctness of the ESG portfolio selection as described later, we also calculate 12-months rolling standard deviations of each stock's returns. Several research papers show that the financial performance of socially responsible investments differs between regions (Friede, et al., 2015). As we want to analyse geographical differences as well, we also retrieve the continent of a firm's headquarter from Thomson Reuters. While it is not necessarily the case that a company headquarter shares the same location as the main targeted market, we are not aware of a better approximation of a firm's geographical focus available to us on Thomson Eikon.

Finally, we exclude from the analysis all those companies which would belong to a 'top' or 'bottom' group in a given month for which we cannot retrieve all the necessary financial data. Thus, all the firms that – in a given month – have only one between ESG score and financial performance are excluded. The resulting, cleaned universe covers 5.674 company stocks in total. On average, the stocks have a monthly excess return of 1,12% and an ESG score of 51 (Table 3.2). We filter the Thomson Reuters database for both active and inactive companies, ensuring the analysis does not suffer from survivorship bias (around 40 stocks became inactive in our analysed period).

Table 3.2

Total number of firms	5.674
Number firms mean	3.352
ESG Score mean	50.68
Excess return mean	1.12%
Book-to-market mean	0.76
Market Cap mean	85.823 MM$
Volatility means	1.58%

3.1.2 *Methodology*

Following the creation of the stock universe, we build stock groups (‘portfolios’) and subsequently compare risk and return of the groups that score well with the groups that engage little in ESG to detect potential differences in financial performance. The analysis of the differences between firms with high and low ESG scores can perform as a first quality test for the two most widely practiced socially responsible strategies among investors, namely positive and negative screening – as it compares the companies singled out by negative screening with the ones grouped by positive screening. In fact, according to the US SIF (the Forum for Sustainable and Responsible Investment) out of 75 asset owners, the majority of them named negative screening as the most common socially responsible investing strategy, accounting for 86% of their ESG incorporation assets (US SIF, 2017). The wide use of these screening techniques makes it necessary to understand how well the firms that are categorically excluded from SRI perform when benchmarked to the ones that are included.

Finally, the comparison of top and bottom portfolios becomes more relevant in the light of the findings of Derwall et al. (2011), which show that both high and low ESG-rated portfolios are able to produce superior returns, but for dissimilar reasons. Derwall et al. (2011) conclude that the heterogeneous results are due to two mutually exclusive investment motivations. One is taken by value-seeking investors who are interested in the ESG surplus of their investment, the other by profit-seeking investors who only focus on the maximization of returns. The authors’ findings support the two hypotheses that 1) controversial stocks yield abnormal returns because they are under-priced due to the value-seeking investors who avoid them and 2) socially responsible stocks can yield superior performance because the underlying cash flows are positively affected by ESG factors, which is not correctly priced by the market. Additionally, as we also know from previous research, ESG factors are not only influencing total stock returns but risk as well. We therefore find it reasonable to directly compare risk and return of firms with high ESG scores to their controversial equivalents on the other end of the ranking.

The rational of our method follows other papers studying the socially responsible/corporate performance relation; Verheyden, et al. (2016) for instance create groups by excluding firms performing poorly in ESG terms. Derwall, et al. (2011) use a combination of negative and positive screens to group stocks, but then apply a regression model to analyse the financial performance of the portfolios; Trinks and Scholtens (2017) rely on a regression model as well for a negative-screened company universe. Usually, multi-factor models are applied to explain risk-adjusted excess returns, and then the intercept or alpha of the model is interpreted as superior, above market returns. Although risk-adjusted measures like the Jensen alpha (1968) are useful in assessing the financial performance of well-diversified portfolios, they are less suited for the stock groups of this analysis.

There are two main reasons behind the decision of not using a regression for the stock analysis. First of all, the purpose of this analysis is not to compare the groups to a market index or to determine their overall performance, but to study the differences in (absolute and risk adjusted) returns between firms that have a high degree of social impact and firms that do not have. Furthermore, the way in which the groups have been clustered results in a poor diversification in the sense of modern portfolio theory, therefore exposing the groups to the idiosyncratic risk of the underlying investments – a risk that could be diversified by combining SRI principles with other financial strategies. Jensen's alpha on the other hand is primarily concerned with the market risk which a portfolio is exposed to (Auer & Schuhmacher, 2016; Eling & Schuhmacher, 2007). Thus, we choose not to apply a multi-factor regression model for our analysis.

In our view, a more appropriate alternative to measure the risk-adjusted performance of our stock groups is the Sharpe ratio, which is based on the total risk of the underlying returns (i.e. idiosyncratic and market risk). Sauer (1997) argues, that for socially responsible investors it is more adequate, because "the Sharpe performance index represents the average risk premium per unit of total risk and, therefore, represents a more relevant risk-adjusted measure of performance for investors that are less than well diversified". Equations below show the computation of the Sharpe ratio (Sharpe, 1994).

Equation 3.1, 3.2, 3.3

$$Sharpe\ Ratio = \frac{\overline{D}}{\sigma_D} \qquad \overline{D} = \frac{1}{T}\sum_{t=1}^{T}(R_{Pt} - R_{rf}) \qquad \sigma_D = \sqrt{\frac{\sum_{t=1}^{T}(D_t - \overline{D})^2}{T-1}}$$

The groups are formed by ranking the stocks in our company universe by ESG score and selecting a percentage of the best and the worst firms, i.e. the companies with the highest and lowest ESG scores, with equal weights for the stocks. The percentiles for the top and bottom portfolios are 5%, 10% and 15%. The portfolio constellation is updated monthly if necessary due to a change in the ranking. The way in which we create the universe and the groups implies the number of firms in the portfolios changes when rated firms drop out of the public equity market or when companies are added to the ESG score rating by Thomson Reuters. As more and more companies are rated by Thomson over time, the number of firms in the portfolios consequently increases as well, while only the percentage of top/bottom firms stays constant every month.

In addition to our main analysis objective, we are also interested in detecting whether the difference in the performance between social and non-social companies are stable across markets or change depending on the geographical area in which the firm operates. Thus, the same procedure of creating the groups is repeated for stocks whose headquarters are located in Asia, Europe and North America, leading to a total sample of 12 top and 12 bottom portfolios (3 percentiles, 4 regional focuses). Whilst stocks from Africa, South America and Oceania are included in our global portfolios, their number is too small to justify the creation of additional regions and thus they are not considered in the groups constructed by geography.

Finally, we want also to make sure that our ESG screens do not lead to similar groups of stocks as other existing portfolio selection strategies would do, e.g. Fama and French factor-based approaches (French, 2018). For this reason, we follow Bali et al. (2011) and Auer and Schuhmacher (2016) and calculate correlations between the ESG score, the Fama-French factors (book-to-market ratio and market capitalization) and the volatility. These factors are often used for investment strategies based on empirical observations, namely Small Minus Big (SMB), High Minus Low (HML) and volatility-based strategies. SMB tries to exploit the fact that small stocks perform better relative to big stocks, while HML derives from findings that stocks with high book-to-market ratios seem to outperform stocks with low book-to-market ratios (French, 2018). Prior research also indicates a ‘low volatility’ anomaly, referring to low-volatility stocks outperforming high-volatility stocks (Dutt & Humphery-Jenner, 2013).

3.1.3 *Statistical tests*

The calculation of certain statistics from a given sample, like excess return or the Sharpe ratio, does not tell us how the distribution of the statistic in the underlying population looks like. To generalize the results obtained by the methods described in the previous chapter – and to compare top and bottom groups – we need to make use of inferential statistical tests. These tests study the probability distribution of the excess returns and Sharpe ratios of the different groups and allow us to analyse, at a certain confidence level, whether there are significant differences between the returns of firms with high and low ESG scores. This way, we are able to understand if the differences observed on the sample level reflect a dominance of one population on the other or if they are merely the result of a casual event. This section will be divided as follows: in the first part, we will study the stochastic dominance of one population on another, with the purpose of understanding whether one group gains higher returns than the other one. We will approach this problem first by discussing advantages and limitations of the different available statistical tests and then by analysing the working process of the test selected as the fittest for our needs. The second part will move its focus on the comparison of the group's Sharpe ratio. Again, we will first discuss which frameworks should be applied and then we will elaborate on the framework selected as the fittest.

3.1.3.1 Statistical tests on Excess Returns

The most widely used statistical test to compare two stocks' returns is the ***t-test*** that – by using the t-distribution and the sample size – tests the null hypothesis that the two samples have equal means. This statistical tool compares the samples distribution with the t-distribution and calculates the probability that the two population have equal means. The null hypothesis states that the difference between the observed values of two samples is not statistically different from zero (Senn, 2008). The t-distribution (i.e. the distribution to which the statistics are compared to) results from the estimation of the mean of a normally distributed population and is symmetric and bell-shaped like a normal distribution – but generally with heavier tails (Hurst & Simon, 2010). The use of the t-distribution as a benchmark for the test implies the assumption of a normal distribution of the populations underlying the samples. However, according to the literature, stock returns follow more a random walk than a normal distribution, making the assumption problematic (Sheikh & Hongtao, 2003). For this reason, we decide to use a non-parametric test to check whether the returns of one group stochastically dominate the other.

A non-parametric alternative to the t-test is given by the *Mann-Whitney* test, checking the null hypothesis that, given two populations, it is equally likely that a randomly selected value from one population will be less than or greater than a randomly selected value from the other population. Here the null hypothesis states that true location shift is equal to zero – i.e. that there is not a population constantly greater than the other (Hettmansperger. & McKean, 1998). Differently from the t-test, the Mann-Whitney method does not require an assumption of a normal distribution of the underlying populations and is still almost as efficient as the t-test in case of a normal distribution of the populations (Mann & Whitney, 1947).

In practice, the test ranks all the "n" observations of the two samples in a single column, from the smallest to the biggest, assigning them an ordinal value (from 1 to 2"n"). Then, it calculates the test statistic (i.e. a quantity derived from the sample through a series of algebraic transformations that is used to perform a statistical test) using the following formula in equation 3.4:

Equation 3.4

$$U_x = R_x - \frac{n_x(n_x + 1)}{2}$$

Where:
R(x) = sum of the ranks in sample 'x'
n = size of sample 'x'

The distribution of "U" is known and tabulated for small samples (n<20), while it has a good approximation for bigger samples in the normal distribution. Since our sample counts more than 20 observations, we can derive the z statistic as follows:

Equation 3.5, 3.6, 3.7

$$z = \frac{U - m_U}{\sigma_U}$$

$$m_U = \frac{n_x n_y}{2} \qquad \sigma_U = \sqrt{\frac{n_x n_y}{12}((n+1) - \sum_{i=1}^{k} \frac{t_i^3 - t_i}{n(n-1)}}$$

Where:
$n=n(x)+n(y)$; $t(i)$= number of subject sharing rank "i",
k = number of distinct ranks.

Even though this test overcomes the limitations of parametric tests, it is not suitable for all the types of data. Indeed, by pooling all the observations together, it fails to preserve important information regarding matching data. Thus, while this does not represent a problem with independent samples, with paired data the test could lead to untruthful results. The literature does not have an official position on whether stock returns should be considered paired or not. In general, two sets of data are defined "paired" when they arise from the same individual at different points in time, while are independent when arise from separate individuals. Following this definition, our groups cannot be defined as paired, since they consist of the returns of two separate groups of stock.

However, the returns of stock groups can be considered as matched data – meaning each data point of one set has a causal relation with one, and only one, data point of the other set. Indeed, given the returns are influenced by the market and external events, it is important not to analyze which group has the highest number of highest values (as the Man-Whitney test does), but how different the groups react to the same external conditions (market crisis, economic expansions etc.). Since randomly comparing the performance of the groups instead of testing who performed better during the same time, would lead to a loss of important information, we decided not to use the Mann-Whitney test.

The most widely used non-parametric test for paired sets of data is the *Sign-Rank test* that compares two related, matched, or paired samples (Wilcoxon, 1945) . It tests the hypotheses that:

H0: difference between the pairs follows a symmetric distribution around zero
H1: difference between the pairs does not follow a symmetric distribution around zero

In practice, the test calculates the absolute difference between the pairs and the sign (positive or negative) of this difference. Then, the absolute values of the differences are sorted from the smallest to the biggest, and a crescent numerical value (called Rank, with values going from 1 to 2N) is assigned. Then, the test statistic W is calculated as follows (equation 3.8):

Equation 3.8

$$W = \sum_{i=1}^{N_r} [sgn(x_i - y_i) * R_i]$$

Where: $sgn(x_i - y_i)$ = sign function of the difference between the paired "i" observations

R(i) = Rank (from 1 to 2N)

N(r) = number of differences (that is, the number of paired observations less observations whose difference is zero)

Under the null hypothesis, W follows a known specific distribution with no simple expression. Therefore, the test compares the empirical W (generated from our sample) with the theoretical W (that there would be if the differences of our paired values followed a symmetric distribution around zero) and calculates the possibility that W follows the same distribution of the empirical W (that is the p-value). More technically, the distribution of W under the null hypothesis has a mean of zero and a standard deviation of:

Equation 3.9

$$\sigma = \sqrt{\frac{N_r(N_r+1)(2N_r+1)}{6}}$$

The value of W can be compared to a reference table. The null hypothesis will be discharged if the absolute empirical W is bigger than the critical W.

As N(r) increases, the distribution of W can be approximated to a normal distribution; therefore, a z-score can be calculated with the following formula:

Equation 3.10

$$z = \frac{W}{\sigma_r}$$

Given the characteristics of our samples and the purpose of our analysis, we decided to use the Sign-rank test to check whether the bottom Portfolios and the top Portfolios gain similar returns (i.e. the difference of their returns is symmetrically distributed around zero) or if one group gains higher returns than the other. To run the sign-rank test we use the statistical program "R". See the appendices (Table "a") for the commands we used to run the test and an example of the output.

3.1.3.2 Statistical tests on Sharpe Ratios

We are now interested in comparing the Sharpe ratios of our groups, which are defined as the ratio between the average excess return of one portfolio divided by the standard deviation of the excess returns. In order to compare the Sharpe ratio of two portfolios, we cannot use any of the methods we exposed before, since those tests are for series of data (or in general for comparing two vectors), while the Sharpe ratio is a single number for the whole period we take into consideration. For this reason, we will need to perform a bootstrap test to create the distribution of the statistic empirically. However, since stock performance generally has a degree of autocorrelation, prior to the bootstrap we have to study the time series, so to understand which degree of autocorrelation (if any) should be included in the analysis in order to avoid loss of important information. The first step in the analysis of time series is to determine whether the series is stationary or not. A stochastic process is defined stationary if its unconditional joint probability distribution does not change through time, and therefore neither mean nor variance does so (Kwiatkowski & Philips, 1992).

Stationary test

To determine the stationarity of a series it is possible to use either decomposition graphs, that however will lead to less accurate analysis, or the Augmented Dickey-Fuller test, that tests the null hypothesis that a unit root is present in a time series sample and uses an alternative hypothesis for the stationarity of the series (Kwiatkowski & Philips, 1992). A unit root is defined as a stochastic trend in a time series that – if present – would result in a systematic but unpredictable pattern in the series' path. The presence of a unit root is one of the main causes of non-stationarity in time series. However, a series can be non-stationary even without any unit root (for example if it is trend-stationary) and therefore it is important to also check for the presence of a trend and of a drift in the time series. The Augmented Dickey-Fuller test performs an autoregressive model, i.e. a model in which the time series is regressed on its own previous values, and tests the presence of unit roots, trends and drifts in the series.

Specifically, the autoregressive function constructed by the test is the following:

Equation 3.11

$$\Delta y_{(t)} = \alpha + \beta t + \gamma y_{(t-1)} + \delta_1 \Delta y_{(t-1)} + \cdots + \delta_{p-1} \Delta y_{(t-p+1)}$$

Where:

α = constant (drift) → if = 0 → no drift

β = time trend coefficient → if = 0 → no trend

γ = unit root → if = 0 → unit root (H0); if < 0 → no unit root (H1)

p = lag of the autoregressive process

The test statistic is calculated as:

Equation 3.12

$$DF_{\tau} = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

If this test statistic is less than the critical value, then the null hypothesis is rejected, and the series will result stationary at the selected confidence level. The idea behind the test is that if the series is stationary its average is constant through time and therefore positive past values will follow negative future ones (and vice versa). Therefore, the sign of the relationship between future and past values should be negative, since a positive value today should preannounce a negative one tomorrow. This is the reason why a stationary time series will have a very negative 'γ', that represents the relationship between past and future values. A further test can be obtained through the graphical decomposition of the time series in its four main components (trend, seasonal, random). In general, seasonality and trend are the factors that make a time-series not stationary. However, the graphic solution – even though it can be a useful tool to better understand the series – is less objective than the ADF test and does not check for unit roots.

Autocorrelation Test

Once we established that the time series are stationary, we can analyze their total and partial autocorrelation. The stationarity assumption is important because only if a series is stationary the statistical dependence between two of its values will imply the statistical dependence between all the pairs with the same lags. There are two main types of autocorrelation that we need to consider: total and partial autocorrelation. Total autocorrelation measures the correlation between a signal with a delayed copy of itself. Differently, partial autocorrelation measures the degree of association between two random variables, with the effect of a set of controlling random variables removed. For example, given two observations y(t-2) and y(t), the total autocorrelation function will calculate the degree of correlation between those values, while the partial autocorrelation function will calculate the degree of correlation between those values after having removed the influence that y(t-1) has on y(t).

Even though it is very important to understand the nature of time-series, the autocorrelation test cannot tell us directly which proper block bootstrap length should be used – for which purpose the construction of an autoregressive model is needed. Nonetheless, the total and partial autocorrelations have a double purpose: first they point us towards the autoregressive model to be selected (as we will see in the next paragraph). Further, in case of two autoregressive models having similar AIC and BIC scores, they will be useful for understanding which model reproduces the series most accurately.

The ARIMA model

The ARIMA (Autoregressive integrated moving average) model provides a description of a stationary stochastic process by analyzing how past values influence future ones. The result is a model that explains the past walk of the time-series and that is able to forecast its future path. The model is the result of the combination of two models (AR, MA) and one integration process (I) and therefore – depending on the value that we assign to its parameters – it can perform an AR test (if q=0), or a MA test (if p=0), or both. The whole process consists in determining from which and with what extent the present (or future) values and model errors are influenced by the past ones. Before running the model, we need to set the value of three parameters: 'p, d, q', where:

- 'p' = determines the number of lags that the AR process will use when calculating the correlation between values. The AR part calculates the extent of the influence of past values on the present ones; the parameter "p" determines the number of lags in which the AR model will test the autocorrelation
- 'd' = determines the number of differencing needed to make the series stationary, i.e. how many times the original values were replaced with the difference between the original value in "t" and the original value in "t-1".
- 'q' = indicates the number of lags that the MA process will use when calculating the correlation between errors. The MA part tests the influence of past errors (i.e. the distance between the forecasted values and the observed ones) on present ones, so to include this error adjustment when making forecasts.

These parameters can be estimates from the stationary and the autocorrelation analysis. Specifically, parameter 'd' is given by the number of times that differencing was needed to make the series stationary. Further, the number of lags in the partial autocorrelation graph suggests the value of 'p', while the number of lags in the total autocorrelation graph suggests the value of 'q'.

However, even though the autocorrelation analysis can give us useful hints which the fittest model is, the most accurate way to select the ARIMA model is to compare the AIC and BIC of models with different parameters. The AIC and BIC are two estimators of the quality of statistical models and are defined as follows:

Equation 3.13

$$AIC = 2k - 2\ln(\hat{L}) \qquad\qquad BIC = \ln(n)\,k - 2\ln(\hat{L})$$

Where:

AIC:

k = number of estimated parameters in the model

L = maximum value of the likelihood function for the model

BIC:

n = the number of data points or equivalently, the sample size

k = number of estimated parameters in the model

L = maximum value of the likelihood function for the model

The model with the lowest AIC and BIC scores is the one that better represents the time series. As we can see, both the models are made of two parts: one depending from the likelihood function of the model (that measures how well the model fits the data) and one that is a penalty for the number of parameters in the model (with the purpose of avoiding overfitting). The only part in which AIC and BIC differ is the latter, with the BIC score being more cautious by having a higher penalty for the number of parameters. Since there is no clear preference in the literature between these two tests, in case of discordant results (i.e. if the ARIMA with the lowest AIC does not have the lowest BIC) we will either use the BIC score (that is indeed more cautious) or choose the model that better reflects the characteristics observed in the total and partial autocorrelation graphs. Finally, a last check on the quality of the model we selected can be given by the analysis of its residual, whose distribution has to be bell-shaped and that should not show any partial or total autocorrelation. As we said before, the main purpose of the model is to produce forecasts on the future possible paths that the series might take. However, given the scope of this paper, we will use this model not to make predictions but to understand which are the lags levels (i.e. the ‘p’ and ‘d’ parameters) that better describe the series, so to include this information in the bootstrap process.

To sum up, in order to determine the length of the bootstrap block that is going to be used to compare the Sharpe ratios, we have to determine the strength with which the autocorrelation characterizes the series. To do so, we first run a stationary test (specifically the Augmented Dickey-Fuller test) to check whether we need to apply a differencing process or not. Then, we study the partial and total autocorrelation graphs to understand the series' behavior and to have some hints on which ARIMA model should be used. Finally, we compare the AIC and BIC values of different ARIMA models to determine which is the fittest and we study the residuals to be assured about the quality of the chosen model. The highest of the 'p' and 'q' parameters of the selected model will be the block length of the bootstrap. Lastly, with the specification obtained by the time-series analysis we will perform a series of bootstraps to empirically retrieve the distribution and confidence interval of the Sharpe ratios of the groups and of the differences in the Sharpe ratios of Top and Bottom groups. The way in which these intervals are constructed will be discussed in the analysis in section 4.

3.2 Funds analysis

The second research question moves the focus from a firm to a fund level. In this section, we will embody a private investor willing to contribute to social and environmental causes – but only if those offer him risk-adjusted returns equal to the ones of peer investments. Thus, two strategies will be compared. The first consists in investing in an actively traded SRI fund, who will use ESG integration strategies to overcome the issues arising from social investing. Embracing this strategy, the investor would be able to gain both the financial return and the positive impact on the society he desired. The second one consists in investing in a market index, a strategy that the investor will embrace only if it would offer returns statistically higher than the former one.

3.2.1 ***Data:***

3.2.1.1 Data gathering and cleaning

Data Cleaning

Data has been retrieved from the Thomson Eikon database. The sample for this dissertation includes all the funds whose investments have been classified by the Thomson database as either 'Ethical, 'Social impact' or 'Green, for a total number of 3.424 funds. In the appendices, it is possible to find the exact specifications that we chose to screen the sample (table 'b'). The sample is further screened by excluding indexed funds and funds of funds: the former because, being indexing a form of passive management, they would fail to reveal the differences between active and passive ESG investing; the latter because they would not be comparable with indices such as the MSCI world that invests in equity only. After this cleaning our sample counts 3.069 active funds. Of those, 2.995 are defined as 'Ethical', 213 as 'Social' and 37 as 'Green' (with some funds having more than one attribute).

In addition to the information collected so far, we are interested in the composition of the funds' portfolios, with the purpose of calculating their weighted ESG scores and the exposure to small, medium and large capitalization stock. The funds' ESG score is calculated following the 'MSCI ESG Fund metrics methodology' (MSCI, 2017) and consists of the weighted average of the ESG scores of the funds' holdings. The composition of the portfolios and the ESG score of the companies in which the funds invested are retrieved using excel add-ins of Thomson Eikon. More information concerning how the score is calculated can be found in section '3.1.1. Data'.

Finally, we are interested to also determine the exposure to small, medium and large stock of our sample of funds. To do so, we retrieve the market capitalization values of all the funds' holdings and we calculate the percentage exposure to each of those categories – where small capitalizations are defined as companies whose market value is lower than 2 billion, medium capitalizations between 2 and 10 billion and large capitalizations higher than 10 billion. However, the information we need to calculate the funds' ESG scores and market exposures are not available for all the 3.069 funds in our sample; therefore, we exclude from the analysis those funds with no information concerning the portfolio composition. This process leaves us with a total of 1.126 funds.

Yet, most of those funds do not have a complete coverage of their portfolio and further cleaning is needed. To determine the minimum portfolio coverage needed to include a fund in the sample, we follow the 'MSCI ESG Fund metrics methodology' (MSCI, 2017) that uses a minimum of 65% ESG coverage for a fund to be included in the database. However, we decided to increase this threshold to 75%, so to have a more reliable and relevant sample. Additionally, we introduce a minimum holding coverage of 85% meaning that, in order to include a fund in our sample, we must know in which companies the 85% of its portfolio is invested and we have to know the ESG score of at least 75% of its overall investments. Further, since our interest lies not only on the ESG score but also on the market capitalization exposure, we exclude those funds for which we do not have market capitalization information for at least 75% of the portfolio.

After having retrieved the information concerning the composition and the exposure of the funds, we are now interested in evaluating their financial performances. To this aim, we download the monthly rolling performance over the last 12 year (from February 2006 to February 2018) and we exclude from the sample all the funds for which we have less than 12 months of performance. Our cleaned sample counts 387 funds (11.3% of the first screening).

Rolling Performance

Rolling performance is retrieved directly from the Thomson database and is defined as the capital gain of an investment at the net of fees, interests and dividends, assuming that dividends are immediately reinvested. Indeed, the comparison between active investments (funds) and passive ones (market index) is only possible if the cost of managing the funds (i.e. fees and interests) is included in the analysis. The data has been retrieved at monthly frequency, since weekly (and daily) data is only available for the last 5 years of activity.

Market Index and Risk-Free Rate

In order to study the competitiveness of our sample, we will compare the funds' performance with a market index. The choice which index to use is complex and can deeply influence the regression results. To solve this problem, Thomson Eikon – the database from which we retrieved the financial and ESG data – calculates the fittest index to use as a benchmark for each fund. The following table summarizes the indices the database selected for our sample:

Table 3.3

Funds' Benchmark	
Index	*Number of Funds*
MSCI	151
Russell	23
S&P	27
FTSE	17
STOXX	29
Swiss Performance Index	5
Other / blank	124
OMX	8
Down Jones	3
TOT	**387**

MSCI is the index most widely utilized as benchmark for our sample. Moreover, the MSCI AC world is mentioned as one of the most well-know and used indices to benchmark fund performance in the respective literature (Bauer, et al., 2002). In view of this, we decide to use MSCI AC world as the benchmark for our analysis. Finally, the risk-free rate – defined as the 1-month T-bill rate – and all the regression factors has been retrieved from the Fama-French website (French, 2018).

Sample Description

Our final sample counts a total of 387 funds, for which we retrieved information concerning the launch date, the domicile, the geographical focus, the NAV, the rolling performance, the pay-out, the market capitalization exposure and the ESG score of their current portfolios. The average age of our sample is 13 years, with the oldest fund of 46 years and the youngest of 1 year. The domicile of the funds is spread across all the continents, even though it is more concentrated in Europe (with 289 funds) and USA (with 69 funds). The presence of 86 funds in Luxemburg suggests that some funds might have chosen their domicile only for a favourable taxation system and therefore the domicile information might be affected by biases. Table ‘c’ in the appendices summarizes the domicile of the funds. The market capitalization exposure is well diversified among the sample, with the average fund having 6,49% invested in small cap, 20,66% in medium and 72,85% in large cap. The clear preference of social funds for large cap stock over small cap could be explained by the fact that small capitalization stocks tend to be riskier and less sustainable than large ones that, having access to more resources, are more likeable for ethical funds (Bauman, 1998).

Table 1.4

Funds Market Capitalization Exposure			
	Small Cap	Medium Cap	Large Cap
Max	62.07%	83.33%	100.00%
Min	0.00%	0.00%	0.00%
Average	6.49%	20.66%	72.85%

To conclude, the average company has a rather high ESG score, of 70/100 and 90% of the funds have a score higher or equal to 60, a result that is consistent with the composition of our sample – entirely made of social and ethical funds.

3.2.1.2 Clustering Process and time frame selection

To reach the scope of our paper, we decide to perform a double grouping of our fund’s sample. In a first step, we will analyse the funds depending on their ESG score, while later we will analyse how fund performance is influenced by the geographical focus.

ESG groups

In this first part of the analysis we divide our sample in 6 groups depending on the ESG score of the funds. The average ESG score of the whole sample is 70.05 and there are only 15 funds with a score lower than 55 and 11 with a score higher than 80. Aiming to create equally distributed groups, three mutually exclusive groups are created: ESG < 70, 70 < ESG < 75 and ESG>75. In addition to these, we further add three groups: two characterized by funds with extreme scores (ESG<55 and ESG>80) and one overall group comprehensive of all the funds without ESG discrimination (that we will call ‘Total Portfolio’).

Geographical focus groups

The second part of the analysis will use a clustering process based on the geographical focus of the funds, to investigate the differences of the social markets across the globe. Indeed, given that social investing is a new-born sector it is reasonable to believe that some countries are more evolved than others in terms of extent and public incentives, leading to more efficient and profitable organizations. Even though Thomson Reuters offers information concerning the domicile of the funds, we decided not to use this information as it might be due to taxation advantages and not to the real operation market of the funds. Thus, following Cortez (Cortez, et al., 2011) and Gregory (Gregory & Whittaker, 2007) we sample the funds depending on their geographical focus. Groups composition is shown in the following table:

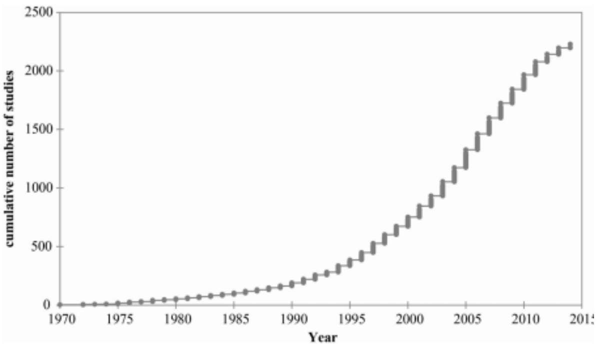
Table 3.4

Geographical Focus	Number of Funds
Global	182
Europe	129
North America	62
Nordic	18
Asia	6
Australia	2
South America	4
South Africa	2

As can be seen from table 3.4, we have few data on Asian, Australian, South American and South African groups. In order to have more robust and reliable results, we decided to run this analysis only on companies whose investment focus is on Global, European, North American and Nordic countries, for a total of 373 funds. Funds in the Nordic group also belong to the European one. The institution of a separated group for funds operating in Nordic countries (i.e. Denmark, Finland, Iceland, Norway and Sweden) is made due to their especially mature and efficient social market, that could lead to interesting results. However, funds in this category are of a limited number and therefore we have to analyse results of this sample with caution, since they might be affected by small-sample biases (Caballero, 2017).

Time frame selection: since 2006

The analysis of social funds will be done across a 12-year period, from February 2006 to February 2018. We picked this time-frame since, being interested in the performance of the funds during crisis and non-crisis periods, it was important to include the largest crisis of recent years in the analysis – which effects started to impact the market in the first half of 2007. We also wanted to compare the pre-crisis and post-crisis performances, and that is why we decided to collect the data since before the subprime crisis. Additionally, social investing aims to achieve long-term results instead of short-term ones (Fulton, et al., 2012) and therefore the analysis needs to be done across a time period long enough to test this strategy. Finally, the market for social investment is very recent, as it can be seen from the number of empirical studies on ESG published over time (figure 3.2) (Friede, et al., 2015). Thus, even though Tomson Eikon would provide us with ESG scores back to 2002, we decided not to consider the first 4 years of scores in our analysis, in order to avoid possible market inefficiencies.



igure 1. Estimated number of empirical studies on the ESG–CFP relation over time.

Figure 3.2

Time Frame Selection: Crisis vs Non-Crisis

The second part of our analysis will test if social fund performance is influenced by market recession. In order to define when a crisis occurred, its starting and ending dates as well as its extent, we used the 'bull and bear' technique – consisting in the analysis of bull and bear phases in the market index. A bull phase is defined as a period of growing returns, while the bear phase is the one with sharply declining prices. A strong market crisis is defined as a 20% downturn in the market index from its last peak in an eight-month period (Pagan & Sossounov, 2002), while a small market crisis is defined as a 15% downturn. We applied this framework to detect bull phases in the MSCI AC world index and we identify two strong crises and two small crises periods.

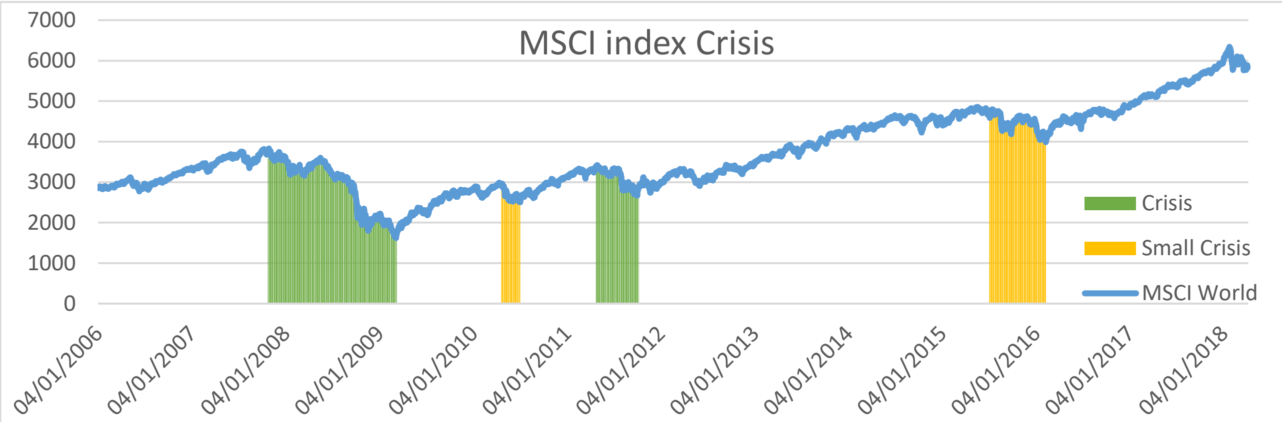


Figure 3.3

The following table summarize the crisis start and end dates, its length and the relative change in the index value:

Table 3.5

Crisis Periods					
Start		End			
Date	Index Value	Date	Index Value	% Change	Length (months)
02/07/2015	4696.4	11/02/2016	3986.1	-15.12%	7
02/05/2011	3422.5	04/10/2011	2670.3	-21.98%	6
26/04/2010	2957.6	05/07/2010	2503.0	-15.37%	3
31/10/2007	3832.1	09/03/2009	1616.5	-57.82%	17

Time Frame Selection: Since 2011

The third and last period we consider in our analysis is 7 year long and goes on from February 2011 to February 2018. The reason why we chose this timeframe is to test whether, in recent years, social funds became more efficient than in the past. As a matter of fact, being a new-born market, it is reasonable to believe that the efficiency of social investments has been increasing through time. Further, the overall number of active funds has been constantly changing, as new funds were introduced in the market every year (figure 3.4). The higher number of active funds from 2011 should give us a more truthful representation of the current profitability of the social investment market, without being influenced by the performance of fewer funds when the market was still in its early phase.

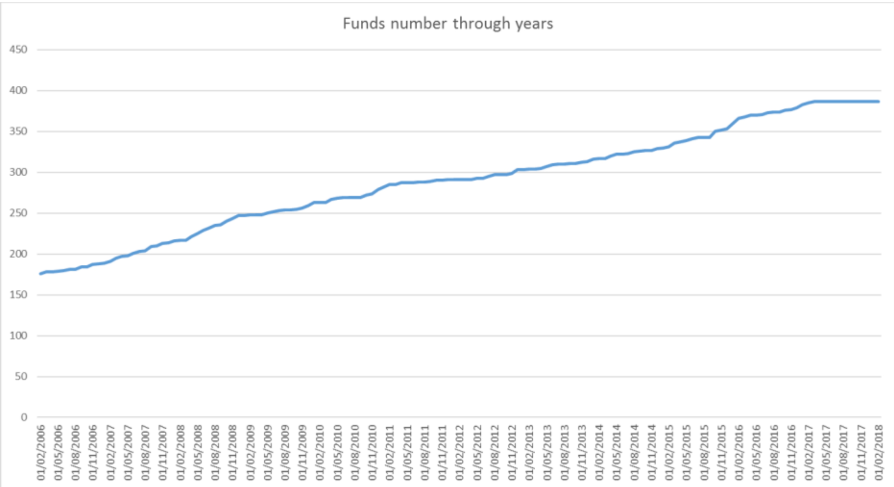


Figure 3.4

Therefore, we decide to run our third analysis from 2011, when 285 funds were already actively trading (while only 176 funds were active in 2006). We expect the sample to perform better when only the last 7 years of returns are considered, thanks to a more mature and efficient market as well as to a higher number of funds. Further, to study the presence of differences among the different time-frames more thoroughly, we will perform a Chow test, that studies whether the true coefficients of two linear regressions are equal through the use of a statistic (called Chow statistic) that follows a F distribution. The null hypothesis of the test is that the true coefficient values of two regressions do not differ from each other; thus, a p-value of less than 0.05 in this test would indicate the presence of a structural break. The following steps show how the test works.

Given a linear regression:

Equation 3.14

$$y_t = a + bx_{1t} + cx_{2t} + \varepsilon$$

We can split our data in two groups creating two regressions:

Equation 3.15; 3.16

$$y_t = a_1 + b_1x_{1t} + c_1x_{2t} + \varepsilon \qquad\qquad y_t = a_2 + b_2x_{1t} + c_2x_{2t} + \varepsilon$$

Given H0: a(1) =a(2), b(1) = b(2) and c(1)=c(2)

The test statistic is given as:

Equation 3.17

$$Test\ statistic = \frac{\left(\frac{S_c - (S_1 + S_2)}{k}\right)}{\left(\frac{S_1 + S_2}{N_1 + N_2 - 2k}\right)}$$

Where:

S(c) → sum of squared residuals for the combined data

S(1) → sum of squared residuals for the first group

S(2) → sum of squared residuals for the second group

N(1) → number observations first group

N(2) → number of observations second group

k → total number of parameters (3 in the example)

Then, the p-value is calculated as for a F distribution with k and N(1) +N(2)-2k degrees of freedom.

3.2.2 *Methodology*

In order to analyse the returns of the funds, a number of econometric models will be utilized. When it comes to performance analysis, the most commonly used framework is the single index model, developed by (Sharpe, 1963):

Equation 3.28

$$\boldsymbol{r_{i,t} - r_f = \alpha_i + \beta_i(r_{m,t} - r_f) + \epsilon_{i,t}}$$

Where:

$r_{i,t}$ = return to stock "i" in period "t"

r_f = the risk-free rate

r_{mt} = return to the market portfolio in period "*t*"

α = stock's alpha, or abnormal return

β = stocks' beta

ε = residual returns. They are assumed to be independent normally distributed

Even though this method has been widely adopted by both theoretical and empirical studies, it has been also subjected to criticism, mainly because it assumes portfolio performance to be fully explained by one market factor. In the analysis of portfolio performance, the model shows whether the portfolio outperformed or underperformed the market index (through a positive or negative alpha) and the degree to which it is sensible to market changes (through the Beta). However, the single-factor model fails to determine which strategy is responsible of the over- (or under-) performance of the investment. Given the purpose of this paper (i.e. to determine how the integration of SRI strategies contributes to the funds' returns), knowing that a fund over- (or under-) performed the market is not sufficient, since we would not be able to understand whether this performance has to be attributed to strategies based on social principles or to other strategies that the funds might have undertaken simultaneously (e.g. small capitalization strategy). For this reason, Elton (1992) proposed to add a small capitalization factor to the model, an alteration further evolved by Fama and French (1992, 1993, 1996) with the introduction of a book-to-market factor.

The model resulting from these changes is called the Fama-French three-factor and reads:

Equation 3.39

$$\boldsymbol{r_{i,t} - r_f = \alpha_i + \beta_i(r_{m,t} - r_f) + \beta_{1,i}SMB_t + \beta_{2,i}HML_t + \epsilon_{i,t}}$$

Where:

<u>SMB_t</u> = is the difference in return between a portfolio of companies with a high book to market ratio and a portfolio of companies with a low book to market ratio. It is calculated as "the average return on the nine small stock portfolios minus the average return on the nine-big stock portfolio" (French, 2018)

Equation 3.20

$$SMB = \frac{Small\,Value\ +\ Small\,Neutral\ +\ Small\,Growth}{3} - \frac{Big\,Value\ +\ Big\,Neutral\ +\ Big\,Growth}{3}$$

<u>HML_t</u> = is the difference in return between a portfolio of companies with a high book to market ratio and a portfolio of companies with a low book to market ratio. It is calculated as "the average return on the two value portfolios minus the average return on the two growth portfolios" (French, 2018):

Equation 3.21

$$HML = \frac{Small\,Value\ +\ Big\,Value}{2} - \frac{Small\,Growth + Big\,Growth}{2}$$

Although this model increases the explanatory power of the regression, it does not account for the momentum anomaly described by Jegadeesh and Titman (1993). To fill this gap, Carhart (1997) extends the three-factor model with the introduction of the Moment factor:

Equation 3.22

$$\boldsymbol{r_{i,t} - r_f = \alpha_i + \beta_i(r_{m,t} - r_f) + \beta_{1,i}SMB_t + \beta_{2,i}HML_t + \beta_{3,i}WML_t} \in_{i,t}$$

Where:

<u>WML</u>: is the Momentum factor and it is the equal-weight average of the returns for the two winner portfolios for a region minus the average of the returns for the two loser portfolios:

Equation 3.23

$$WML = \frac{Small\,High\ +\ Big\,High}{2} - \frac{Small\,Low + Big\,Low}{2}$$

Additionally, in 2014 Fama and French decided to further extend their original model by introducing two additional factors (profitability and investment), asserting that there was strong evidence that those factors influence average portfolio returns (2014). This new model reads:

Equation 3.24

$$\boldsymbol{r_{i,t} - r_f = \alpha_i + \beta_i(r_{m,t} - r_f) + \beta_{1,i}SMB_t + \beta_{2,i}HML_t + \beta_{3,i}RMW_t + \beta_{4,i}CMA_t} + \in_{i,t}$$

Where:

<u>*RMW*</u> = is the difference between the returns of firms with robust (high) and weak (low) operating profitability

Equation 3.45

$$RMW = \frac{Small\ Robust\ +\ Big\ Robust}{2} - \frac{Small\ Weak + Big\ Weak}{2}$$

<u>*CMA*</u> = is the difference between the returns of firms that invest conservatively and firms that invest aggressively

Equation 3.56

$$CMA = \frac{Small\ Conservative\ +\ Big\ Conservative}{2} - \frac{Small\ Aggresive + Big\ Aggresive}{2}$$

In order to have a more complete understanding of the funds' performance, we will run a regression for each of the models we introduced above. The presence of a positive and relevant alpha in the two-factor model could be, for example, due to a mixed strategy of small capitalization and robust operating profitability. If this is the case, the relevant alpha found in the two-factor model would lose its relevance when analysed with the 5-factor model. Since our funds have been clustered on their characteristic of being social, the presence of a statistically relevant alpha in all the regression models (where different strategies are tested) would increase the probabilities that the reason behind superior (or inferior) returns is the social side of the investments.

Additionally, as we anticipated, given the evidence pointed out by Nofsinger and Varma (2014) concerning superior performances of social funds during market crisis, we decide to run four additional regressions (one for each of the models introduced before) where, through the introduction of a dummy variable, we will test the performance of the funds during crisis and non-crisis periods. For instance, the three-factor model after the dummy variable has been added is defined as:

Equation 3.27

$$\boldsymbol{r_{i,t} - r_f = \alpha_i + \beta_1(r_{m,t} - r_f) + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + D * \alpha_i + D * \beta_4(r_{m,t} - r_f) + D * \beta_{5,i}SMB_t + D * \beta_{6,i}HML_t} + \in_{i,t}$$

Where: D = dummy variable → 1 if crisis

0 if not crisis

The interpretation of the regression's output is the following: the coefficients of the factors multiplied for the dummy variable (for example $\beta_{5,i}$) display the <u>differences</u> in the exposure between crisis and non-crisis times. Instead, the absolute exposure of the groups to certain factors during market crisis will be given by the coefficient of the factor (not multiplied for the dummy variable) plus the coefficient of the factor multiplied for the Dummy variable. Finally, the exposure during times of market stability will be given, as for a regression without dummy variables, by the coefficient of the factor not multiplied for the dummy variable. Since the 'crisis' coefficients are the sum of two coefficients of the regression, we must run a hypothesis testing to calculate a p-value, where the sum of the two coefficients is tested against the null hypothesis of being zero.

Additionally, following the new evidence pointed out by Cortez et. al (2011), Gregory and Whittaker (2007) and Bauer et al. (2007) concerning the influence of local markets in ESG funds' performances, we will introduce a local factor in the regression, described as the difference between a global market index and the local value of the same index (in our case, it will be the difference between the MSCI AC world and the local MSCI indices). This latter analysis is run only for the groups created using geographical focus as discriminating variable (and not for the one estimated by ESG score).

4. Analysis

4.1 Financial Performance of stocks

In the next sections we proceed with the comparison of the returns of social and non-social stocks. In doing so, first we study whether the two categories of stocks offer different excess returns; then, through the Sharpe ratios, whether this difference is justified by a difference in risk. The core study will be preceded by a descriptive analysis, that will help us to gain a first understanding of stocks' behaviour over time and to graphically identify possible differences in the returns of social and non-social firms.

4.1.1 *Descriptive analysis*

Our global 5% (10%; 20%) portfolio consisting of the best ESG stocks has an average score of 85 (81; 79), compared to the bottom portfolio with an ESG score of 18 (22; 24). Over the observed time-frame, the ESG score of the 'top' portfolios (and especially the one of the 15% top-rated stocks) has constantly increased year after year, while the controversial stocks' scores had been falling until the beginning of 2015 and have improved since then. Moreover, some geographical differences arise. As can be seen in the graph below (figure 4.1), European stocks show the highest ESG scores, in the top as well as in the bottom portfolios, followed by North America and Asia.

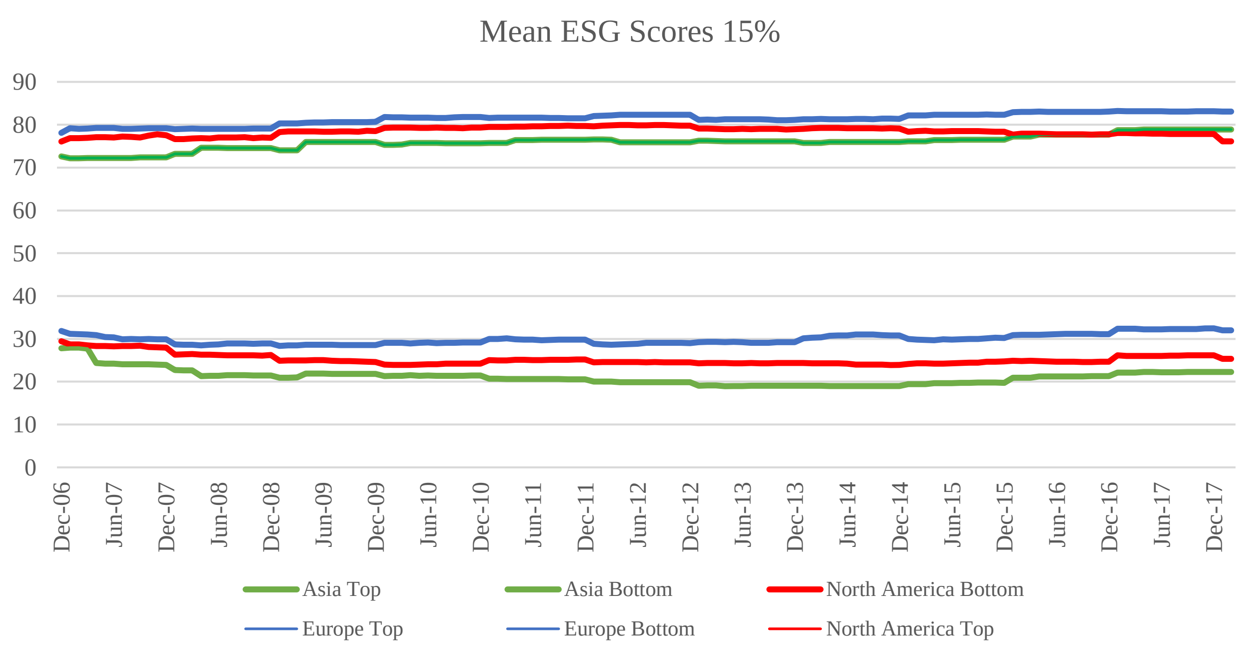


Figure 4.1

The regional differences within ‘Top’ and ‘Bottom’ groups are consistent with differences observed by other ESG rating providers, e.g. Morningstar (Lefkovitz, 2018). Asia has not only the lowest ESG scores in every group, but also the greatest differences in the scores between the top and bottom portfolios. We can also see that top portfolios in all three markets had been experiencing increasing scores until the end of 2014, but socially responsible firms from North America were not able – in contrast to their European and Asian counterparts – to continue improving afterwards and subsequently fell behind the other markets (figure 4.1). This trend is valid for all percentiles, except for the North American “Top 5%”, that still have a slightly better ESG score than the Asian peer. Looking at the bottom portfolios in each region, we notice that the differences between the regions are bigger than in the top portfolios, with European stocks ranking higher than North American stocks, followed by Asia.

Comparing the mean excess returns of the global stock groups, it becomes apparent that, on average, bottom portfolios yielded higher returns than top portfolios, suggesting that public equity investments in controversial stocks gained higher returns than comparable investments in socially responsible stocks during our observation period (table 7). The differences between the bottom and top 5% (10%, 15%) stocks are 0,20 (0,16, 0,18) percentage points in favor of non-social stocks (table 4.2). However, the time-series visualization of the excess returns shows no distinct trend in their development, with no portfolio outperforming the other over a significant period (figure 4.2).

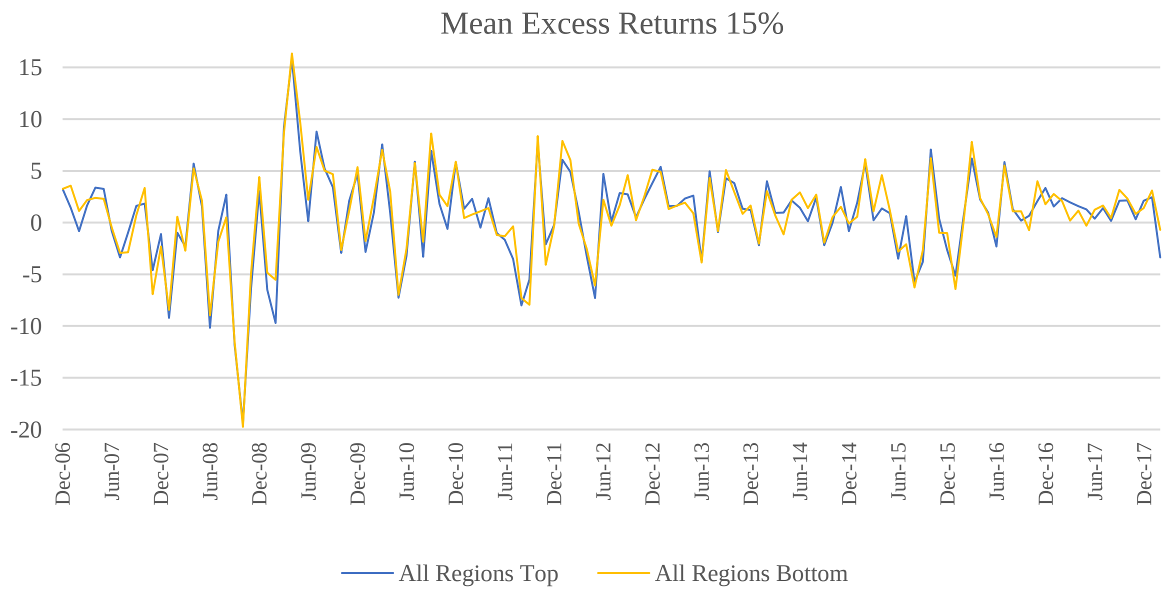


Figure 4.2

4.1.2 *Numerical analysis*

Prior to the analysis of the main statistics, we will perform a series of correlation tests. These tests serve a double purpose. First, they tell us if and how the ESG score might influence other statistics. Further, we use these tests to check whether the ESG screens we applied do not mirror other portfolio selection strategies as the ones associated to market capitalization and book to market ratios.

Table 4.1

Correlation Analysis			
	Book-to-market	Market Cap	Volatility
ESG Score	0.46	0.41	-0.06

As can be seen (table 4.1), the correlation between the ESG score and all factors are quite low, although higher than the ones found by Auer and Schuhmacher (2016), who report a correlation of 0,29 for Book-to-market, 0,39 for market capitalization and 0,29 for volatility factor. This difference in results is probably due to the application of a different ESG score from the ones of Auer and Schumacher, who retrieve it from the Sustainalytics database. Further, the different composition of the stock universe and the longer time-frame of our analysis might also influence the correlation results. Despite comparatively higher correlations, we still regard them as sufficiently low to conclude that our ESG screens constitute a portfolio selection strategy distinct from the afore mentioned. The values indicate that creating portfolios on the basis of these factors will generate groups of stocks substantially different from those generated by an ESG score-based approach.

The global picture is largely reflected at the regional level; in North America, mean excess returns are higher for the lowest-rated firms than for the top-rated ones, namely 0.45; 0.42 and 0.35 percentage points for the 5%, 10% and 15% percentiles. In Europe, the differences are less significant but have the same sign (0.14; 0.2; 0.04). Only in Asia, top-portfolios show higher return means than bottom-portfolios, at the 5% and 15% threshold. The differences of 0.08 and 0.02 percentage points are quite low though, and bottom portfolios perform better at the 10% percentile (Table 4.2).
The following table summarizes the main statistics for the different groups:

Table 4.2

Summary of groups' statistics						
			ESG score	Excess Return (%)	St. Dev.	Sharpe Ratio
All	5	Top	84.99	0.67	0.046	0.143
		Bottom	18.48	0.89	0.044	0.203
	10	Top	81.44	0.65	0.044	0.145
		Bottom	21.78	0.85	0.044	0.192
	15	Top	78.74	0.65	0.044	0.148
		Bottom	24.16	0.83	0.044	0.188
Asia	5	Top	82.1	0.54	0.054	0.100
		Bottom	16.14	0.49	0.045	0.108
	10	Top	78.6	0.52	0.051	0.102
		Bottom	18.97	0.54	0.045	0.119
	15	Top	75.91	0.54	0.050	0.109
		Bottom	21.11	0.53	0.045	0.119
Europe	5	Top	86.65	0.40	0.054	0.074
		Bottom	22.82	0.67	0.055	0.121
	10	Top	83.69	0.50	0.053	0.095
		Bottom	26.9	0.59	0.051	0.116
	15	Top	81.38	0.56	0.050	0.113
		Bottom	29.94	0.62	0.051	0.121
North America	5	Top	85.3	0.91	0.046	0.199
		Bottom	20.51	1.35	0.052	0.260
	10	Top	81.43	0.77	0.045	0.170
		Bottom	23.25	1.21	0.051	0.238
	15	Top	78.45	0.81	0.045	0.179
		Bottom	25.2	1.13	0.050	0.227

Table 4.2 draws a picture of non-social companies (i.e. the ‘bottom’ groups) gaining higher returns than the social firms. Further, it appears that as the social factor becomes more extreme, so do the differences between the Top and Bottom groups. On the other hand, volatility seems to have a negative relation with the ESG score – with social firms showing lower standard deviation. However, the reduced volatility does not seem to justify lower returns, as the Sharpe ratio is constantly lower for the ‘top’ portfolios.

Even though this first analysis already yields some valuable insights, we have to remember that what we are observing now is only a small sample generated from a wider population, and therefore we will statistical tools as described in the methodology section to proper compare the results we present.

4.1.3 *Statistical Tests*

As we have introduced in the “methodology” chapter, we will now run a series of statistical tests to understand whether there is a real difference between the statistics of interest (i.e. returns and Sharpe ratio) of Top and Bottom groups.

4.1.3.1 Comparison of group’s excess returns

To compare the excess returns of the groups with high and low ESG scores we use the non-parametric Sign-rank Wilcoxon test, that tests the null hypothesis of absence of differences between the populations. To run this test, we run the statistical software “R”; the command input and the output for one group of stock can be found in appendix (table ‘d’). The following table summarizes the output values of the test for the 12 paired groups:

Table 4.3

Summary of tests on Excess Return

	Confidence Interval		
	Lower Boundary	***Upper Boundary***	*P-value (bootstrap)*
All_5	**-0.1300**	**0.496**	0.2457
All_10	**-0.0850**	**0.405**	0.1931
All_15	**-0.0530**	**0.393**	0.1226
N_5	**-0.5770**	**0.298**	0.5536
N_10	**-0.4260**	**0.314**	0.7024
N_15	**-0.3990**	**0.252**	0.6153
E_5	**-0.3950**	**0.565**	0.7268
E_10	**-0.4240**	**0.272**	0.6465
E_15	**-0.3880**	**0.238**	0.5954
A_5	**-0.0290**	**0.756**	0.0681
A_10	**0.0230**	**0.667**	0.0380
A_15	**0.0240**	**0.569**	**0.0322**

The purpose of the test is to check whether the returns of low-ESG scores groups are statistically higher than the ones of a high EGS score. This is done by creating a confidence level for the difference of excess returns. With a p-value < 0,05 we can reject – at the 95% confidence level – the null hypothesis and accept that the difference pairs do not follow a symmetric distribution around zero. Hence, the observed difference would be significant, and we could confirm that one stock group outperforms another in terms of excess return.

From the table above (table 4.3), we can see that there is no statistical difference in the excess returns of social and non-social firms operating in Asia, Europe and more in general in the World – and therefore we cannot reject the null hypothesis. The failure to reject the null hypothesis does not prove that there is no actual difference between the two populations, but it implies that – given the observed values – these differences are not strong enough to determine a statistical dominance. However, only considering firms operating in North America a statistically relevant difference arises, with low-rated firms having higher returns than high-rated firms. Thus, for these groups we do reject the null hypothesis.

This first result would confirm that, in most of the geographical areas, there is no statistical dominance of the returns of non-social companies over the ones of social firms and that therefore it is not certain and necessary to sacrifice returns when investing in companies with green and social policies. This is different for North America, where firms with low ESG scores gain statistically higher returns than social stocks. However, this analysis is not sufficient to reach satisfying conclusions: a higher or lower return could be justified by a different level of risk, or volatility, of the groups and therefore further analyses is needed. Hence, we will compare the Sharpe ratios of the different groups in the next part.

4.1.3.2 Comparison of group's Sharpe ratio

We will now compare the Sharpe ratio of social and non-social stock portfolios, which is defined as the ratio between the average excess return of one portfolio divided by the standard deviation of the excess returns. As introduced in the methodology, the process needed to compare the Sharpe ratios of two groups is more complicated than the one we required for excess returns, since – being the ratio a single number covering the whole time-frame and not a vector – it is necessary to empirically retrieve its distribution through a bootstrap process. However, before running the bootstrap it is important to study the autocorrelation degree of the time series, so not to lose any important information regarding the structure of data during the bootstrap process.

Stationarity test

The first step in the determination of the autocorrelation level is given by a stationary analysis, that we will perform using the Augmented Dickey-Fuller test – whose null hypothesis assumes the presence of a unit root, a drift or a trend in the series. We demonstrate the process of the stationary analysis for the group 'All_Bottom_5' step-by-step, while for the other groups we will summarize the final results in table 4.5, as the methodology is the same for all portfolios. However, it is possible to find a more detailed analysis for all the groups in the appendix ('Autocorrelation'). To run the Augmented Dickey-Fuller test we choose the statistical program 'R'.

Among the many output statistics, we are interested in the tau3, intercept and tt values. The tau3 is the statistic of the test, where the null hypothesis is the presence of a unit root in the series. The 'tt' indicates the presence of a trend while a significant intercept would show the presence of a drift. Finally, the z.lag.1 is the 'gamma' term and – if significantly lower than zero – indicates the absence of a unit root (the tau3 test is on this variable). The other variable (z.diff.lag) is the error term of the autoregression. Table 4.4 show the 'R' output for the ADF test on the stock group 'All_Bottom_5'.

Table 4.4

Augmented Dickey-Fuller test “All_Bottom_5”

```
ur.df(y, type = c("none", "drift", "trend"), lags = 1, selectlags = c("Fixed", "AIC", "BIC"))

# Augmented Dickey-Fuller Test Unit Root Test #
Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
     Min       1Q   Median       3Q      Max
-0.17507 -0.01662  0.00269  0.01893  0.11342

Coefficients:
             Estimate   Std. Error  t value  Pr(>|t|)
(Intercept)  1.987e-04  7.421e-03   0.027    0.979
z.lag.1     -7.486e-01  1.029e-01  -7.273 3.03e-11 ***
tt           9.039e-05  9.550e-05   0.946    0.346
z.diff.lag   7.905e-02  8.751e-02   0.903    0.368
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04206 on 129 degrees of freedom
Multiple R-squared:  0.3524,	Adjusted R-squared:  0.3374
F-statistic:  23.4 on 3 and 129 DF,  p-value: 3.687e-12

Value of test-statistic is: -7.273 17.6446 26.4664
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.99 -3.43 -3.13
phi2  6.22  4.75  4.07
phi3  8.43  6.49  5.47
```

In this example, the tau3 statistic is equal to -7.273, significantly lower than its critical values at any confidence level (the lowest being -3.99). Additionally, neither the intercept nor the tt value are relevant (with a P(>|t|) equal to 0.979 and 0.346 respectively). Finally, the z.lag.1 – as we could expect from the tau3 value – is significantly lower than 0 at any confidence level. Therefore, we can assert that the series does not have unit root, drift and trend and that therefore it is stationary.

As we introduced before, an alternative to the Augmented Dickey-Fuller test is a graphical decomposition of the series, also possible to run through the 'R' software. The following graph (figure 4.3) shows the output for the series 'All_Bottom_5':

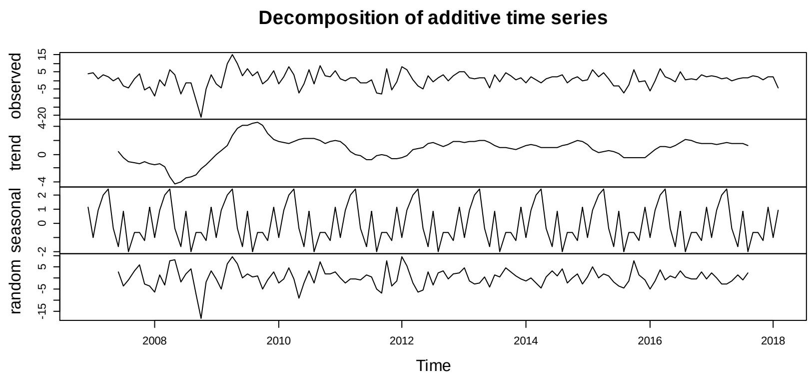


Figure 4.3

We can see that the trend of the series does not have a clear path (generally if it is not stationary it decreases or increase steadily) and it is quite constant over time. Moreover, the seasonality part has a very low impact on the series, while the random part is the one mostly influencing the overall path the most. Even though the decomposition can be a useful tool, from now on we will perform it only in the case of non-stationarity of the series, to better understand the reasons behind it. In the appendix it is possible to find the stationarity test for all the groups ('Autocorrelation'). The following table summarizes the main outputs of the 'ADF' test for the 24 groups:

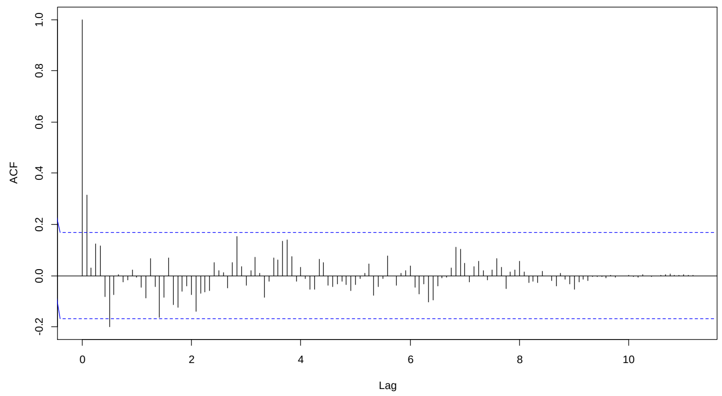
Table 4.5

Augmented Dickey-Fuller Test results			
	Tau3	***Intercept PR(>|t|)***	***tt PR(>|t|)***
All_Top_5	-8.56	0.5601	0.1330
All_Bottom_10	-7.59	0.9660	0.3030
All_Top_10	-8.45	0.5676	0.1378
All_Bottom_15	-7.65	0.9370	0.2930
All_Top_15	-8.40	0.5492	0.1291
Asia_Bottom_5	-7.23	0.7860	0.4030
Asia_Top_5	-7.68	0.3190	0.1050
Asia_Bottom_10	-7.63	0.8330	0.3890
Asia_Top_10	-7.66	0.3200	0.1020
Asia_Bottom_15	-8.40	0.5492	0.1291
Asia_Top_15	-7.67	0.3500	0.1080
Europe_Bottom_5	-7.82	0.8120	0.3850
Europe_Top_5	-9.08	0.5698	0.2600
Europe_Bottom_10	-7.83	0.7040	0.3170
Europe_Top_10	-8.78	0.5050	0.1830
Europe_Bottom_15	-7.71	0.7040	0.3100
Europe_Top_15	-8.66	0.5729	0.1906
North_America_Bottom_5	-6.97	0.5500	0.5540
North_America_Top_5	-8.07	0.8220	0.1640
North_America_Bottom_10	-7.36	0.5880	0.5240
North_America_Top_10	-7.96	0.8240	0.2260
North_America_Bottom_15	-7.44	0.6980	0.4400
North_America_Top_15	-8.02	0.9500	0.2750

As we can see, all the Tau3 are far below their critical value (3.99 for a 1% confidence level), and the p-values of ‘intercept’ and ‘tt’ are always higher than 0.05. Thus, we can assert that all the groups are already stationary and thus we can proceed in our analysis without any transformation.

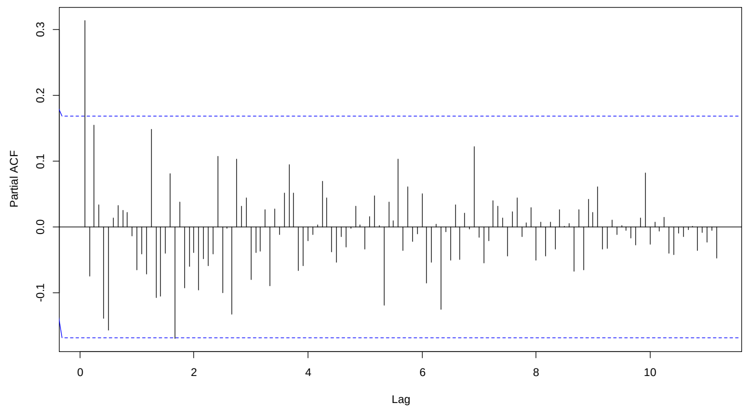
Autocorrelation Test

Once we established that the time series are stationary, we can analyze their total and partial autocorrelation, a process that will be of great help in the determination of the fittest ARIMA model to use in the next step. The next graphs show the autocorrelation in the group ‘All_Bottom_5’. It is possible to find the graphs for the other groups in the appendix (‘Autocorrelation’).



Total autocorrelation

Figure 4.4



Partial autocorrelation

Figure 4.3

When interpreting the partial (pacf) and total (acf) autocorrelation graphs, it is important to keep in mind that the acf function starts with a lag=0, while the pacf starts with a lag=1 (so we have to compare the first line of pacf with the second line of acf). In our example, we can see that the acf graph shows a ‘cut-off’ (i.e. it becomes statistically not different from 0) after a lag=1, while the pacf graph never shows a relevant correlation.

Given this first result and following the theory exposed in the methodology, we should choose a bootstrap block length equal to 1 (the highest lag among the autocorrelations). However, in order to have a more accurate indication on how past values influence future ones we will construct an autoregressive model. As anticipated, the total and partial autocorrelation analysis has a double purpose: first it points us towards the autoregressive model we should use, as we show in the next paragraph. Further, in case of two autoregressive models having similar AIC and BIC scores, it will be useful for understanding which model reproduces the series most accurately.

The ARIMA model

We now have to determine the ARIMA(p,d,q) model that better fits the time series we are analyzing. A rule of thumb is to find the 'p' and 'q' coefficients from the total and partial autocorrelation graphs, where 'p' would be given by the lag at which the pacf function cuts off and 'q' by the lag at which the acf function cuts off. Even though this analysis could indicate a good starting point to test the models, we find it more convenient and accurate to run the ARIMA models for several 'p' and 'q' combinations and to subsequently select the one with the lowest AIC and BIC scores.

We then run the ARIMA model on the 'All_Bottom_5' group. The same steps for all the other groups can be found in the appendix ('Autocorrelation'). From the graphs above (figure 4.4, 4.5) we can see that the ACF clearly cuts off at a lag=1, that will likely result in q=1, while the PACF – not being relevant – would indicate a p=0. Further, since the series was stationary and did not require any differencing transformation, we must set d=0. This would result in an ARIMA model (0,0,1), that can also be read as a MA(1) model with zero degrees of differencing. However, the autocorrelation analysis can only point us towards a plausible solution but cannot show which is the fittest model for the time series. Thus, we calculate the BIC, AIC and AICc scores for different sets of parameters. Table 4.6 summarizes the results.

Table 4.62

Arima Model for "All_Bottom_5"		
C(1,0,0)-> ACF=-470.29	AICc= -470.11	BIC= -461.58
C(1,0,1)-> ACF= -470.04	AICc= -469.73	BIC= -458.42
C(1,0,2) -> ACF= -468.15	AICc= -467.69	BIC= -453.62
C(1,0,3) -> ACF= -467.71	AICc= -467.05	BIC= -450.27
C(1,0,4) -> ACF= -468.04	AICc= -467.16	BIC= -447.71
C(2,0,0) -> ACF= -469.04	AICc= -468.73	BIC= -457.41
C(2,0,1) -> ACF= -468.15	AICc= -467.69	BIC= -453.63
C(2,0,2) -> ACF= -466.7	AICc= -466.05	BIC= -449.27
C(2,0,3) -> ACF= -471.55	AICc= -470.67	BIC= -451.21
C(2,0,4) -> ACF= -469.62	AICc= -468.47	BIC= -446.37
C(3,0,0) -> ACF= -470.26	AICc= -469.79	BIC= -455.73
C(3,0,1) -> ACF= -468.31	AICc= -467.65	BIC= -450.88
C(3,0,2) -> ACF= -469.63	AICc= -468.75	BIC= -449.29
C(3,0,3) -> ACF= -468.66	AICc= -467.52	BIC= -445.42
C(3,0,4) -> ACF= -467.64	AICc= -466.2	BIC= -441.49
C(4,0,0) -> ACF= -468.41	AICc= -467.75	BIC= -450.98
C(4,0,1) -> ACF= -466.72	AICc= -465.83	BIC= -446.38
C(4,0,2) -> ACF= -470.35	AICc= -469.2	BIC= -447.1
C(4,0,3) -> ACF= -468.57	AICc= -467.13	BIC= -442.42
C(4,0,4) -> ACF= -466.65	AICc= -464.87	BIC= -437.59
C(0,0,1) -> ACF= -472.01	**AICc= -471.83**	**BIC= -463.3**
C(0,0,2) -> ACF= -470.03	AICc= -469.72	BIC= -458.41
C(0,0,3) -> ACF= -468.49	AICc= -468.02	BIC= -453.96
C(0,0,4) -> ACF= -469.99	AICc= -469.33	BIC= -452.55

We can see that the model that maximizes the AIC, AICc and BIC is the one with parameters=c(0,0,1), which is consistent with the observations in the autocorrelation analysis. While the decision in this case was straightforward, it can happen that the AIC and the BIC scores point towards different models. In such a situation, it is a rule of thumb to choose the model that better explains the autocorrelation graphs (partial and total) so not to lose any important information of the series.

Therefore, for this example we choose the ARIMA(0,0,1) and we use a block length equal to the highest of the 'p' and 'q' parameters (in this case, block length = 1). To have further confirmation that the selected model fits the sample, we can test whether the residuals of the model are made from a white noise process. The residuals are the part of the time series that the model fails to predict, and we want to test whether they are totally random or if there are some trends or correlations that the model still fails to forecast.

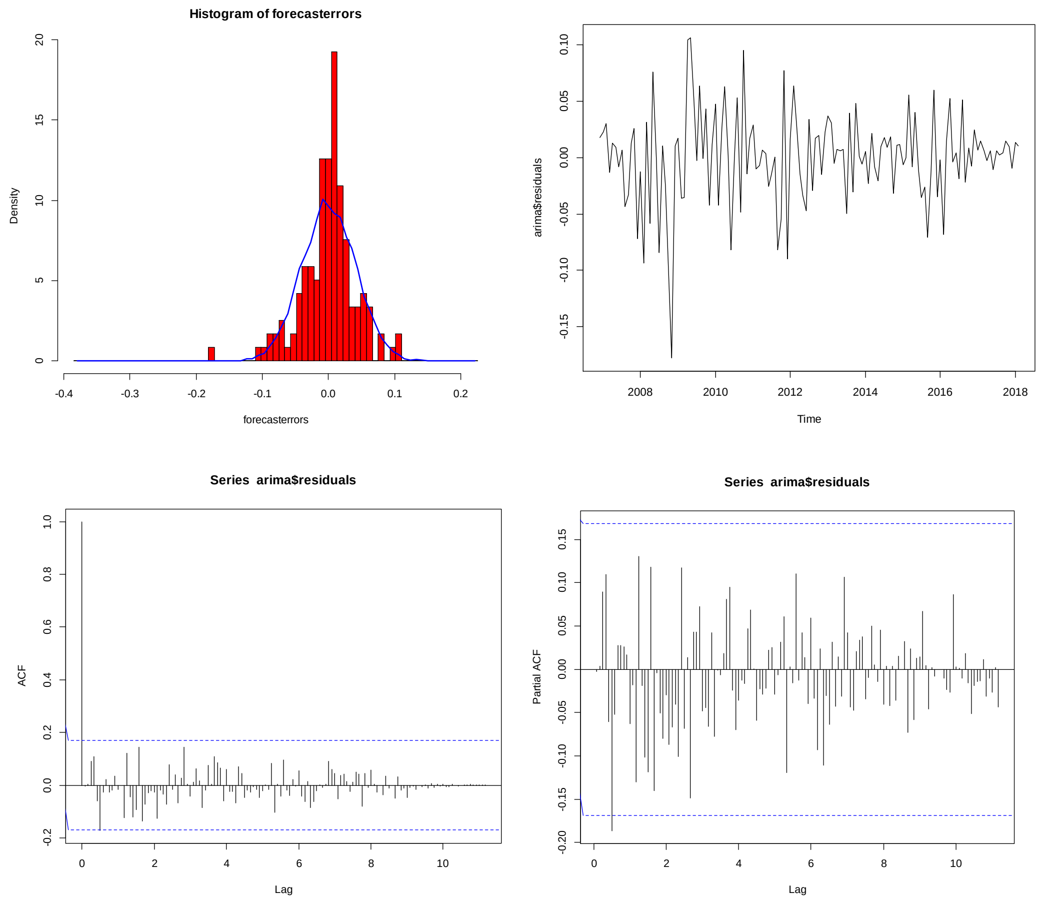


Figure 4.6

The residuals of our model are well approximated to a white noise with normal distribution (mean 0) and clearly stationary (that can be further confirmed by running an ADF test). Autocorrelation graphs show a weak pacf at lag 6, probably a small semiannual seasonality effect, that however does not undermine the analysis result. Also, the total correlation is never significant. We therefore accept the model and use a lag=1 for our circular block bootstrap. An alternative to manually test the different mix of parameters is using a command in R that automatically chooses the model with the lowest BIC or AIC. It is possible to find the specific command we used and the output of R, that coincides with the result we showed above, in the appendix (table ‘g’).

Thus, after having tested the groups for stationarity, autocorrelation and having selected the fittest ARIMA model, we can define which is the autocorrelation degree in each of the 24 time series. The following table summarizes the block length that we will utilize during the bootstrap process:

Table 4.73

Group Name	Bootstrap Block Length
All_Bottom_5	1
All_Top_5	3
All_Bottom_10	1
All_top_10	3
All_Bottom_15	1
All_Top_15	3
Europe_Bottom_5	3
Europe_Top_5	3
Europe_Bottom_10	3
Europe_Top_10	3
Europe_Bottom_15	3
Europe_Top_15	3
Asia_Bottom_5	1
Asia_Top_5	1
Asia_Bottom_10	1
Asia_Top_10	1
Asia_Bottom_15	1
Asia_Top_15	1
NAmerica_Bottom_5	1
NAmerica_Top_5	3
NAmerica_Bottom_10	1
NAmerica_Top_10	1
NAmerica_Bottom_15	1
NAmerica_Top_15	1

Block Bootstrap

We now perform a series of bootstraps to retrieve the confidence interval of a Sharpe ratio and of the difference between two Sharpe ratios. Specifically, we will do the latter to compare the risk adjusted performances of social and non-social firms, while we will construct the confidence interval for the single Sharpe ratios to better understand the performance of the stocks in absolute terms.

To calculate the confidence interval for a single statistic (in our case the Sharpe ratio), we proceed as follows: first, we construct a circular block bootstrap with replacement, using the block length previously found (Table 4.7). Then, we create a vector containing the statistic of interest for each of the random samples generated by the bootstrap. The vector will represent the empirical distribution of the Sharpe ratio. The confidence interval is calculated as a quantile interval and is given by the values occupying the '$\left((1-\alpha)*N\right)th$' and '$(\alpha*N)th$' position of the sorted vector (with N being the number of bootstrap replicates and alpha the confidence level). The table below shows the steps of this process for the 'Europe_Bottom_5' group.

First, we perform a block bootstrap test for the series, in our example 'Europe_Bottom_5', with a block length of 3 (see table 4.7) and a number of samples created equal to 5.000, as indicated by Ledoit and Wolf (2008) .

Table 4.84

```
BLOCK BOOTSTRAP FOR TIME SERIES

Fixed Block Length of 3

Call:
tsboot(tseries = E_Bottom_5_excess, statistic = Sharpe, R = 5000, l = 3,
    sim = "fixed", orig.t = TRUE)

Bootstrap Statistics:
     original        bias           std. error
t1* 0.1204814     0.003785371      0.100628
```

Then, we sort the 5.000 Sharpe ratios from the smallest to the biggest and we calculate the 95% confidence interval by taking the 150th (that is 0.05*5000) and the 4850th (that is 0.95*5000) values of the sorted statistic. The following table summarizes the results:

Table 4.9

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

CALL :
boot.ci(boot.out = E_Bottom_5_boot, conf = 0.95, type = c("perc"))

Intervals :
Level     Percentile
95%   (-0.0772,  0.3134 )
Calculations and Intervals on Original Scale
```

Thus, the confidence interval for the Sharpe ratio of the 'Europe_Bottom_5' group is given by: [-0.0772; 0.3134]. The following table summarizes the Sharpe ratio confidence level:

Table 4.10

	Lower bound	***Upper bound***
All_Bottom_5	0.0296	0.4063
All_Top_5	-0.0480	0.3604
All_Bottom_10	0.0175	0.3931
All_Top_10	-0.0462	0.3691
All_Bottom_15	0.0077	0.3867
All_Top_15	-0.0442	0.3748
Asia_Bottom_5	-0.0622	0.2994
Asia_Top_5	-0.0668	0.2882
Asia_Bottom_10	-0.0507	0.3068
Asia_Top_10	-0.0645	0.2943
Asia_Bottom_15	-0.0473	0.3124
Asia_Top_15	-0.0558	0.3037
Europe_Bottom_5	-0.0791	0.3207
Europe_Top_5	-0.1011	0.2647
Europe_Bottom_10	-0.0782	0.3227
Europe_Top_10	-0.0869	0.2947
Europe_Bottom_15	-0.075	0.324
Europe_Top_15	-0.0722	0.3191
North_America_Bottom_5	0.0847	0.4594
North_America_Top_5	0.0030	0.4276
North_America_Bottom_10	0.0635	0.4361
North_America_Top_10	-0.0037	0.3596
North_America_Bottom_15	0.0559	0.4175
North_America_Top_15	0.0095	0.3728

From the results we can see that not all the groups are able to produce a Sharpe ratio statistically different from zero at the 95% confidence level, which implies the average returns are higher than the risk-free rate. Specifically, groups from Asia and Europe do not gain a Sharpe ratio statistically different from zero, while groups belonging to the ‘North America’ category generally do. Finally, when we consider the overall market (i.e. the ‘All’ groups), we can see that only firms with low ESG scores do not contain zero in their confidence interval, while firms with high ESG score do.

Even though there are some differences in the confidence intervals between the groups, it is not possible to reach any conclusion concerning whether one group is gaining risk-adjusted returns higher than others. In fact, the comparison of confidence intervals allows to conclude that a statistic is higher than another only when the related confidence intervals are not overlapping – while it is not possible to reach any conclusion when they are not. Since all the Sharpe ratios’ confidence intervals do overlap, we will run a more specific test to determine whether there is a statistical difference in the risk-adjusted returns of different groups. Thus, we will now run a series of bootstrap tests to retrieve the empirical distributions of the difference between Sharpe ratios. From this distribution we will calculate the 95% confidence interval, that will show whether the Sharpe ratio of one group stochastically dominates another.

First, we have to calculate the empirical distribution of the Sharpe ratio of each group by using a paired circular block bootstrap as well as using ‘Table 4.7’ to select the appropriate block length – so to preserve the autocorrelation structure of the time series data. The necessity of performing a paired bootstrap arises from the transformations we conduct in the second part of the analysis, when we subtract the bootstrapped statistics of one group from the ones of the matching group. The paired bootstrap associates the returns from one group with the ones from the matching group and computes a block bootstrap with replacement on the first sample. Then, the permutation on the second group is made by associating each value of the bootstrapped samples to the matched value from the paired group. Finally, the statistic (the Sharpe ratio) for each bootstrapped sample from both groups is calculated and the difference between these statistics is performed. The resulting vector will represent the empirical distribution of the difference between the Sharpe ratios.

Thus, we sort the vector in an ascending order; the N*(alpha) number in the sorted vector will represent the lower boundary of the ‘1-alpha’ confidence interval, while the N*(1-alpha) number will represent the upper boundary.

Finally, we will calculate the p-value for the null hypothesis of the equality of the Sharpe Ratios (i.e. H0= sharpe(i) – sharpe(j) = 0), that is defined as the lowest confidence level at which the confidence interval does not contain zero. Since the pairs of samples need to share the same specifications for the bootstrap, we will use the highest between two samples in case the two have a different autocorrelation structure (and therefore a different block length), so to avoid losing important information concerning the structure of the series.

Additionally, even though the bootstrapping process we just explained will give us a good insight on the differences in the Sharpe ratios among the groups, the state-of-the-art estimation is given by Ledoit and Wolf (2008). In their paper, they test the null hypothesis of equal risk adjusted performance, (i.e. H(0) = n(i)-n(j) =0, where n(i) and n(j) are the true Sharpe ratios of two portfolios "I" and "j") by creating a studentized confidence interval through a circular block bootstrap (Ledoit & Wolf, 2008). The main difference from the process we explained above is given by the studentification of the bootstrap and by a complex estimation of standard errors adjusted for heteroskedasticity. The punctual description on how the model works goes beyond the scope of this paper, and for this reason we will base our conclusion only on the results obtained by the bootstrap process previously described. However, since the statistical software R offers a specific command ('sharpeTesting') to automatically run this test, we will also present these results. By showing the outcome of this blackbox-like function, we hope to offer the reader a more truthful picture on the real differences characterizing social risk-adjusted returns from non-social ones. Thus, table 4.11 summarizes both the confidence intervals found thought the bootstrap process and the p-values generated by the 'sharpeTesting' command.

Table 4.115

	Statistics of Sharpe ratios						
	Original Sharpe ratios			***Confidence Interval***			
	Bottom Group	*Top Group*	*Original difference*	***Lower Boundary***	***Upper Boundary***	*P-value (bootstrap)*	*P-value (Ledoit and Wolf)*
All_5	0.2030	0.1426	0.0604	**-0.0149**	**0.1541**	0.1208	0.1088
All_10	0.1922	0.1453	0.0469	**-0.0107**	**0.1113**	0.1108	0.1228
All_15	0.1877	0.1475	0.0402	**-0.0127**	**0.0955**	0.1256	0.1196
N_5	0.2601	0.1986	0.0616	**-0.0265**	**0.1535**	0.1680	0.1532
N_10	0.2381	0.1698	0.0682	**0.0012**	**0.1412**	0.0472	0.0571
N_15	0.2272	0.1793	0.0479	**-0.0094**	**0.1060**	0.0928	0.1200
E_5	0.1209	0.0737	0.0472	**-0.0365**	**0.1288**	0.2728	0.2845
E_10	0.1163	0.0951	0.0212	**-0.0504**	**0.0901**	0.4588	0.5462
E_15	0.1208	0.1127	0.0081	**-0.0586**	**0.0655**	0.8416	0.7910
A_5	0.1077	0.0998	0.0080	**-0.0717**	**0.0874**	0.8704	0.8424
A_10	0.1187	0.1024	0.0164	**-0.0550**	**0.0867**	0.6652	0.6462
A_15	0.1195	0.1092	0.0103	**-0.0653**	**0.0755**	0.7988	0.7702

The differences in the Sharpe ratios between most of the groups are not statistically different from zero at the 95% confidence level, as the zero is contained in most of the confidence intervals and the vast majority of the related p-values are higher than 0.05. Thus, for these pairs we cannot reject the null hypothesis of equalities of the Sharpe ratios – which implies the absence of a significant dominance of one statistic on the other. However, the rejection of the null hypothesis does not imply that the difference between the statistics will be zero on average, as each confidence interval contains a variety of other possible values and the distribution over these values is unknown. The only pair that gains a statistically relevant difference between the Sharpe ratios, in favour of the non-social group, is ‘North_America_10’, for which we reject the null hypothesis in favour of the alternative one. Finally, it is interesting to notice how the process we adopted seems to be a good approximation of the estimation made by Ledoit and Wolf (2008), with the values derived from the ‘sharpeTesting’ function being very similar to the p-values calculated by the bootstrap process.

4.2 Financial performance of funds

The second part of the analysis is concerned with the profitability of social funds when compared to a market index. Our sample counts 387 funds that are analysed using two different clustering processes (by geographical focus and by ESG score), during three different timeframes (from 2006, from 2011 and during crisis periods) and over four different factor models (two factors, three factors, four factors and five factors). The asterisks next to regression outputs indicate the statistical relevance of the factors' value: three are for a p-value lower than 0.01 (99% confidence interval), two for a p-value lower than 0.05 (95% confidence level) and one for a p-value lower than 0.1 (90% confidence interval), while no mark indicates that the value is not statistically relevant for any of the confidence intervals considered. Finally, prior to the regression, a descriptive analysis on each group is run.

4.2.1 *ESG-score groups*

4.2.1.1 Descriptive Analysis

This section helps us to identify trends and differences among timeframes and to gain a wider understanding on which results we should expect from the regression analyses. The following table summarizes the main statistics of the different groups:

Table 4.126

Statistics of ESG groups

	Total Portfolio	*ESG<55*	*ESG < 70*	*ESG 70 < x < 75*	*ESG > 75*	*ESG>80*
ESG Av.	70.1	48.6	62.6	72.5	77.7	81.4
Rolling Av.	0.51%	0.57%	0.46%	0.44%	0.34%	0.30%
St. Dev. Rolling	2.93%	3.10%	2.51%	2.41%	2.58%	2.59%
Number of funds	387	15	150	135	102	11

The three mutually exclusive groups (ESG<70, 70<ESG<75, ESG>75) are characterized by a similar number of members and by a similar standard deviation. However, the average rolling performance results to be higher for lower ESG scores – even though this level of analysis is not sufficient to assert a statistical dominance of a group on another. Further, when we also consider the two extreme groups and the Total portfolio, it is possible to detect an inverse correlation between ESG score and standard deviation. The same conclusion is reached through a correlation analysis between funds' ESG score, rolling performance and standard deviation:

Table 4.137

Correlation Analysis

	ESG score	Standard Deviation		ESG score	Rolling performance
ESG score	1		ESG score	1	
Standard Deviation	-0.26548	1	Rolling performance	-0.09068	1

The result indicates that, on average, the increment of one point in the ESG score results in a decrement of 0.26 points in the standard deviation and of 0.09 points in the rolling performance of the fund. The inverse relation between social score and return volatility is in line with what is anticipated by the literature, that sees ESG as a tool for qualitative screening. Further, the inverse relation between social score and rolling performance might be explained by the fact that funds with high ESG scores might be more willing to sacrifice part of their return to have a higher social impact.

The following table and graph summarize the exposure of each group to small, medium and large market capitalization.

Table 4.14

Market capitalization Exposure

	ESG < 50	*ESG < 70*	*70< x < 75*	*ESG > 75*	*ESG > 80*
Small	21.34%	11.87%	3.67%	2.31%	2.40%
Medium	50.89%	30.14%	17.37%	11.07%	10.45%
Large	27.76%	57.99%	78.96%	86.62%	87.14%

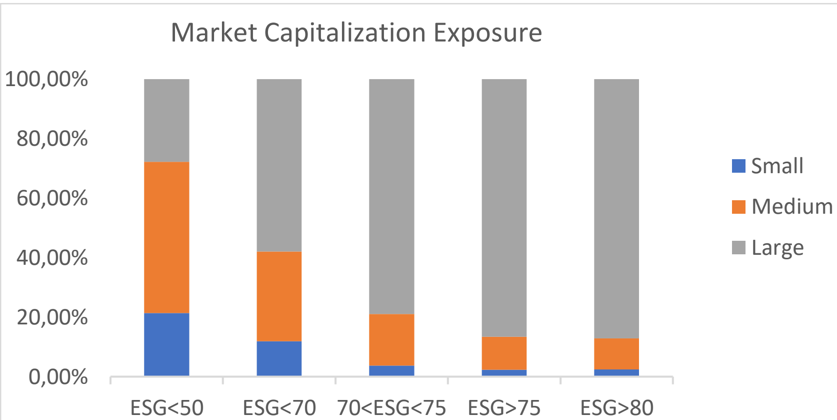


Figure 4.4

Likewise, what we observed in the stock analysis, the ESG score seems to be highly correlated with the capitalization size of the firms in which social funds invest. From the graph and table above, we can see how funds with high ESG scores tend to invest almost 90% of their capital in large capitalization stocks – against the 27% invested by funds with a low score. This result might be explained by the riskier nature and the smaller amount of capital available for social purposes of small stocks that might lead to funds with a strong social purpose to prefer large market capitalization stocks.

We will now provide a graphical visualization of the returns of the funds. In the following graph it is possible to find the normalized (in base 100) performances of all the groups plotted against the market index. Moreover, a visual identification of the crisis and non-crisis periods is included. Finally, to have a better understanding of the groups' trends, in the Appendix (figures a-e) six graphs are available where the groups are plotted individually against the market index.

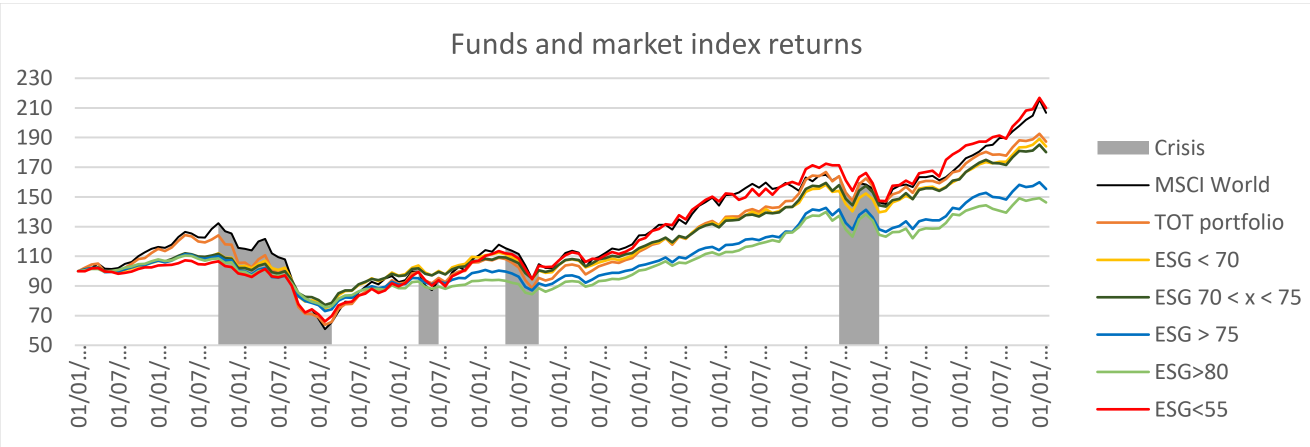


Figure 4.5

The graph shows higher average returns for the MSCI World index over most of the social funds. In general, we can distinguish between two different paths. The first one is followed by funds with a low/medium ESG score (ESG<55; ESG<70; 70<ESG<75) and is characterized by both volatility and average return resembling the ones of the market index. When compared to the MSCI World, the funds belonging to this category show a similar behaviour during both market crisis and market stability. Moreover, these funds gained absolute returns lower than the market index up to 2013 and higher afterwards.

The second path is followed by funds with a high ESG score (ESG>80; ESG>75) and is characterized by lower but steadier returns than the ones of the market index. This difference continues also during market crisis, with MSCI world suffering less during the two smaller crises and with social funds being less affected by the bigger ones. These results are consistent with our correlation analysis, where we found an inverse relationship between ESG score and both rolling performance and standard deviation.

The total portfolio, which includes all the social funds in our sample, seems to behave similarly to the MSCI World during both market and non-market crises, with a slightly lower volatility during the first crisis (2007) and with a better financial performance from 2014 to 2017.

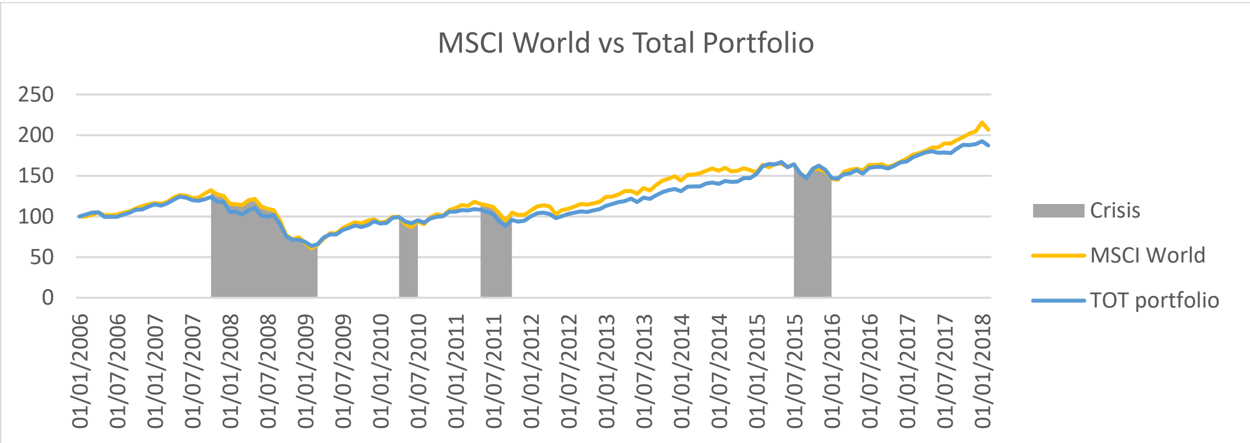


Figure 4.6

To conclude, from the core analysis we do not expect to have significant differences between the returns gained from the totality of social funds (i.e. the ‘Total Portfolio’) and the market index (i.e. the ‘MSCI world’). However, the descriptive analysis suggests possible significant differences in the sub-categories in terms of both absolute returns and betas – with high ESG score groups having lower returns but lower exposure to systematic risk.

4.2.1.2 Regression Analysis

Statistical tests

As described in the methodology, we will use different linear regressions to study the performance of funds and to compare them to the market index. However, prior of running the core analysis we need to perform a series of statistical tests to validate the regressions' results. In general, there are three components that are important to test when performing a regression: 1) the statistical relevance of the model and of the estimated variables, 2) the goodness of fit, 3) the absence of multicollinearity among the independent variables.

The *statistical relevance* concerns both the overall model and the single factors. When running a regression, we are interested mainly in two aspects: that the model itself actually explains the dependent variable and that the estimates of the independent coefficients are statistically relevant. The statistical relevance of the *model* is given by the F statistic and tests the following hypothesis:

> *H0: The fit of the intercept-only model and your model are equal*
> *H1: The fit of the intercept-only model is significantly reduced compared to your model*

Thus, if the p-value associated at the F-statistic is <0.05 it means that the model we estimated can predict the dependent variable better than a model without any predictor (called intercept-only). In practice, it means that the relationship between the model and the response variable is statistically relevant.

The statistical relevance of the estimated *parameters* is given by the single t-statistic associated to each estimated parameter. It tests the hypothesis:

> *H0: the estimated parameter is not statistically different from zero*
> *H1: the estimated parameter is statistically different from zero*

Thus, a p-value lower than 0,05 would mean that the estimated parameter (i.e. the relation between the dependent and the independent variables) is statistically different from zero.

These two statistics are both already present in the summary of the output of the regression obtained in R with the command ‘lm’. The only exception is for the regression model where the dummy variables are used. In this case, the coefficients during crisis periods will be the sum of the coefficients of the factor and the coefficient of the factor multiplied by the Dummy variable. Thus, the p-value for this sum of coefficients has to be calculated separately through a hypothesis test, testing the null hypothesis of this sum being zero. It is possible to run this test by using the command ‘hypothesisTesting’ in R.

The *goodness of fit* of a linear regression measures how well the model fits the observed values. While the F-test measures the relevance of the model, the goodness of fit measure the extent of this relevance. The most widely applied statistic to test the goodness of fit is the R square (or R^2), that is the proportion of the variance in the dependent variable predicted from the independent variables. The R^2 vary in a range from 0 to 1, where 1 means that the model perfectly fits the observed data. The main issue with R^2 is that it increases by adding more independent variables (even though these do not explain the dependent one) and therefore it is exposed to overfitting biases. For this reason, we will use the adjusted R^2 – that has a penalty for the number of parameters in the model. This statistic is also present in the summary of the output of the regression obtained in R with the command ‘lm’.

The last test we need to run before proceeding with the analysis is the multicollinearity test. *Multicollinearity* is the phenomena where a number of independent variables in a multiple regression are closely related to each other. Indeed, if the explanatory variables of the regression are highly correlated with each other, the regression coefficients can become unstable and the estimated parameters unreliable. Additionally, the presence of multicollinearity can lead to a higher (but distorted) F-test and can (rarely) also affect the R^2 – increasing the risk of accepting a model that does not fit the observed data. The most widely used approach to test the variables for multicollinearity is through the calculation of a ‘*Variance inflation factor’* (or VIF), defined as the ratio of variance in a model with multiple terms divided by the variance of a model with one term alone. In practice, it measures how much the variance of the estimated regression coefficient has increased due to collinearity.

Given a linear regression $y = a + b(1)x(1) + b(2)x(2) +b(3)x(3)$,

the VIF test on x(1) calculates the R^2 of the regression: $x(1) = a + b(2)x(2)+b(3)x(3)$

Then, the VIF score is calculated as:

Equation 4.1

$$VIF = \frac{1}{1 - R^2}$$

Therefore, the VIF calculates how much of an independent variable can be explained by the other independent variables. Since R^2 goes from a minimum of 0 to a maximum of 1, the VIF goes from a minimum of 1 (if R^2 is 0) to an upper limit of infinite (as R^2 approaches 1). Generally, multicollinearity is defined not strong if the VIF score is <5 (and therefore the R^2 is <0.8). Since the test of the multicollinearity does not depend on the dependent variable, it is enough to do it once for each model we will utilize. The following table summarizes the VIF for the different factor models:

Table 4.15

	VIF score					
	MSCI_World	*SMB*	*HML*	*MWL*	*RMW*	*CMA*
3 factor	1.039664	1.001707	1.039480			
4 factor	1.142982	1.004733	1.199981	1.319171		
5 factor	1.641551	1.154679	1.796739		1.653959	1.841614

As we can see, all the VIF scores are lower than 2; therefore, the multicollinearity is not strong enough to influence the validity of the regression results.

Regression: "since 2006":

Table 4.16

Since 2006

2 factor

	Intercept	*MSCI excess*	*Adj R^2*
TOT Portfolio	0.0001	0.7934 ***	83.5%
ESG < 55	0.0012	0.7154 ***	81.3%
ESG <70	0.0007	0.5971 ***	84.3%
70< ESG <75	0.0009	0.5215 ***	78.8%
ESG > 75	-0.0001	0.5217 ***	71.1%
ESG > 80	-0.0002	0.4536 ***	59.9%

3 factor

	Intercept	*MSCI excess*	*SMB*	*HML*	*Adj R^2*
TOT Portfolio	0.0624	0.8004 ***	0.1495	0.1164	83.8%
ESG < 55	0.0012	0.7110 ***	0.3828 ***	0.0047	83.5%
ESG <70	0.0006	0.6040 ***	0.2353 ***	-0.127**	85.9%
70 < ESG< 75	0.0008	0.5296 ***	-0.0081	-0.1092 *	78.9%
ESG > 75	-0.0001	0.5243 ***	-0.1021	-0.0209	71.0%
ESG > 80	-0.0002	0.4551 ***	-0.0668	-0.0115	59.5%

4 factor

	Intercept	*MSCI excess*	*SMB*	*HML*	*MWL (MOM)*	*Adj R^2*
TOT Portfolio	0.0003	0.7899 ***	0.1438	-0.151 *	-0.0472	83.8%
ESG < 55	0.0012	0.7102 ***	0.3824 ***	0.0022	-0.0035	83.3%
ESG <70	0.0008	0.5968 ***	0.2314 ***	-0.152 **	-0.0328	85.9%
70 < ESG <75	0.0009	0.5229 ***	-0.0117	-0.131 **	-0.0300	78.9%
ESG > 75	0.0397	0.5160 ***	-0.1060	-0.4780	-0.3709	70.9%
ESG > 80	-0.0001	0.4499 ***	-0.0696	-0.0287	-0.0237	59.3%

5 factor

	Intercept	*MSCI world excess*	*SMB*	*HML*	*RMW*	*CMA*	*Adj R^2*
TOT Portfolio	0.0007	0.7584 ***	0.0997	-0.0219	-0.0471	-0.2525 **	84.0%
ESG < 55	0.0020	0.7108 ***	0.3437 ***	-0.1559 *	-0.3430 **	0.1286	84.2%
ESG <70	0.0013	0.5794 ***	0.1931 ***	-0.1270 *	-0.1435	-0.1038	86.0%
70 < ESG < 75	0.0013	0.5131 ***	-0.038 *	-0.1147	-0.1092	-0.0649 *	78.8%
ESG > 75	0.0008	0.4803 ***	-0.1683	0.0198 *	-0.1731	-0.2173 *	71.4%
ESG > 80	0.0009	0.3977 ***	-0.1452	0.0747	-0.1556	-0.3107 **	60.6%

Intercepts' p-values

	2 factor	3 factor	4 factor	5 factor
Tot Portfolio	0.915	0.962	0.837	0.621
ESG<55	0.344	0.329	0.329	0.117
ESG<70	0.464	0.501	0.411	0.202
70<ESG<75	0.386	0.416	0.351	0.242
ESG>75	0.922	0.921	0.975	0.530
ESG>80	0.873	0.874	0.934	0.560

The groups' returns result to be well explained by the models, with most of the adjusted R^2 having satisfying values. The explanatory power of the regression is higher for models with more factors – suggesting that the added factor does have an explanatory power – and for groups with a low-medium ESG score. The Beta value is statistically relevant and lower than 1 for all the groups, indicating an exposure lower than the market index to systematic risk. Further, a clear path can be detected concerning the correlation between Beta and ESG score, with funds with higher social score being characterized by lower Beta. This trend is consistent with our finding in the descriptive analysis, where we showed an inverse relationship between volatility and ESG score. None of the intercepts, that can be interpreted as Jensen's alpha, are statistically different from zero, but their value (and their p-value) increases (decreases) for lower ESG scores, a result in line with the inverse correlation between the funds' scores and their rolling performances.

The size factor (SMB) results positive and relevant for low ESG scores, while negative and not relevant for higher scores (except for one case where it is negative and relevant). This result is a good reflection of the exposure to small, medium and big capitalizations of our groups, since groups with low ESG scores (<55 and <70) have an overall exposure to medium and small stocks higher (73% and 44% respectively) than groups with high ESG scores (around 18%). The value factor (HML) is relevant mainly for groups with a low ESG score and is mostly negative for all the groups, suggesting that social funds are more exposed to low book-to-market firms (instead of high book-to-market ratio as it would result in a positive HML). This result is not in line with the stock analysis, where we found a positive correlation between ESG and low book-to-market ratio.

Finally, both the profitability (RMW) and the investment (CMA) factors are negative, though only the latter is statistically relevant most of the times. These results draw a picture of social funds as aggressive investors mainly interested in companies with weak operating profits. This profile is consistent with social investing, where companies do not worry only for their main business (that would explain a weak operating profit) and where – due to a lack of diversification – investments can be more aggressive. The F-test is lower than '2.2e-16' for all the regressions and therefore all the adjusted R^2 are statistically different from zero.

Regression “Crisis/non-Crisis”

Table 4.17

Dummy/Non Crisis

2 factor

	Intercept		MSCI world excess		Adjusted R^2
	Dummy	Non-Crisis	Dummy	Non-Crisis	
TOT Portfolio	-0,0048	0,0025	0,1264*	0,6951***	0,8393
ESG < 55	-0,0038	0,0023	-0,0023	0,7066***	0,8127
ESG <70	-0,0033	0,0019	0,0334	0,5655***	0,8434
70 < ESG < 75	-0,0031	0,0019	0,0209	0,4988***	0,7873
ESG > 75	-0,0043	0,0015	0,0416	0,4817***	0,7121
ESG > 80	-0,0050	0,0016	0,0504	0,4059***	0,6014

3 factor

	Intercept		MSCI world excess		SMB		HML		Adjusted R^2
	Dummy	Non-Crisis	Dummy	Non-Crisis	Dummy	Non-Crisis	Dummy	Non-Crisis	
TOT Portfolio	-0,0057 *	0,0025	0,1337 **	0,699***	0,1064	0,0926	-0,2039	-0,0516	0,8426
ESG < 55	-0,0031	0,0021	-0,0118	0,7099***	-0,2270	0,4746***	0,1280	-0,0576	0,8345
ESG <70	-0,0037 *	0,0019	0,0311	0,5744***	0,0482	0,2101 **	-0,0562	-0,1165 *	0,8587
70 < ESG < 75	-0,0039 *	0,0020	0,0241	0,5056***	-0,0360	-0,0018	-0,1027	-0,0848	0,7882
ESG > 75	-0,0052 *	0,0015	0,0570	0,4779***	-0,0716	-0,0884	-0,2446	0,0504	0,7127
ESG > 80	-0,0059 *	0,0016	0,0698	0,3993***	-0,0351	-0,0709	-0,3134	0,0836	0,6011

4 factor

	Intercept		MSCI world excess		SMB		HML		MWL (MOM)		Adjusted R^2
	Dummy	Non-Crisis	Dummy	Non-Crisis	Dummy	Non-Crisis	Dummy	Non-Crisis	Dummy	Non-Crisis	
TOT Portfolio	-0,0070 **	0,0023	0,0887	0,71***	0,0734	0,0948	-0,3479 *	-0,0340	-0,1497	0,0252	0,8461
ESG < 55	-0,0035	0,0017	-0,0378	0,7158***	-0,2457	0,479***	0,0357	-0,0263	-0,1051	0,0450	0,8341
ESG <70	-0,005 **	0,0020	0,0088	0,5721***	0,0316	0,2085 **	-0,1181	-0,1292 *	-0,0551	-0,0183	0,8602
70 < ESG < 75	-0,005 **	0,0023	0,0064	0,5015***	-0,0494	-0,0045	-0,1466	-0,1070	-0,0331	-0,0320	0,7890
ESG > 75	-0,0066 **	0,0018	0,0347	0,4727***	-0,0885	0,0919	-0,3001 *	0,0227	-0,0422	-0,0398	0,7140
ESG > 80	-0,0071 **	0,0019	0,0524	0,3949***	-0,0483	-0,0739	-0,3558 *	0,0595	-0,0309	-0,0345	0,5992

5 factors

	Intercept		MSCI world excess		SMB		HML		RMW		CMA		Adjusted R^2
	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	
TOT Portfolio	-0,0067 *	0,0029	0,1258	0,6831***	0,1091	0,0591	-0,1043	-0,0404	0,2832	-0,1143	0,0029	-0,1449	0,8404
ESG < 55	-0,0010	0,0029*	-0,0135	0,6906***	-0,2147	0,4201***	0,2339	-0,2455 **	-0,1833	-0,2959 *	-0,2682	0,2499	0,8399
ESG <70	-0,0029	0,0022*	0,0297	0,5608***	0,0510	0,1813 **	-0,0727	-0,1089	-0,0689	-0,0994	0,0772	-0,1196	0,8566
70 < ESG < 75	-0,0037	0,0022	0,0344	0,4958***	-0,0167	-0,0221	-0,1364	-0,0718	-0,0022	-0,0662	0,1162	-0,1027	0,7831
ESG > 75	-0,0046	0,0019	0,0453	0,4597***	-0,0828	-0,1240	-0,2541	0,1045	-0,0194	-0,0988	0,1311	-0,2533	0,7098
ESG > 80	-0,0051	0,0019	0,0345	0,3823***	-0,0838	0,3824	-0,2963	-0,1014	-0,0419	0,1740	0,0900	-0,0614	0,6003

Table 8

Intercepts' p-value

	2 factor		3 factor		4 factor		5 factor	
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*
Tot Portfolio	0.1060	0.1320	-0.059*	0.1300	-0.024**	0.1920	0.1220	0.1030
ESG<55	0.2010	0.1610	-0.2670	0.1740	-0.2370	0.3020	0.1653	0.073*
ESG<70	0.1380	0.1190	-0.0809*	0.1087	-0.031**	0.1020	0.2420	0.077*
70<ESG<75	0.1820	0.1330	-0.098*	0.1237	-0.0432**	0.1010	0.1820	0.1100
ESG>75	0.1290	0.3530	-0.0722*	0.3477	-0.03**	0.2780	0.1700	0.2720
ESG>80	0.1160	0.3540	-0.0658*	0.3580	0.037**	0.3100	0.6890	0.3100

The two tables above show the main outputs of the regression. As introduced before, the coefficient associated with the factors multiplied with the dummy variable does not show the absolute exposure of the funds during market crisis, but the differences in this exposure between crisis and non-crisis. Indeed, the absolute exposure of a group to a factor during market crisis is given by the sum of the dummy coefficient with the coefficient of the same factor when it is not multiplied for the dummy variable. It is possible to find these data (i.e. the exposure of the funds during market crisis) in Table 'd' in the appendix.

With the introduction of the dummy variable we are able to discern the performance during crisis periods from the ones during times of market stability. Differently from what expected, social funds did not show superior performance during crisis. Instead, the dummy's alphas are always negative and mostly statistically relevant, suggesting a poorer performance of the funds during crisis than during market stability. Further, from table 'd' in the appendix, it is possible to see that not only the crisis negatively impacted the funds, but that their losses were heavier than the ones suffered by the market index. With exception of Jensen's alpha, all the other dummy's coefficients are not statistically different from zero, suggesting that the funds' exposure to the other factors (market index, SMB, HML, MWL, CMA, RMW) is not influenced by the presence of a market crisis.

On the other hand, during market stability the funds performed either equally to or better than the market, with low-ESG ones (ESG<55) being able to reach a statistically positive alpha. Most of the other variables keep the same trend of the previous regression, with Beta decreasing at the increment of the ESG score and with the SMB split between positive values for low ESG scores and negative ones for the more 'social' groups.

Finally, after the introduction of the dummy variable, the explanatory power of the model results in similar (and sometimes higher) to previous regressions and the F-test is always lower than 2.2e-16.

However, in order to truly understand whether there is a statistical difference in the funds' behaviour between market crisis and market stability, we will perform a hypothesis test. This will test the null hypothesis that the sum of all the dummy variables is equal to zero. The acceptance of the alternative hypothesis (i.e. that the sum of the dummy variables is not zero) would indicate the presence of a different behavioural path followed by the funds during times of crisis. The following table summarize the p-values of the test results for the three different regression models used.

Table 4.18

Statistical Test for Dummy Variables (p-values)

	Two factor	Three factor	Four factor	Five Factor
Tot Portfolio	0.0617	0.9027	0.2824	0.5032
ESG<55	0.9238	0.6278	0.2386	0.4254
ESG<70	0.5287	0.9135	0.5446	0.9755
70<ESG<75	0.7239	0.5468	0.367	0.9862
ESG>75	0.5456	0.2736	0.1934	0.7536
ESG>80	0.5087	0.2905	0.2609	0.6429

As we can see from table 4.18, the p-value is always higher than 0.05 and therefore we fail to refuse the null hypothesis. This means that even though the regressions show differences in the funds' behaviour between crisis and non-crisis times, these differences are not strong enough to determine a statistically different behaviour in the regressions' coefficients.

Regression: Since 2011

Table 4.19

Since 2011

	2 factor		
	Intercept	*MSCI world excess*	*Adjusted R^2*
TOT Portfolio	0.00078	0.7707 ***	78.3%
ESG < 55	0.00180	0.8105 ***	77.6%
ESG <70	0.00085	0.6704 ***	80.7%
70 < ESG < 75	0.00131	0.6164 ***	73.9%
ESG > 75	0.00060	0.6110 ***	63.4%
ESG > 80	0.00092	0.5657 ***	53.5%

	3 factor				
	Intercept	*MSCI world excess*	*SMB*	*HML*	*Adjusted R^2*
TOT Portfolio	0.00068	0.7720 ***	-0.03144	-0.11332	78.2%
ESG < 55	0.00169	0.8340 ***	0.4660 ***	-0.03480	80.6%
ESG <70	0.00068	0.6845 ***	0.2189 **	-0.138 *	82.0%
70 < ESG < 75	0.00120	0.6137 ***	-0.12255	-0.13272	74.6%
ESG > 75	0.00063	0.5990 ***	-0.2468 *	-0.00329	64.0%
ESG > 80	0.00099	0.5541 ***	-0.21889	0.03565	53.4%

	4 factor					
	Intercept	*MSCI world excess*	*SMB*	*HML*	*MWL (MOM)*	*Adjusted R^2*
TOT Portfolio	0.00081	0.7685 ***	-0.03351	-0.12659	-0.01914	77.9%
ESG < 55	0.00143	0.8412 ***	0.4702 ***	-0.00761	0.03924	80.4%
ESG <70	0.00082	0.6807 ***	0.2167 **	-0.153 *	-0.02116	81.8%
70 < ESG < 75	0.00141	0.6082 ***	-0.12583	-0.15372	-0.03031	74.4%
ESG > 75	0.00099	0.5893 ***	-0.2526 *	-0.04051	-0.05371	63.7%
ESG > 80	0.00165	0.5362 ***	-0.22960	-0.03294	-0.09900	53.5%

	5 factors						
	Intercept	*MSCI world excess*	*SMB*	*HML*	*RMW*	*CMA*	*Adjusted R^2*
TOT Portfolio	0.00234	0.6940 ***	-0.21022	-0.13981	-0.3730 *	-0.28598	79.4%
ESG < 55	0.0030 *	0.7839 ***	0.3422 **	-0.19461	-0.3726 *	0.08044	80.9%
ESG <70	0.00159	0.6388 ***	0.11700	-0.10919	-0.17731	-0.24936	82.5%
70 < ESG < 75	0.00233	0.5564 ***	-0.2498 **	-0.08516	-0.21195	-0.3343 *	75.7%
ESG > 75	0.00233	0.5122 ***	-0.43926 ***	0.07449	-0.31567	-0.5169 **	66.6%
ESG > 80	0.00259	0.4655 ***	-0.4088 **	0.21843	-0.22611	-0.719 ***	57.5%

Table 4.20

Intercepts p-value

	2 factor	3 factor	4 factor	5 factor
Tot Portfolio	0.608	0.656	0.616	0.1586
ESG<55	0.274	0.27004	0.37537	0.0810*
ESG<70	0.494	0.5683	0.5134	0.223
70<ESG<75	0.344	0.378	0.328	0.1162
ESG>75	0.733	0.7156	0.5872	0.20837
ESG>80	0.643	0.619	0.428	0.22010

The analysis of this last time-frame is made with the purpose of testing whether the efficiency of the social market increased in recent years, when more funds approached the market and more SRI-related studies where published. This hypothesis seems to be partially confirmed by the results. The alphas for all the groups are positive and have a significantly lower standard error than the ones in the regression ‘since 2006’. Additionally, one of the groups, the ESG<55, has a statically relevant and positive alpha in the 5-factor model. This result is in line with our observations in the descriptive analysis where, during the period from 2011 to 2018, ESG<55 gained higher returns than the market index. However, the majority of the groups shows a non-negative, non-relevant alpha.

Furthermore, the reduction of the time-frame led to a loss in the explanatory power of the models (the adjusted R^2), that however remains high enough to provide a realistic picture of how the funds performed and how they have been influenced by the independent variables. Finally, the Beta keeps its inverse relationship with the ESG score and all the other factors show trends similar to the ones observed in the previous regressions, with the SMB gaining in relevance.

Even though there are some differences between the regressions whose timeframe starts in 2006 and the ones whose timeframe starts in 2011, we cannot yet conclude that there has been a structural change in the funds’ profitability after 2011. In fact, to determine if 2011 has been a point of structural break we need to perform a Chow test, that studies whether the true coefficients of two linear regressions are equal or statistically different. The null hypothesis is that there is no statistical difference in the value of the coefficients of the two regressions, while the alternative hypothesis states the presence of statistically different coefficients. The specification of the test has been explained in the methodology chapter; the following table summarizes its results:

Table 4.219

Chow Test p-value

	2 factor	*3 factor*	*4 factor*	*5 factor*
Tot Portfolio	0.7478	0.4048	0.5071	0.2627
ESG<55	0.0136	0.0037	0.0058	0.0767
ESG<70	0.0141	0.0169	0.0236	0.0181
70<ESG<75	0.0013	0.0014	0.0024	0.0006
ESG>75	0.0164	0.0238	0.0372	0.0057
ESG>80	0.0030	0.0045	0.0062	0.0005

Except for the Total Portfolio, the p-values of the sub-groups are always lower than 0.05 (5%). Thus, for these groups we reject the null hypothesis in favour to the alternative one, which identifies 2011 as a point of structural break for the regression. The reason that might be at the root of the difference between the p-value of the overall sample (total portfolio) and the ones of its sub-groups can be found in figure 4.8 of the descriptive analysis. Indeed, here it is possible to observe that, while in the first part of the time-frame the performances of the groups were mixed, from 2011 groups with high and low ESG scores took different paths. Thus, it is possible that the overall portfolio is not able to catch the behavioural change because, by pooling together all the funds in one cluster, the performance of low and high ESG-score groups off-sets each other's.

To conclude, it is important to remember that the Chow test only highlights the presence of a change – without stating whether this change is positive or negative. However, the increment of statistically positive alphas in the latter regression suggests an increment in the funds' risk-adjusted performances in recent years.

4.2.2 Geographical Focus Groups

4.2.2.1 Descriptive Analysis

The second part of our analysis will focus on the performance of social funds investing in different geographical areas. As for the previous part, we first conduct a descriptive analysis. The purpose of this section is to have a general insight on the composition and on the graphical behaviour of the groups we will regress, so to have more tools to understand the regression results. The following table summarizes the main statistics for our samples:

Table 4.22

	Global	**North America**	**EU**	**Noridc**	Australia	Asia	South America	South Africa
ESG Average	69.72	65.9	73.74	69.41	71.76	65.75	66.43	65.09
Rolling Average	38.70%	58.30%	33.20%	71.30%	32.50%	42.10%	47.80%	40.20%
St. Dev. Rolling	2.60%	3.10%	2.80%	3.80%	1.60%	2.30%	4.10%	3.30%
N	176	58	109	18	2	6	4	2

From the table we can see that there are few differences in ESG score among the countries, as it varies from a maximum of 73 to a minimum of 65 points, with Europe leading the rank of social commitment. From what observed in Table 4.22 and if we consider only the groups we will use in the analysis, i.e. Global, North America, EU and Nordic, there is not a clear correlation between ESG scores and Rolling Performance. For example, Nordic funds are able to gain both an above-average ESG score and rolling performance. Differently, the standard deviation keeps its trend of decreasing at the increase of the ESG score.

The following graph and table summarize the exposure of the groups to small, medium and large market capitalization.

Table 4.23

	Market Capitalization Exposure		
	Small	*Medium*	*Large*
Global	5.70%	18.28%	76.02%
EU	8.51%	22.79%	68.70%
North America	5.67%	23.43%	70.90%
Nordic	4.41%	14.22%	81.37%
Asia	1.23%	17.75%	81.02%
South America	3.20%	22.98%	73.83%
South Africa	7.88%	16.29%	75.82%
Australia	16.99%	39.77%	43.24%

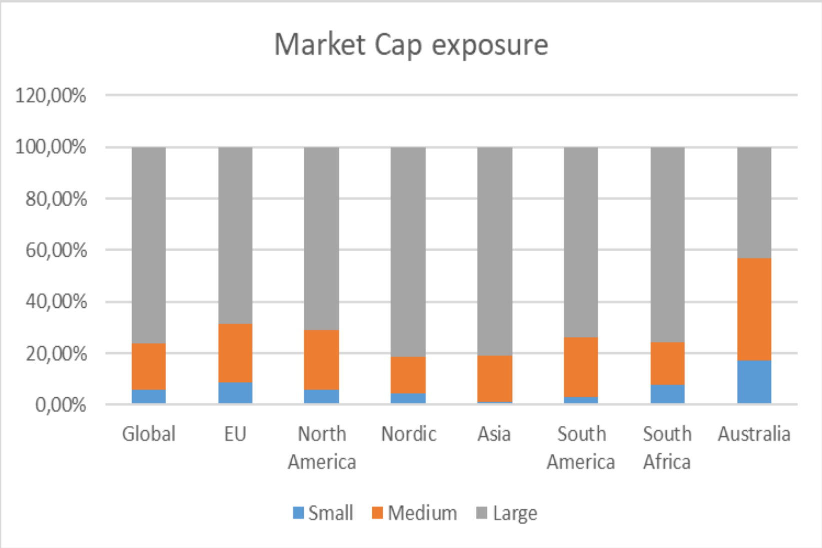


Figure 4.7

Differently from the previous analysis, when Geographical focus is used as discriminator, the resulting groups do not differ substantially in their exposure to market capitalization. Further, the positive relation between high social scores and high exposure to large capitalization stock seems to be inverted, with Europe having both the highest ESG score and the highest exposure to small stock. This apparent inversion of trend could be due to differences in the local market or in the local state incentives.

To better understand the trend the groups followed over the timeframe, we plot the normalized (in base 100) performance of the groups against the market index. Further, a visual identification of the crisis and non-crisis periods is included. In the Appendix (figures 'g-l') it is possible to find four graphs where the groups are plotted individually against the market index.

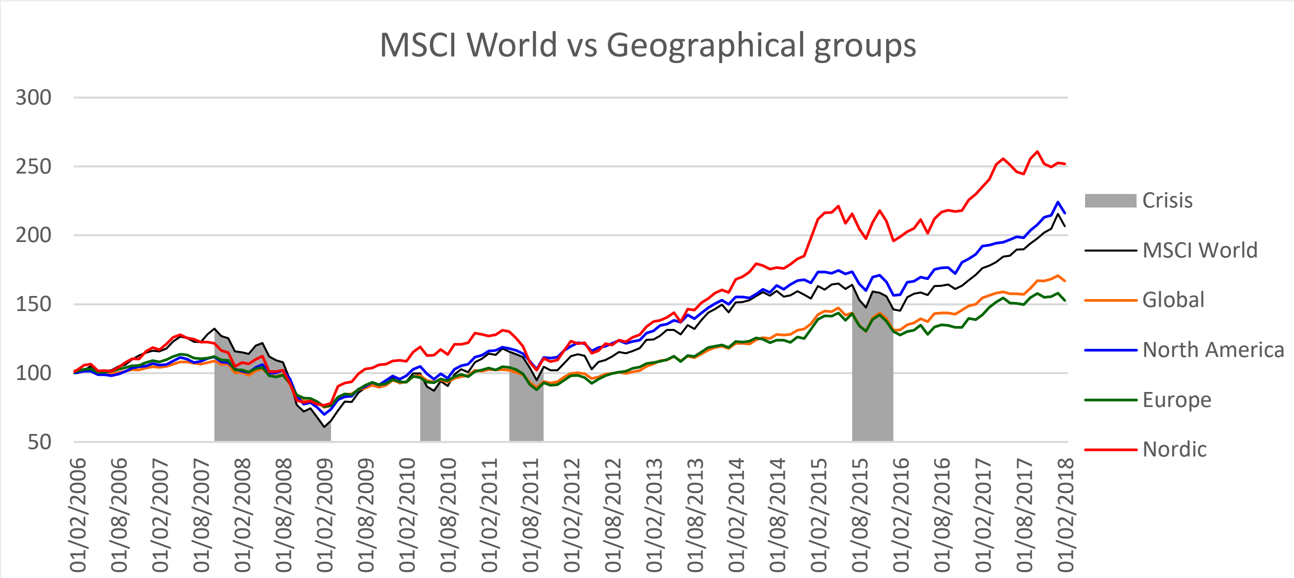


Figure 4.8

As we can notice, there are important differences among the groups' performance. Global and European funds show a volatility lower than the market during crises, a fact that can be seen clearly during the 2007 crisis, but in recent years they follow a path similar to the one of MSCI world. Their performance resembles the ones of high ESG-score groups that we saw in the previous analysis, a result consistent with the composition of EU and Global groups, which are the ones with the highest social scores.

Funds focusing their investments in North America and Nordic countries seem, especially in recent years, to be gaining higher returns than the market. These funds are still heavily affected by crisis, with losses comparable to the ones of the other groups, but Nordic countries in 2009 and North America in 2010 have been able to grow faster than the market. Moreover, Nordic funds seem to be affected by a lower volatility of returns during the 2007 crisis. However, it is not clear whether North American and Nordic funds performed better due to the whole local market growing, or because social investing strategies were effective. Therefore, in the next graph we plot three local MSCI indices and the world MSCI index (normalized in base 100) against each other, to see if also the local markets performed differently.

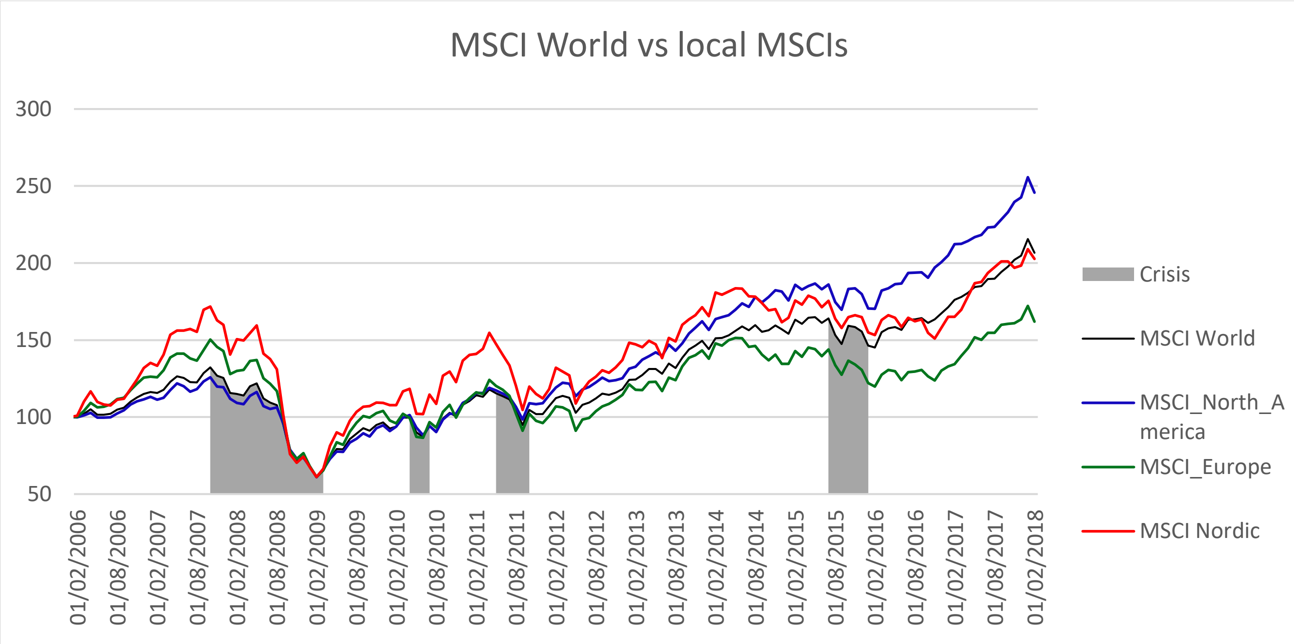


Figure 4.9

This graph shows two main trends: from 2006 to 2011 the local indices of North America and Europe followed a path very similar to the World index, while the Nordic MSCI gained higher returns. After 2011 Europe constantly underperformed the MSCI World, North America gained higher returns and Nordic countries followed a similar path. The high performance of MSCI North America suggests that the local funds performed better than the global index not because of a specific advantage given by the social policy, but thanks to a favourable time for the local market.

Differently, the poor performance of the MSCI Nordic in recent years suggests that Nordic funds were able to maximize their performances despite a declining local market. Further analysis can be found in the appendix (figures 'm-o'), where we plot each group against its local market index. From those graphs, it is possible to see even clearer that North America's excess returns are highly correlated to the ones of MSCI North America, while Nordic funds produce results constantly higher than the ones of MSCI Nordic. In the next part, we first analyse the groups with the factor models and in the timeframes utilized in the previous regression; then, the analysis is repeated with the addition of a local factor – defined as the difference between the local MSCI and the MSCI AC world.

4.2.2.2 Regression Analysis

Statistical tests on the validity of the regression

As we have done for the previous regression, we have first to test that the results we are obtaining from the analysis will be statistically valid. In general, we will follow the same rules explained in the previous section; however, we have to run the multicollinearity test with the introduction of the local factor again:

Table 4.2410

VIF score Europe

	MSCI_World	*MSCI_Europe*	*SMB*	*HML*	*MWL*	*RMW*	*CMA*
2 factor + local	1.327897	1.327897					
3 factor + local	1.360724	1.341177	1.008493	1.043074			
4 factor + local	1.444154	1.345764	1.012083	1.201322	1.323683		
5 factor + local	1.941430	1.351921	1.160890	1.815710		1.659743	1.847950

VIF score North America

	MSCI_World	*MSCI_North_America*	*SMB*	*HML*	*MWL*	*RMW*	*CMA*
2 factor + local	1.1568	1.1568					
3 factor + local	1.1950	1.1850	1.0260	1.0395			
4 factor + local	1.2945	1.1853	1.0287	1.2000	1.3195		
5 factor + local	1.7578	1.2205	1.1790	1.8219		1.6688	1.8732

VIF score Nordic

	MSCI_World	*MSCI_Nordic*	*SMB*	*HML*	*MWL*	*RMW*	*CMA*
2 factor + local	1.3702	1.3702					
3 factor + local	1.4576	1.4607	1.0522	1.0572			
4 factor + local	1.5339	1.4681	1.0537	1.2262	1.3259		
5 factor + local	1.9779	1.4811	1.1970	1.7986		1.6545	1.8653

As we can see, all the VIF scores are lower than 2; therefore, the multicollinearity is not strong enough to influence the validity of the regression results.

Regression 'since 2006'

Table 4.25

Since 2006

2 factor

	Intercept		*MSCI world*		*Local factor*	*Adj R^2*	
	No local	*With local*	*No local*	*With local*		*No local*	*With local*
Global	0,0004	/	0,5077***	/	/	74%	/
Europe	-0,0003	0,0003	0,5424***	0,4904***	0,26700***	71%	73%
North_America	0,0015	0,0001	0,6717***	0,7485***	0,9689***	88%	95%
Nordic	0,0027	0,0031*	0,6886***	0,5392***	0,43555***	62%	70%

3 factor

	Intercept		*MSCI world*		*SMB*		*HML*		*Local Factor*	*Adj R^2*	
	No local	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*		*No local*	*With local*
Global	0,0003	/	0,5169***	/	0,1110	/	-0,1413*	/	/	75%	/
Europe	-0,0003	0,0002	0,5471***	0,4962***	-0,0949	-0,0716	-0,0508	-0,0650	0,2657***	71%	73%
North_America	0,0015	0,0000	0,6713***	0,7505***	0,0967	0,1953***	-0,0079	-0,0037	1,017***	88%	96%
Nordic	0,0026	0,0030*	0,6988***	0,5537***	0,29475**	0,1359	-0,1803	-0,1018	0,4101***	64%	70%

4 factor

	Intercept		*MSCI world*		*SMB*		*HML*		*MWL*		*Local Factor*	*Adj R^2*	
	No local	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*		*No local*	*With local*
Global	0,0004	/	0,5116***	/	0,1082	/	-0,15858*	/	-0,0238	/	/	75%	/
Europe	-0,0002	0,0004	0,5387***	0,4901***	-0,0994	-0,0755	-0,0785	-0,0870	-0,0382	-0,0306	0,2620***	71%	73%
North_America	0,00166*	0,0002	0,6634***	0,7419***	0,0925	0,1908***	-0,0336	-0,0327	-0,0354	-0,0399**	1,0186***	88%	96%
Nordic	0,0032	0,0035*	0,6688***	0,5323***	0,27860**	0,1264	-0,2785**	-0,1855	-0,1353**	-0,1127*	0,4008***	65%	71%

5 factor

	Intercept		*MSCI world*		*SMB*		*HML*		*RMW*		*CMA*		*Local Factor*	*Adj R^2*	
	No local	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*		*No local*	*With local*
Global	0,0011	/	0,4794***	/	0,0579	/	-0,0930	/	-0,1187	/	-0,1965	/	/	76%	/
Europe	0,0004	0,0011	0,5187***	0,4690***	-0,1448	-0,1223	-0,0545	-0,0874	-0,1755	-0,2034	-0,1161	-0,0931	0,2692***	71%	73%
North_America	0,00194*	0,0003	0,6753***	0,7422***	0,0781	0,1744***	-0,1140	-0,0323	-0,20191*	-0,10536*	0,1024	-0,0088	1,0064***	88%	96%
Nordic	0,0034	0,0036*	0,6631***	0,5341***	0,2401	0,0961	-0,1508	-0,1258	-0,1484	-0,1684	-0,1740	-0,0648	0,4097***	63%	70%

Table 4.26

Intercepts p-value

	2 factor		*3 factor*		*4 factor*		*5 factor*	
	No local	*With local*	*No local*	*With local*	*No local*	*With local*	*No local*	*With local*
Global	0.711	/	0.7676	/	0.6990	/	0.3627	/
Europe	0.804	0.83954	0.794	0.85967	0.896	0.78395	0.755	0.44492
North_America	0.102	0.908	0.105	0.965	0.0768	0.7837	0.0522	0.5631
Nordic	0.169	0.0832	0.1830	0.0926	0.1017	0.0509	0.112	0.0614

The regression without the local factor shows a good explanatory power for all the groups, with the adjusted R^2 fluctuating from a minimum of 62% (for Nordic Countries) to a maximum of 86% (North America). Among the groups, only the funds focusing their investments in North America are able to gain statistically higher returns than the market index, with an alpha different from zero at the 10% confidence level (and a p-value equal to 0,0768 in the four-factor model and 0,0522 in the five-factor model). The Beta seems to be mostly stable across the sample, with North America and Nordic groups being more exposed to systematic risk than Europe and Global ones.

The remaining regressions factors are rarely statistically different from zero, a result that might be grounded in the differences between grouping by ESG score and by geographical focus. Indeed, while ESG score tends to cluster funds with similar strategies, geographical focus creates more heterogenous groups that are less affected by single strategies (since they might be composed by funds with opposite strategies that balance each other out). For example, the low relevance of the SMB was foreseeable from the descriptive analysis, where all the groups had similar (and balanced) exposure to small, medium and large stocks (differently from the ESG groups where high scores were associated to different market capitalization exposure). These results suggest that, while investment strategies as buying shares of low, weak and aggressive firms are common among all of the social funds (since HML, RMW, CMA are similar in most of the groups no matter the clustering process), the funds exposure to market capitalization and to systematic risk is influenced by both their geographical focus and their ESG score.

The introduction of the local factor in the regressions leads to an increment of the explanatory power of the analysis, with the adjusted R^2 ranging between 70% (Nordic) to 96% (North America). The rejection of the null hypothesis for all the confidence levels shows the relevance of local factors in the explanation of fund returns and confirms Cortez findings (Cortez, et al., 2011). The relationship between funds and local index is particularly strong in North America, where the increment of one point in the local MSCI is associated to an increment of 1,18 points in the associated group of funds.

Additionally, the presence of local indices in the regression has different implications on the alpha values. On one side, the superior returns gained from North American funds are now mostly explained by the local index and therefore no alpha is statistically relevant, since the cause of higher performance has been added to the equation. On the other side, the presence of the local index helps to explain the returns of Nordic funds during downturn periods, when the funds followed a course similar to that of the local index, but not during growth ones (when the funds overperformed both the local and the global indices). This leads to a positive and statistically relevant alpha for Nordic funds, no matter which regression model is adopted. These results are in line with what we observed in the descriptive analysis, where North America and MSCI North America seem to follow a similar path, while Nordic countries were disjointed from their local index. Finally, all the other variables gain in statistical relevance and keep the trends shown in the previous regression.

We will now run the same analysis but setting 2011 as starting point of our time-frame, so to understand whether these relationships changed through time.

Regression 'since 2011'

Table 4.27

Since 2011

2 factor

	Intercept		MSCI world		Local factor	Adj R^2	
	No local	With local	No local	With local		No local	With local
Global	0.0012		0.5842***			68,5%	
Europe	-0.0002	0.0016	0.6591***	0.5578***	0.4161***	63,1%	67,6%
North_America	0.0022**	-0.0004	0.7133***	0.8059***	0.9986***	86,3%	94,7%
Nordic	0.0022	0.0042*	0.7572***	0.6387***	0.4285***	54,8%	61,7%

3 factor

	Intercept		MSCI world		SMB		HML		Local Factor	Adj R^2	
	No local	With local	No local	With local	No local	With local	No local	With local		No local	With local
Global	0.0011		0.5886***		0.0292		-0.12135			68,4%	
Europe	-0.0003	0.0015	0.6501***	0.5512***	-0.228	-0.1901	-0.0852	-0.1212	0.4172***	63,6%	68,1%
North_America	0.0021*	-0.0005	0.7208***	0.8205***	0.1180	0.2040***	-0.0711	-0.0303	1.04***	86,4%	95,5%
Nordic	0.0019	0.0039	0.765***	0.6463***	0.0387	-0.0133	0.2371	-0.1596	0.4131***	54,9%	61,4%

4 factor

	Intercept		MSCI world		SMB		HML		MWL		Local Factor	Adj R^2	
	No local	With local	No local	With local	No local	With local	No local	With local	No local	With local		No local	With local
Global	0.0012		0.5848***		0.0269		-0.1358		-0.0208			68,1%	
Europe	0,0000	0.0016	0,642***	0.5474***	-0,2334	-0.1930	-0,1165	-0.1379	-0,0450	-0.0243	0.4148***	63,3%	67,8%
North_America	0.0022*	-0.0004	0.7179***	0.8164***	0.1162	0.2016***	-0.0823	-0.0465	-0.0160	-0.0233	1.041***	86,3%	95,5%
Nordic	0.0028	0.0048*	0.7425***	0.6235***	0.0253	-0.027	-0.3233*	-0.2465	-0.1243	-0.1255	0.4134***	55,0%	61,5%

5 factor

	Intercept		MSCI world		SMB		HML		RMW		CMA		Local Factor	Adj R^2	
	No local	With local	No local	With local	No local	With local	No local	With local	No local	With local	No local	With local		No local	With local
Global	0.0023		0.5259***		-0.1092		-0.0571		-0.2204		-0.388*			69,9%	
Europe	0.0013	0.0031	0.5679***	0.4712***	-0.4096**	-0.3774**	0.0015	-0.0809	-0.2862	-0.3343	-0.5135**	-0.4292*	0.4046***	65,6%	69,9%
North_America	0.0029**	0,0000	0.6879***	0,7955***	0.0364	0,1469**	-0.1814*	-0,0322	-0.2494*	-0,1142	0.0627	-0,1047	1,042***	86,6%	95,7%
Nordic	0.0043	0.0059**	0.6528***	0.5522***	-0.2153	-0.2349	-0.2245	-0.1460	-0.4895	-0.4264	-0.5052	-0.4560	0.3961***	56,6%	62,6%

Table 4.28

Intercepts p-value								
	2 factor		3 factor		4 factor		5 factor	
	No local	With local	No local	With local	No local	With local	No local	With local
Global	0.429	/	0.478	/	0.446	/	0.160	/
Europe	0.904	0.389	0.885	0.409	0.988	0.385	0.516	0.113
North_America	0.044	0.592	0.052	0.401	0.052	0.562	0.014	0.975
Nordic	0.396	0.088	0.447	0.109	0.302	0.064	0.129	0.027

The relationships between the funds and the factors change slightly when a shorter period of time is considered. The risk premium's Betas show an increment in value, suggesting an increment in the exposure to systematic risk by social funds in recent years. However, the values of the Betas remain lower than 1, confirming the lower volatility and risk exposure characterizing social funds when compared to the market index.

Jensen's alphas are relevant for North America when the local factor is not included and for Nordic funds when the local factor is included, a result that does not seem to be dependent on the regression timeframe. However, a difference from the previous regression is given by the values of the alphas that, when relevant, are constantly higher than the ones calculated over 12 years of returns. Europe results to be more influenced by its local index, with the value of the local factor increasing from 0.26 (when the whole period is analysed) to 0.41 (when only the last 7 years of data is taken in consideration). However, its alpha never results to be statistically relevant. North America keeps the trends showed in the previous regression, with a high degree of explanatory power (95.5%) and a strong dependency with its local MSCI. As previously observed, the introduction of the local index in the regression leads to a significant loss in the statistical relevance of the alphas, since the superior returns of the funds are now explained by the local factor.

Finally, the introduction of the local factor increases the adjusted R^2 and affects both the Betas (that are lower for Europe and Nordic funds but higher for North America) and the Jensen's alphas (that become statistically not relevant for North America and statistically different from zero for Nordic funds) – results that are in line with the ones obtained from the previous regression. All the other variables keep the main trend already observed in the previous regressions, with SMB having mixed results and with negative values for HML, RMW, WML and CMA.

As we did before, we perform a Chow Test to investigate whether 2011 represented a point of structural break for the regressions’ factors.

Table 4.2911

Chow Test p-value

	2 factor	*3 factor*	*4 factor*	*5 factor*
Global	0.0103	0.0224	0.0361	0.0083
North_America	0.0188	0.0009	0.0007	0.0026
Europe	0.0014	0.0015	0.0026	0.0006
Nordic	0.0802	0.0867	0.0937	0.0427

The p-values of the Chow test are always lower than 0.05 (5%). Thus, we accept the alternative hypothesis that 2011 is a point of structural break of the regression. This implies that the coefficients of the regressions before and after 2011 are statistically different from each other. However, it is important to remember that the Chow test only makes explicit the presence of a change – without stating whether this change is positive or negative. Nonetheless, the increment of statistically positive alphas when changing the time-frame from “2006” to “2011” could indicate an increment in the funds’ performances in recent years thanks to a more mature and organized social market.

We will now analyse how the groups performed during crisis and non-crisis periods, through the use of a set of dummy variables. In this last part of analysis, we will not distinguish between models with and without the local factor, as this will be used as default.

Regression 'Crisis'

Table 4.3012

Dummy/Non Crisis

2 factor

	Intercept		*MSCI world excess*		*Local factor*		*Adjusted R^2*
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	/
Global	-0,0042*	0,0022	0,0671	0,4506***	/	/	0,7479
Europe	-0,004*	0,0020	0,0742	0,4458***	-0,3127*	0,3693***	0,7315
North America	-0,0001	0,0000	-0,0054	0,7518***	-0,0075	0,9714***	0,9516
Nordic	-0,0044	0,0051**	0,0760	0,4635***	0,1010	0,403***	0,7024

3 factor

	Intercept		*MSCI world excess*		*SMB*		*HML*		*Local factor*		*Adjusted R^2*
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	/
Global	-0,0052*	0,0022	0,0723	0,4579***	0,0437	0,0814	-0,1697	-0,0923	/	/	0,7563
Europe	-0,0057*	0,0020	0,0919	0,4484***	-0,1345	-0,0328	-0,1548	-0,0469	-0,36263*	0,3746	0,7314
North America	0,0003	-0,0001	-0,0130	0,758***	-0,0804	0,2294***	0,0550	-0,0232	-0,0421	1,027***	0,9585
Nordic	-0,0047	0,0051**	0,0822	0,4762***	0,0072	0,1158	-0,0290	-0,0971	0,0770	0,3887***	0,6975

4 factor

	Intercept		*MSCI world excess*		*SMB*		*HML*		*MWL (MOM)*		*Local factor*		*Adjusted R^2*
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	/
Global	-0,0061**	0,0021	0,0468	0,458***	0,0249	0,0815	-0,2470	-0,0917	-0,0763	0,0008	/	/	0,7573
Europe	-0,00682**	0,0025	0,0751	0,4445***	-0,1411	-0,0383	-0,1745	-0,0812	-0,0025	-0,0517	-0,3372*	0,3626***	0,7305
North America	-0,0001	0,0000	-0,0309	0,7551***	-0,0942	0,2277***	0,0163	-0,0408	-0,0316	-0,0257	-0,0650	1,032***	0,9599
Nordic	-0,0075*	0,006**	0,0602	0,461***	0,0042	0,1057	-0,1230	-0,1691	-0,0511	-0,1045	0,0302	0,3929***	0,7041

5 factors

	Intercept		*MSCI world excess*		*SMB*		*HML*		*RMW*		*CMA*		*Local factor*		*Adjusted R^2*
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	/
Global	-0,00495*	0,0024*	0,0606	0,4435***	0,0333	0,0524	-0,1465	-0,0630	0,0336	-0,0887	0,0575	-0,1718	/	/	0,7525
Europe	-0,0055	0,0025	0,1180	0,429***	-0,0850	-0,0773	-0,2068	-0,0503	0,0232	-0,1607	0,1982	-0,1465	-0,3636*	0,3698***	0,7269
North America	0,0012	0,0001	-0,0149	0,7489***	-0,0797	0,2089***	0,0368	-0,0286	-0,1004	-0,0777	0,0437	-0,0605	-0,0596	1,03***	0,9585
Nordic	-0,0043	0,0056**	0,1166	0,4591***	0,0677	0,0756	-0,1075	-0,1233	-0,0219	-0,1609	0,2139	-0,0840	0,0897	0,3858***	0,6907

Table 4.31

Intercepts p-value								
	2 factor		*3 factor*		*4 factor*		*5 factor*	
	Dummy	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*	*Dummy*	*Non-Crisis*
Global	0.0922*	0.1239	0.0401	0.1162	0.0207**	0.1426	0.0984*	0.0925*
Europe	0.0978*	0.2141	0.0514	0.2016	0.02704**	0.1500	0.1145	0.1421
North_America	0.9630	0.9510	0.7690	0.8110	0.6710	0.9550	0.4031	0.9326
Nordic	0.2754	0.0251**	0.2553	0.0274**	0.0806*	0.0143**	0.3749	0.0237

During non-market crisis, the regression shows similar results to the ones observed when no distinction between crisis and non-crisis periods is made, a result that was foreseeable since the majority of observations are from non-crisis periods (105) rather than during crisis ones (41).

Thus, during times of market stability the funds are strongly influenced by both the local and the global index and their alphas are generally positive but not relevant, except for Nordic funds and Global funds that are able to gain positive and statistically relevant alphas. However, differently from what we expected, the dummy variable shows a negative relationship between the presence of a market crisis and the performance of social funds, with all the dummy's alphas being negative and most of them being statistically relevant. Thus, social funds perform better during market stability than during market crisis.

While the dummy variables in table 4.30 show how differently the funds performed during market crisis and market stability, the alphas and the coefficients in table 'e' in the Appendix show the absolute performance of social funds during the times of market crisis. From this table, it is possible to see that not only did the crisis impact the funds negatively, but that their losses were heavier than the ones suffered by the market index. This results in having negative alphas for all the groups, with Europe and Global ones being statistically relevant.

Also, it is interesting to note the presence of a statistically negative coefficient for the dummy variable when multiplied with MSCI Europe – indicating an inverse relationship between European funds and their local index during times of market instability. Most of the other factors show little or no statistical relevance when the dummy variable is introduced, a result that is in line with what we found in the "ESG crisis regression" and that shows no or little influence of the crisis on the fund's exposure to the main factors (market index, HML, SMB, RMW, CMA, MWL).

Finally, as we have done in the previous chapter, we run a series of hypothesis tests testing the null hypothesis that the sum of all the dummy variables in the regressions is equal to zero. The purpose of this test is to determine whether there is a statistical difference in funds' behaviour between crisis and non-crisis times. The following table summarizes the p-values of the test:

Table 4.32

Statistical Test for Dummy Variables (p-values)

	Two	Three factor	Four factor	Five Factor
Global	0.2528	0.7801	0.3416	0.9481
Europe	0.8661	0.2196	0.4003	0.8665
North America	0.989	0.3434	0.9061	0.9666
Nordic	0.2894	0.5649	0.7723	0.6243

As we can see from the results, the p-value is always higher than 0,05 and therefore we cannot accept the alternative hypothesis. This means that even though the regressions show some differences in the performances between crisis and non-crisis times, these differences are not strong enough to have a statistical relevance.

5. Limitations and further research

Prior to presenting the conclusion of the thesis, it is important to describe some relevant limitations to our results. First of all, the results of the analysis are strictly related to the data we chose to utilize. The inclusion or exclusion of certain companies and funds rely on the ESG scores provided by Thomson Reuters, that might be affected by biases. As a matter of fact, Thomson Reuters calculations are based on the past activities of the firms and mainly rely on voluntary reports disclosed by the companies, a process that might require time to be updated (Lee & Moscardi, 2018). This might entail a mismatch between the updating of the score and the financial impact of the news on the company's stock, with the result of both having the wrong companies labelled as 'ethical' and failing to relate company performance with their social score.

Further, even though Thomson Reuters is acknowledged as a reliable database in providing financial and social data, it is not the only one on the market and the choice of a different database could result in reaching different conclusions about SRI performance. Indeed, the same company can be rated quite differently depending on the methodology used, subjective interpretation, or an individual agency's agenda, revealing biases ranging from market cap size to industry sector or location – all due to lack of uniform disclosure (Doyle, 2018). Finally, by combining a wide range of environmental, social and governance criteria into a single ESG score, there is the risk of opposing effects cancelling each other out (Escrig-Olmedo, et al., 2014).

A second restriction is given by the novelty of the social market itself. While it is possible that at the present day the market has not priced the social impact yet, it is realistic to assume that this will happen in the future. As this segment will increase in size, the market will reach a new equilibrium where, according to the market efficiency theory, investors will be required to discount their financial performance to gain utility from social impact. Therefore, the absence of differences between social and non-social investments, as we found, is unlikely to last, and the fast growth of the social market might speed up the pricing process of the social side of ethical stock.

Further, the representativeness of the sample might also constitute another bias. First of all, the analysis took into account only the companies and the funds that had a score assigned by Thomson Reuters. In addition, the lack of financial information on the database forced us to heavily screen the universe of companies and funds: while we screened 3.424 funds labelled as either ethical or social, only 10% (386) had all the financial information needed to run the analysis. Even though this is a more than satisfying amount when compared with previous studies (Friede, et al., 2015) and although the statistical tests helped us to reach conclusions that go beyond the sample level, it is difficult to say how results might change with full information on the whole database.

Finally, the results we reached through the regression analysis might be flawed. The factor models got several weaknesses (e.g. the assumption of the existence of a risk-free rate, unlimited borrowing and lending, no transaction costs in trading etc.) and the assumption that the variance is an adequate measure of risk implies that returns are normally distributed (or distributed in any two-parameter way). Even though we avoided normality assumption in the stock analysis (by using non-parametric tests), we were not able to find an alternative to factor models to decompose funds' strategies. Therefore, we were forced to the normality assumption in the second part of the analysis. Additionally, the presence of positive and relevant alphas might be due to other strategies – different from SRI – that we did not include among the regression factors.

We suggest further research to focus on niche industries or in more narrowly defined regions, where other valuable results might be found. Further, an analysis of the relation between performance and social impact in the private equity sector could be of great interest – as disadvantages deriving from a lack in diversification would be reduced and more opportunities for investors to buy or sell mispriced assets might arise. Studies on ESG performance during times of recession might yield valuable insights as well, as our findings do not support the argument that socially responsible investments perform better than their peers in a financial crisis. We also suggest future research to focus on the different components of ESG, as we believe it is possible that combined score calculations mix up factors affecting financial performance positively with factors that influence it negatively. Further, more information could arise by the creation of a social factor (by shorting low ESG stock and investing in high ESG stock) and then by studying how the Beta of this factor influences the returns and the Jensen' alpha of social companies.

6. Conclusion

This thesis aims to shed light on the financial performance of socially responsible investments over different timeframes and geographical areas. The analysis is made up of two sections, each studying a different aspect of SRI performance. The first section studies SRI at its core, i.e. without mixing it with other strategies. Here we investigate the differences in returns between socially responsible and socially non-responsible stocks, whose social scores are retrieved from the Thomson Reuters database. The groups are compared over two performance measurements: excess return and Sharpe ratio. For the former, we use the Augmented Dickey-Fuller test to examine whether the groups differ statistically from each other, while for the latter we create a confidence level for the difference between the groups' Sharpe ratios by using a circular paired block bootstrap.

The results show that the difference in the performance of social and non-social stocks has no statistical validity. The absence of a statistical impact of the ESG score on company performance is in line with about 20% of all the empirical studies (Friede, et al., 2015) and represents a point in favour of SRI. Indeed, when embracing ethical strategies, stockholders do not aim to gain superior results, but rather to avoid losses. Following the results, with the purchase of ethical stocks investors are able to derive a financial utility equal to that generated by non-ethical stocks – plus an additional utility derived by the social side of the investment. Therefore, on a stock level and without taking in consideration the constraints relative to portfolio creation and diversification, social stocks result to be more attractive than their non-social peers.

Additionally, even though it would be necessary to run further statistical tests, we find a negative correlation between ESG score and volatility – a result consistent with several other empirical studies (Porse, et al., 2017; Kumar, et al., 2016; Verheyden, et al., 2016; Chaudhry, et al., 2016) that strengthens the idea of SRI as a tool for qualitative screening. On the other side, the absence of higher risk-adjusted returns from non-social stocks is not in line with other academic studies, for which addressing resources to ESG-related issues leads to poorer performances (Mcwilliams & Siegel, 1997; Hillman & Keim, 2001; Brammer & Millington, 2008). Furthermore, the absence of superior performance from social stock goes against the findings of Eccles et al. (2014), who find SRI leading to measurable improvements within a company and thus to have a positive impact on its stock returns.

The second section analyses SRI when combined with other investing strategies, as it happens in ethical funds. By shifting the focus from the stock to the fund level we are able to study whether the issues arising from the creation of social portfolios (e.g. lack in diversification), affect the after-fee-returns of social funds. The aim is to understand if it is possible for a social investor to gain risk-adjusted returns after fees equal to that generated by a passive investment in a market index. To do so, we analyse the presence of positive and negative Jensen alphas by regressing the rolling performance of social funds against the market index 'MSCI world'. Further, in order to avoid other strategies (i.e. SMB, HML, WML, RMW and CML) to influence the size and sign of the alphas, we run the regression over four factor models, namely CAPM, Fama-French three-factor, Carhart four-factor and Fama-French five-factor. Finally, the regressions are run over different time-frames and the groups are sub-divided depending on their ESG score and geographical focus, in order to exploit potential niche inefficiencies.

The results show neither superior nor inferior performance for most of the groups and during most of the time-frames, with the majority of the alphas being not statistically different from zero. Furthermore, in contrast to our expectations, social funds underperformed the market index during crisis periods, with the alphas of 'Global', 'Europe' and of high-ESG score groups being statistically lower than zero. Even though the poor performance during market crisis goes against Becchetti's theory (Becchetti, et al., 2015), that sees SRI funds as an opportunity to insure investments from economic downturns, it does not confute its findings. Indeed, the conclusions of Becchetti et al. are based on the comparison between social and non-social funds, and do not account for the potential that all funds might underperform the market during periods of instability. Thus, the inadequacy of ethical funds for insurance purposes does not contrast with their ability to gain higher risk-adjusted returns than common funds which might be affected by the crisis even more severely.

On the other hand, when we set the beginning of the time-frame to 2011, we find a statistically relevant improvement in fund performance, suggesting an increment in the profitability of social investing in the years following the financial crisis. This result is strengthened by the Chow tests, that identifies a point of structural break for regression coefficients in 2011. Although other factors might have influenced this change, the consistency of the results across both groups and regression models suggests that it is the social factor that drives the superior performance of the funds. This change in trend might be due to an improvement in the efficiency of the social market that, thanks to the increment of the number of social funds and social stocks, is now able to contain the downturn effects of social investing, e.g. differentiation. Finally, in line with the overall literature, the local factor appears to be extremely relevant for the explanation of the results of socially responsible funds, although an explicit comparison of developed and emerging markets might yield more insights into regional differences (Friede, et al., 2015).

Contrary to the suggestion of Friede, et al. (Friede, et al., 2015), we cannot conclude that lack of diversification, management and trading costs of socially responsible funds leads to underperformance, since we find no evidence that the SRI fund performance is inferior to the market return benchmark. However, as our results also do not show that SRI funds perform statistically better than the market, the argument of Friede, et al. (Friede, et al., 2015), that opposing ESG effects might cancel each other out in an ESG-screened portfolio, is not weakened in our view.

To conclude, thanks to the evolution of the social investing infrastructure and to the constant increment of social stocks available on the market, SRI has become a competitive investment strategy. The absence of a statistical difference between the returns of social and common investments supports the view that overall the ESG dimension of an investment is not priced by the market (Eccles, et al., 2014). Assuming that social impact has a not estimable but positive effect on the utility curve of investors, the utility produced through SRI will be either equal or higher to that of common investments. Therefore, SRI represents an attractive strategy for any investor, especially for those that derive positive utility from social impact.

7. Bibliography

Auer, B. & Schuhmacher, F., 2016. Do socially (ir)responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance* , Volume 59, pp. 51-62.

Balia, T. G., Cakici, N. & Whitelaw, R. F., 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics,* 99(2), pp. 427-446 .

Barton, D., 2011. Capitalism for the Long Term. *Harvard Business Review,* Volume March.

Bauer, Koedijk & Otten, 2002. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance.*

Bauer, R., Derwall, J. & Otten, R., 2007. The Ethical Mutual Fund Performance Debate: New Evidence from Canada. *Journal of Business Ethics,* 70(2), pp. 111-124.

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

Becchetti, L., Ciciretti, R., Dalò, A. & Herzel, S. R., 2015. Socially responsible and conventional investment funds: performance comparison and the global financial crisis. *Applied Economics,* Volume February, pp. 1-22.

Bello, Z. Y., 2005. Socially Responsible Investing And Portfolio Creation. *Journal of Financial Research,* 28(1), pp. 41-57.

Bénabou, R. & Tirole, J., 2010. Individual and Corporate Social Responsibility. *Economica,* 77(305), pp. 1-19.

Bialkowski, J. & Starks, L. T., 2016. SRI Funds: Investor Demand, Exogenous Shocks and ESG Profiles. *Working paper, University of Canterbury, Department of Economics and Finance.*

Bloomberg , 2017a. *Best of Sustainable Finance Report,* s.l.: s.n.

Bloomberg , 2017b. *The growing role of ESG investing in portfolio management.* [Online]
Available at: https://www.bloomberg.com/professional/blog/growing-role-esg-investing-portfolio-management/
[Accessed July 2018].

Bollen, N. P. B., 2007. Mutual Fund Attributes and Investor Behavior. *The Journal of Financial and Quantitative Analysis,* 42(3), pp. 683-708.

Brammer, S. & Millington, A., 2008. Does It Pay to Be Different? An Analysis of the Relationship between Corporate Social and Financial Performance. *Strategic Management Journal,* 29(12), pp. 1325-1343.

Caballero, 2017. Small Sample Bias And Adjustment Costs. *MIT Economics.*

Caplan, L., Griswold, J. S. & Jarvis, W. F., 2013. *From SRI to ESG: The Changing World of Responsible Investing,* s.l.: Commonfund Institute.

Carhart, M. M., 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance,* 52(1), pp. 57-82

Chaudhry, K. et al., 2016. *ESG - Environmental, Social & Governance Investing: A Quantitative Perspective of how ESG can Enhance your Portfolio,* s.l.: J.P. Morgan.

Coleman, L., 2011. Losses from Failure of Stakeholder Sensitive Processes: Financial Consequences for Large US Companies from Breakdowns in Product, Environmental, and Accounting Standards. *Journal of Business Ethics,* 98(2), pp. 247-258.

Cortez, Silva & Areal, 2011. socially responsible investing in the global market: the performance of us and european funds. *International Journal of Finance & Economics.*

Davis Polk & Wardwell LLP, 2017. *ESG Reports and Ratings: What They Are, Why They Matter?,* s.l.: Davis Polk & Wardwell LLP.

Derwall, J., Koedijk, K. & Horst, J. T., 2011. A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance,* Volume 35, p. 2137–2147.

Diltz, J. D., 1995. Does Social Screening Affect Portfolio Performance?. *The Journal of Investing,* 4(1), pp. 64-69.

Dimson, E., Karakaş, O. & Li, X., 2015. Active Ownership. *Review of Financial Studies,* 28(12), pp. 3225-3268.

Doyle, T. M., 2018. *Ratings That Don't Rate - The Subjective World Of ESG Rating Agencies,* Washington D.C.: American Council For Capital Formation.

Dutt, T. & Humphery-Jenner, M., 2013. Stock return volatility, operating performance and stock returns: International evidence on drivers of the ‘low volatility’ anomaly Author links open overlay panel. *Journal of Banking & Finance,* 37(3), pp. 999-1017 .

Eccles, R. G., Ioannou, I. & Serafeim, G., 2014. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science,* 60(11), p. 2835–2857.

Eling, M. & Schuhmacher, F., 2007. Does the choice of performance measure influence the evaluation of hedge funds?. *Journal of Banking & Finance,* Volume 31, p. 2632–2647.

Elton, 1992. Optimal investment strategies with investor liabilities. *Journal of Banking and Finance*

Escrig-Olmedo, E., Muñoz-Torres, M., Fernández-Izquierdo, M. & Rivera-Lirio, J., 2014. Lights and shadows on sustainability rating scoring. *Review of Managerial Science,* 8(4), pp. 559-574.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance,* 25(2), pp. 383-417.

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

Fama & French, 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance.*

Fama & French, 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics.*

Fama & French, 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance.*

Fama & French, 2014. A five-factor asset pricing model. *Journal of Financial Economics.*

Flood, C., 2018. *Strong ESG policies are no protection against scandal.* [Online]
Available at: https://www.ft.com/content/08a3420e-05bf-11e8-9650-9c0ad2d7c5b5
[Accessed May 2018].

Fombrun, C. J., 2005. A World of Reputation Research, Analysis and Thinking — Building Corporate Reputation Through CSR Initiatives: Evolving Standards. *Corporate Reputation Review,* 8(1), pp. 7-12.

Fombrun, C. & Shanley, M., 1990. What's in a Name? Reputation Building and Corporate Strategy. *The Academy of Management Journal,* 33(2), pp. 233-258.

Freeman, R. E., Harrison, J. S. & Wicks, A. C., 2007. *Managing for stakeholders: survival, reputation, and success.* New Haven: Yale University Press.

French, K. R., 2018. *Data Library.* [Online]
Available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[Accessed 1st April 2018].

Freshfields Bruckhaus Deringer, 2005. *A legal framework for the integration of environmental, social and governance issues into institutional investment,* s.l.: United Nations Environmental Programme Finance Initiative.

Friede, G., Busch, T. & Bassen, A., 2015. ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment,* 5(4), pp. 210-233.

Fulton, Kahn & Sharples, 2012. *Sustainable Investing: Establishing Long-Term Value and Performance,* s.l.: Deutsche Bank Group.

Garvey, G. T. et al., 2016. A Pitfall in Ethical Investing: ESG Disclosures Reveal Vulnerabilities, Not Virtues. *SSRN Electronic Journal.*

Girard, E., Rahman, H. & Stone, B., 2007. Socially Responsible Investments: Goody-Two-Shoes or Bad to the Bone?. *Journal of Investing,* 16(1), pp. 96-110.

Glode, V., 2011. Why mutual funds "underperform". *Journal of Financial Economics.*

Goldreyer, E. F. & Diltz, J. D., 1999. The performance of socially responsible mutual funds: incorporating sociopolitical information in portfolio selection. *Managerial Finance,* 25(1), pp. 23-36.

Gregory, A., Matatko, J. & Luther, R., 1997. Ethical Unit Trust Financial Performance: Small Company Effects and Fund Size Effects. *Journal of Business Finance & Accounting,* 24(5), pp. 705-725.

Gregory & Whittaker, 2007. Performance and Performance Persistence of 'Ethical' Unit Trusts in the UK. *Journal of Business Finance & Accounting.*

Grene, S. & Marriage, M., 2015. *VW investors ignored corporate governance warnings.* [Online] Available at: https://www.ft.com/content/6af071a8-6205-11e5-a28b-50226830d644 [Accessed 7th June 2018].

GSIA, 2017. *2016 Global Sustainable Investment Review,* s.l.: GSIA.

Halbritter, G. & Dorfleitner, G., 2015. The wages of social responsibility — where are they? A critical review of ESG investing. *Review of Financial Economics,* Volume 26, pp. 25-35.

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), pp. 62-66.

Hettmansperger. & McKean, 1998. *Robust nonparametric statistical methods.* s.l.:Kendall's Library of Statistics.

Hillman, A. J. & Keim, G. D., 2001. Shareholder Value, Stakeholder Management, and Social Issues: What's the Bottom Line?. *Strategic Management Journal,* 22(2), pp. 125-139.

Hurst & Simon, 2010. *The Characteristic Function of the Student-t Distribution.* [Online].

Jegadeesh & Titman, 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market. *The Journal of Finance.*

Jensen, M. C., 1968. The Performance Of Mutual Funds In The Period 1945-1964. *The Journal of Finance,* 23(2), pp. 389-416.

Jensen, M. C., 2002. Value Maximization, Stakeholder Theory, and the Corporate Objective Function. *Business Ethics Quarterly,* 12(2), pp. 235-256.

Khan, M., Serafeim, G. & Yoon, a. A., 2017. Corporate Sustainability: First Evidence on Materiality. *The Accounting Review,* 91(6), pp. 1697-1724.

Kosowski, 2011. Do mutual funds perform when it matters most to investors? US mutual fund performance. *The Quarterly Journal of Finance.*

Kreander, N., Gray, R., Power, D. & Sinclair, C., 2005. Evaluating the Performance of Ethical and Non-ethical Funds: A Matched Pair Analysis. *Journal of Business Finance & Accounting,* 32(7-8), pp. 1465-1493.

Kumar, N. C. A. et al., 2016. ESG factors and risk-adjusted performance: a new quantitative model. *Journal of Sustainable Finance & Investment,* 6(4), pp. 292-300.

Kwiatkowski & Philips, 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of Econometrics.*

Ledoit & Wolf, 2008. Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance.*

Lee, L.-E. & Moscardi, M., 2018. *2018 ESG Trends To Watch,* s.l.: MSCI.

Lefkovitz, D., 2018. *Education Emerging Markets Rank Highly in ESG Ratings.* [Online]
Available at: http://www.morningstar.co.uk/uk/news/166715/emerging-markets-rank-highly-in-esg-ratings.aspx
[Accessed 05 July 2018].

Mallin, C., Saadouni, B. & Briston, R., 1995. The Financial Performance Of Ethical Investment Funds. *Journal of Business Finance & Accounting,* 22(4), pp. 483-496.

Mann & Whitney, 1947. *On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other.* s.l.:The Annals of Mathematical Statistics.

Markowitz, H., 1952. Portfolio Selection. *Journal of Finance,* 7(1), pp. 77-91.

Mcwilliams, A. & Siegel, D. S., 1997. The Role of Money Managers in Assessing Corporate Social Responsibility. *The Journal of Investing,* 6(4), pp. 98-107 .

MSCI, 2017. *MSCI ESG Fund metrics methodology,* s.l.: s.n.

Nofsinger & Varma, 2014. Socially responsible funds and market crises. *Journal of Banking & Finance,.*

O'Toole, J. & Vogel, D., 2011. Two and a Half Cheers for Conscious Capitalism. *California Management Review,* 53(3), pp. 60-76.

Pagan & Sossounov, 2002. A Simple Framework for Analyzing Bull and Bear Markets. *Journal of Applied Econometrics.*

Porse, E. et al., 2017. *Cracking the ESG code (Nordea Equity Research),* s.l.: Nordea Bank.

PRI Association, 2018a. *About the PRI.* [Online] Available at: https://www.unpri.org/pri [Accessed July 2018].

PRI Association, 2018b. *United Nations Principles of Responsible Investment.* [Online] Available at: https://www.unpri.org/asset-owners/investment-strategy
[Accessed 1st June 2018].

PRI, 2016. *A practical guide to ESG integration for equity investing,* s.l.: s.n.

Renneboog, L., Horst, J. T. & Zhang, C., 2008a. Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance,* Issue 32, p. 1723–1742.

Renneboog, L., Horst, J. T. & Zhang, C., 2008b. The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance,* Volume 14, p. 302–322.

Revelli, C. & Viviani, J.-L., 2015. Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. 24(2), pp. 158-185.

Roselle, P., 2016. The Evolution of Integrating ESG Analysis into Wealth Management Decisions. *Journal of Applied Corporate Finance,* 28(2), pp. 75-80.

Sauer, D. A., 1997. The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini Equity Mutual Fund Author links open overlay panel. *Review of Financial Economics,* 6(2), pp. 137-149.

Schröder, M., 2004. The Performance of Socially Responsible Investments: Investment Funds and Indices. *Financial Markets and Portfolio Management,* 18(2), pp. 122-142.

Senn, S., 2008. The t-test tool. *The Royal Statistical Society.*

Sharpe, 1963. A Simplified Model for Portfolio Analysis. *Management Science.*

Sharpe, W. F., 1994. The Sharpe Ratio. *The Journal of Portfolio Management,* 21(1), pp. 49-58.

Sheikh & Hongtao, 2003. *Non-normality of Market Returns,* s.l.: JP Morgan.

Thomson Reuters, 2018. *Thomson Reuters ESG Scores.* [Online]
Available at:
https://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/esg-scores-methodology.pdf
[Accessed 20th July 2018].

Trinks, P. J. & Scholtens, B., 2017. The Opportunity Cost of Negative Screening in Socially Responsible Investing. *Journal of Business Ethics,* Volume 140, p. 193–208.

Turban, D. B. & Greening, D. W., 1997. Corporate Social Performance and Organizational Attractiveness to Prospective Employees. *The Academy of Management Journal,* 40(3), pp. 658-672.

US SIF, 2017a. *Report on US Sustainable, Responsible and Impact Investing Trends,* s.l.: s.n.

US SIF, 2017b. *USSIF.* [Online]
Available at: https://www.ussif.org/esg
[Accessed June 2018].

US SIF, 2017. *USSIF.org.* [Online]
Available at: https://www.ussif.org/esg
[Accessed June 2018].

Vandekerckhove, W. et al., 2011. *Responsible Investment in Times of Turmoil.* 1 ed. Dordrecht: Springer.

Verheyden, T., Eccles, R. G. & Feiner, A., 2016. ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification. *Journal of Applied Corporate Finance,* 28(2), pp. 47-55.

Wilcoxon, 1945. Individual comparisons by ranking methods. *International Biometric Society.*

8. Appendix

8.1. Tables:

Table "a": R output of Sign-Rank Wilcoxon test for All_Bottom_15 and All_Top_15

Wilcoxon signed rank test
data: All_Bottom_15 and All_Top_15
V = 5294, p-value = 0.1226
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-0.0005302016 0.0039282190
sample estimates:(pseudo)median:
0.00172623

Table "b": specifications for fund screening

Asset status is: *Active*
(Asset Universe is: *Closed End Funds,*
Asset Universe is: *Hedge Funds,*
Asset Universe is: *Mutual Funds,*
Asset Universe is: *Pension Funds*)
(Social Criteria is: *yes,*
Ethical is: *yes,*
Green is: yes)
Tot: 3069

Table "c": Domicile of the Funds

Domicile			
Australia	4	*Malaysia*	3
Austria	10	*Mexico*	1
Belgium	9	*Netherlands*	6
Brazil	2	*Norway*	3
Denmark	35	*South Africa*	4
Finland	4	*Spain*	3
France	52	*Sweden*	26
Germany	15	*Switzerland*	12
Hong Kong	1	*Taiwan*	1
Ireland	15	*UK*	16
Italy	1	*USA*	69
Japan	7		
Liechtenstein	2		
Luxembourg	86		

Table "d": Regression results "Crisis/Non-Crisis", ESG groups

Crisis/Non Crisis

2 factor

	Intercept		MSCI world excess		Adjusted R^2
	Crisis	Non-Crisis	Crisis	Non-Crisis	
TOT Portfolio	-0,0023	0,0025	0,8215***	0,6951 ***	83,9%
ESG < 55	-0,0015	0,0023	0,7042***	0,7066 ***	81,3%
ESG <70	-0,0013	0,0019	0,5989***	0,5655 ***	84,3%
70 < ESG < 75	-0,0012	0,0019	0,5196***	0,4988 ***	78,7%
ESG > 75	-0,0028	0,0015	0,5233***	0,4817 ***	71,2%
ESG > 80	-0,0033	0,0016	0,4562***	0,4059 ***	60,1%

3 factor

	Intercept		MSCI world excess		SMB		HML		Adjusted R^2
	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	
TOT Portfolio	-0,0030	0,0025	0,8327***	0,6990 ***	0,1990	0,0926	-0,2554*	-0,0516	84,3%
ESG < 55	-0,0010	0,0021	0,6981***	0,7099 ***	0,2476	0,4746 ***	0,0705	-0,0576	83,5%
ESG <70	-0,0018	0,0019	0,6054***	0,5744 ***	0,258291**	0,2101 **	-0,1727*	-0,1165 *	85,9%
70 < ESG < 75	-0,0020	0,0020	0,5297***	0,5056 ***	-0,0377	-0,0018	-0,1875*	-0,0848	78,8%
ESG > 75	-0,0037	0,0015	0,5349***	0,4779 ***	-0,1600	-0,0884	-0,1941	0,0504	71,3%
ESG > 80	-0,0042	0,0016	0,4691***	0,3993 ***	-0,1060	-0,0709	-0,2298	0,0836	60,1%

4 factor

	Intercept		MSCI world excess		SMB		HML		MWL (MOM)		Adjusted R^2
	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	
TOT Portfolio	-0,0047*	0,0023	0,7987***	0,710 ***	0,1682	0,0948	-0,3819***	-0,0340	-0,1245**	0,0252	84,6%
ESG < 55	-0,0017	0,0017	0,6780***	0,7158 ***	0,2333	0,479 ***	0,0094	-0,0263	-0,0601	0,0450	83,4%
ESG <70	-0,0029	0,0020	0,5809***	0,5721 ***	0,2400**	0,2085 **	-0,2473**	-0,1292 *	-0,0733*	-0,0183	86,0%
70 < ESG < 75	-0,0027	0,0023	0,5079***	0,5015 ***	-0,0539	-0,0045	-0,2536*	-0,1070	-0,0651	-0,0320	78,9%
ESG > 75	-0,0047*	0,0018	0,507***	0,4727 ***	0,0034	0,0919	-0,2773**	0,0227	-0,082	-0,0398	71,4%
ESG > 80	-0,0052*	0,0019	0,4472***	0,3949 ***	-0,1222	-0,0739	-0,296**	0,0595	-0,0654	-0,0345	59,9%

5 factors

	Intercept		MSCI world excess		SMB		HML		RMW		CMA		Adjusted R^2
	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	Crisis	Non-Crisis	
TOT Portfolio	-0,0038	0,0029	0,8089***	0,6831 ***	0,1682	0,0591	-0,1447	-0,0404	0,1689	-0,1143	-0,1419	-0,1449	84,0%
ESG < 55	0,0019	0,0029 *	0,67710***	0,6906 ***	0,2054	0,42014 ***	-0,0116	-0,2455 **	-0,4792	-0,2959 *	-0,0183	0,2499	84,0%
ESG <70	-0,0007	0,0022 *	0,5904***	0,5608 ***	0,2323**	0,1813 **	-0,1816	-0,1089	-0,1682	-0,0994	-0,0424	-0,1196	85,7%
70 < ESG < 75	-0,0015	0,0022	0,5301***	0,4958 ***	-0,0388	-0,0221	-0,2081	-0,0718	-0,0684	-0,0662	0,0135	-0,1027	78,3%
ESG > 75	-0,0028	0,0019	0,5049***	0,4597 ***	-0,2067	-0,1240	-0,1495	0,1045	-0,1182	-0,0988	-0,1222	-0,2533	71,0%
ESG > 80	-0,0032	0,0019	0,4168***	0,3823 ***	0,2986	0,3824	-0,3977	-0,1014	0,1321	0,1740	0,0286	-0,0614	60,0%

Table "e": Regression results "Crisis/Non-Crisis", Geographical focus groups

Crisis/Non Crisis

2 factor

	Intercept		*MSCI world excess*		*Local factor*		*Adjusted R^2*
	Crisis	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	/
Global	-0,0020	0.0022	0,5177***	0.4506***	/	/	74,8%
Europe	-0,0020	0.0020	0,52***	0.4458***	0,0566	0.3693***	73,2%
North America	0,0000	0.00004	0,7464***	0.7518***	0,9639***	0.9714***	95,2%
Nordic	0,0007	0.0051**	0,5395***	0.4635***	0,504***	0.403***	70,2%

3 factor

	Intercept		*MSCI world excess*		*SMB*		*HML*		*Local factor*		*Adjusted R^2*
	Crisis	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	/
Global	-0,0030	0.00218	0,53022***	0.4579***	0,1251*	0.0814	-0,262	-0.0923	/	/	75,6%
Europe	-0,0037	0.0020	0,5403***	0.4484***	-0,1673	-0.0328	-0,2017	-0.0469	0,0120	0.3746	73,1%
North America	0,0002	-0.0001	0,745***	0.7580***	0,149*	0.2294***	0,0318	-0.0232	0,984***	1.027***	95,9%
Nordic	0,0004	0.0051**	0,5584***	0.4762***	0,1230	0.1158	-0,12607	-0.09707	0,4657***	0.3887***	69,8%

4 factor

	Intercept		*MSCI world excess*		*SMB*		*HML*		*MWL (MOM)*		*Local factor*		*Adjusted R^2*
	Crisis	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	/
Global	-0,0040	0.0021	0,5048***	0.4580***	0,1064	0.0815	-0,3387*	-0.0917	-0,0755*	0.0008	/	/	75,7%
Europe	-0,004*	0.0025	0,51955***	0.4445***	-0,1794	-0.0383	-0,2557*	-0.0812	-0,0542	-0.0517	0,0254	0.3626***	73,1%
North America	0,0000	0.00004	0,7242***	0.7551***	0,1335*	0.2277***	-0,0245	-0.0408	-0,0573*	-0.0257	0,9670***	1.032***	96,0%
Nordic	-0,0015	0.0060**	0,5212***	0.4610***	0,1099	0.1057	-0,2921	-0.1691	-0,1556*	-0.1045	0,423***	0.3929***	70,4%

5 factors

	Intercept		*MSCI world excess*		*SMB*		*HML*		*RMW*		*CMA*		*Local factor*		*Adjusted R^2*
	Crisis	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	*Crisis*	*Non-Crisis*	/
Global	-0,0026	0.0024*	0,5041***	0.4435***	0,0857	0.0524	-0,2095	-0.0630	-0,0551	-0.0887	-0,1143	-0.1718	/	/	75,3%
Europe	-0,00299	0.0025	0,547***	0.429***	-0,1623	-0.0773	-0,2571	-0.0503	-0,1375	-0.1607	0,0517	-0.1465	0,0062	0.3698***	72,7%
North America	0,0013	0.00006	0,734***	0.7489***	0,1292*	0.2089***	0,0082	-0.02856	-0,1781	-0.07765	-0,0168	-0.0605	0,9704***	1.030***	95,9%
Nordic	0,0013	0.0056**	0,5757***	0.4591***	0,1433	0.0756	-0,2308	-0.1233	-0,1828	-0.1609	0,1299	-0.0840	0,4755***	0.3858***	69,1%

Table "f": Hypothesis testing for Dummy variables, group: "Tot Portfolio", model: "five factor model"

```
reg<-
lm(Tot_Portfolio~MSCI_World+SMB+HML+RMW+CMA+Dummy_Crisis+Dummy_Crisis*MSCI_World+Dummy_Crisis
*SMB+Dummy_Crisis*HML+ Dummy_Crisis*RMW+Dummy_Crisis*CMA)
F<-c(c(0,0,0,0,0,0,1,1,1,1,1,1))
linearHypothesis(reg,hypothesis.matrix=F)
```

```
Linear hypothesis test


Hypothesis:
Dummy_Crisis + MSCI_World:Dummy_Crisis + SMB:Dummy_Crisis + HML:Dummy_Crisis +
RMW:Dummy_Crisis + CMA:Dummy_Crisis = 0


Model 1: restricted model
Model 2: Tot_Portfolio ~ MSCI_World + SMB + HML + RMW + CMA + Dummy_Crisis +
   Dummy_Crisis * MSCI_World + Dummy_Crisis * SMB + Dummy_Crisis *
   HML + Dummy_Crisis * RMW + Dummy_Crisis * CMA


  Res.Df      RSS Df  Sum of Sq      F Pr(>F)
1    134 0.031447
2    133 0.031341  1 0.00010621 0.4507 0.5032
```

Table "g": auto-Arima output

```
auto.arima(All_Bottom_5,ic=c("bic"),stepwise=FALSE,approximation=FALSE)

ARIMA(0,0,1) with non-zero mean

Coefficients:
        ma1    mean
     0.3548  0.0089
s.e.  0.0817  0.0048

sigma^2 estimated as 0.001721:  log likelihood=239.01
AIC=-472.01   AICc=-471.83   BIC=-463.3
```

8.2. Figures:

Figure a: Normalized returns of "ESG<55" plotted against MSCI world with visual identification of crisis periods

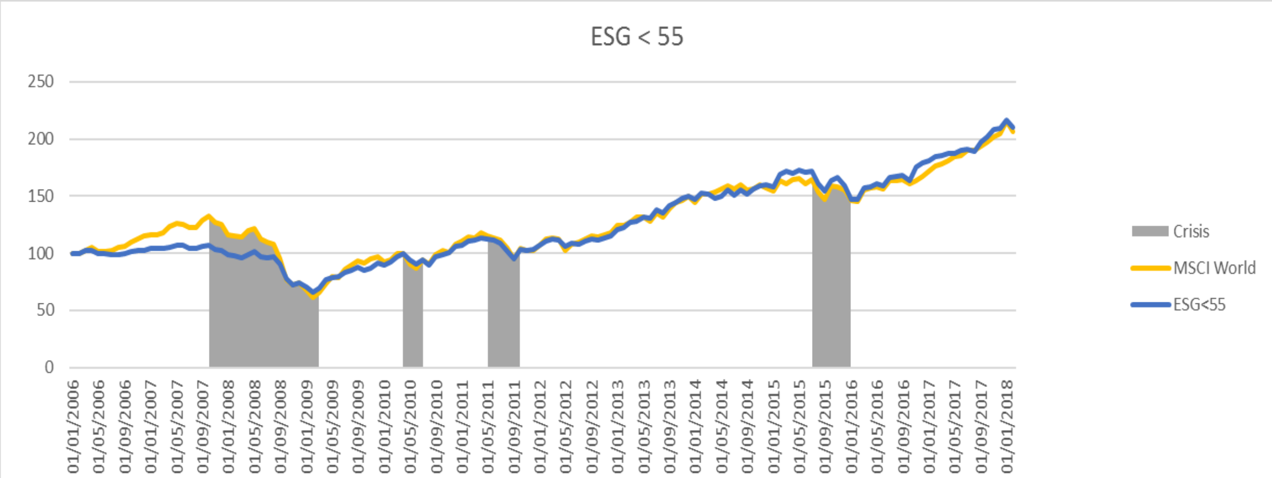


Figure b: Normalized returns of "ESG<70" plotted against MSCI world with visual identification of crisis periods

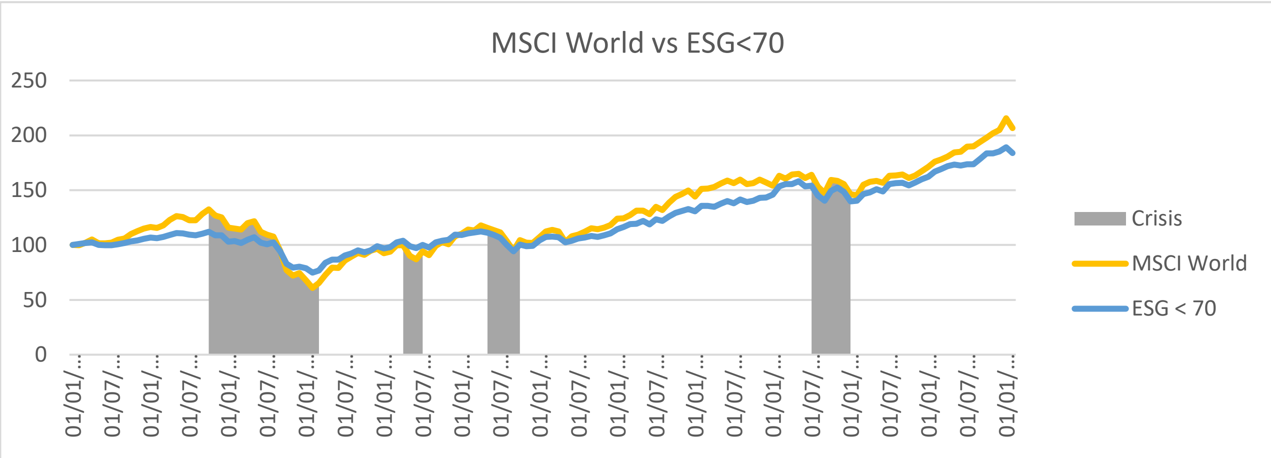


Figure c: Normalized returns of "70<ESG<75" plotted against MSCI world with visual identification of crisis periods

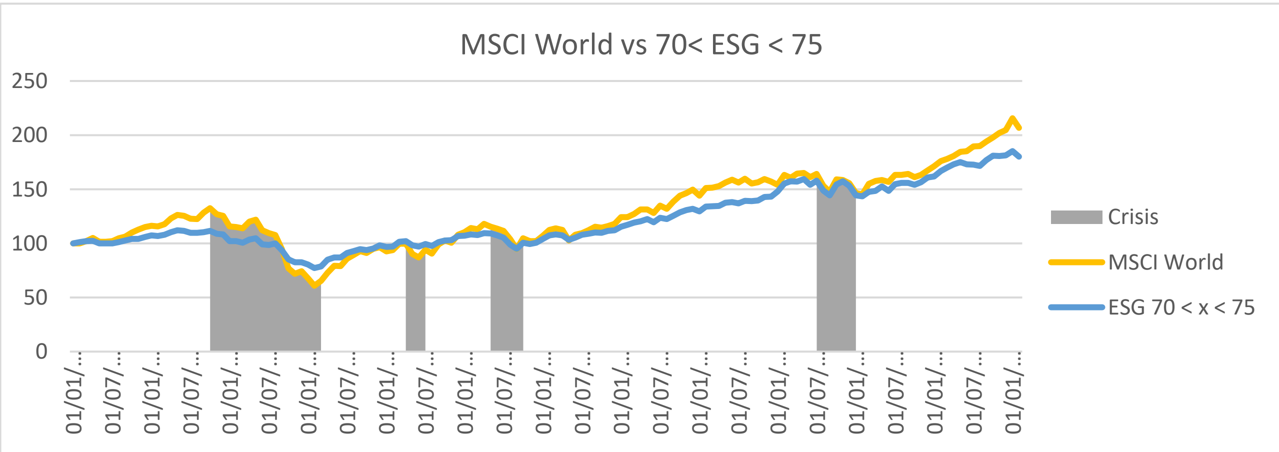


Figure d: Normalized returns of "ESG>75" plotted against MSCI world with visual identification of crisis periods

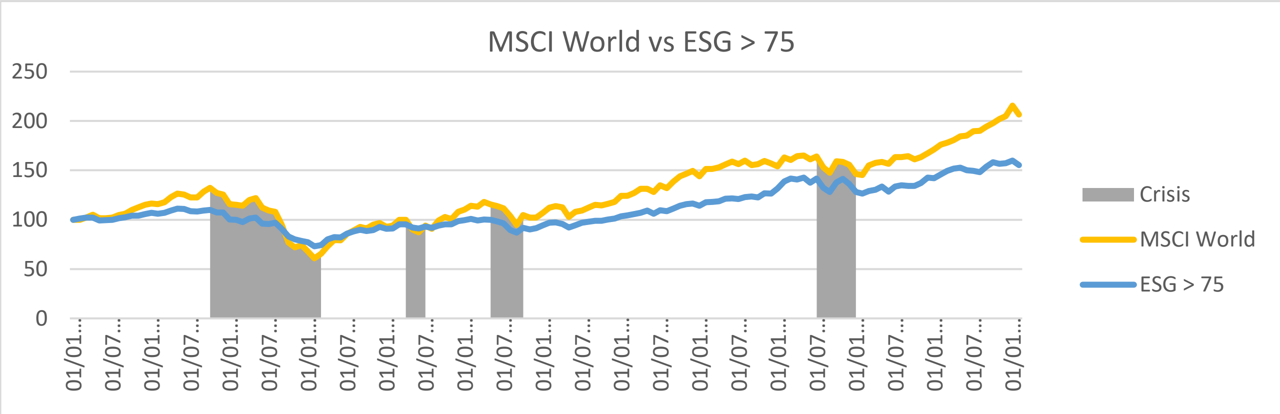


Figure e: Normalized returns of "ESG>80" plotted against MSCI world with visual identification of crisis periods

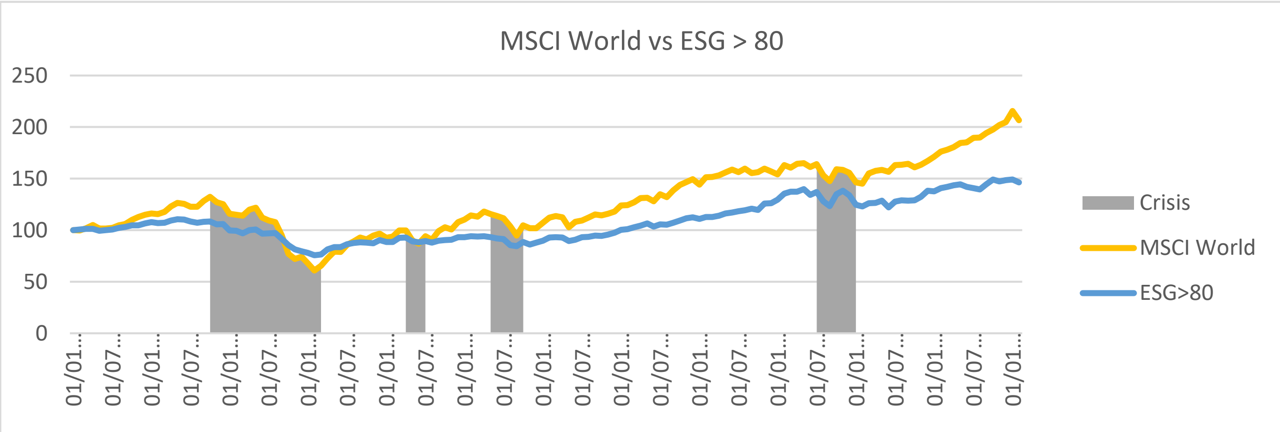


Figure f: Normalized returns of "Total Portfolio" plotted against MSCI world with visual identification of crisis periods

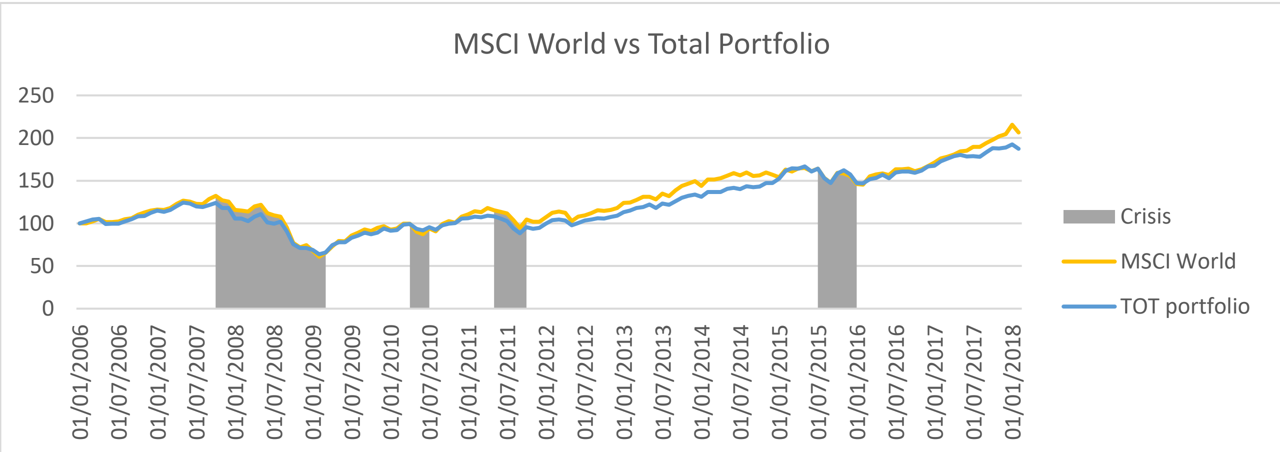


Figure g: Normalized returns of "Global" plotted against MSCI world with visual identification of crisis periods

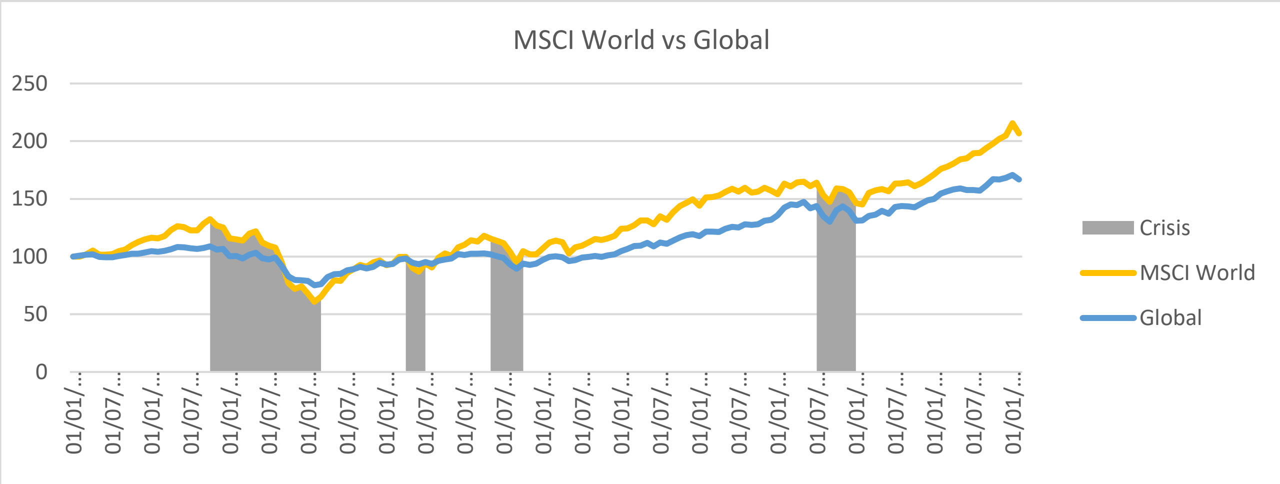


Figure h: Normalized returns of "North_America" plotted against MSCI world with visual identification of crisis periods

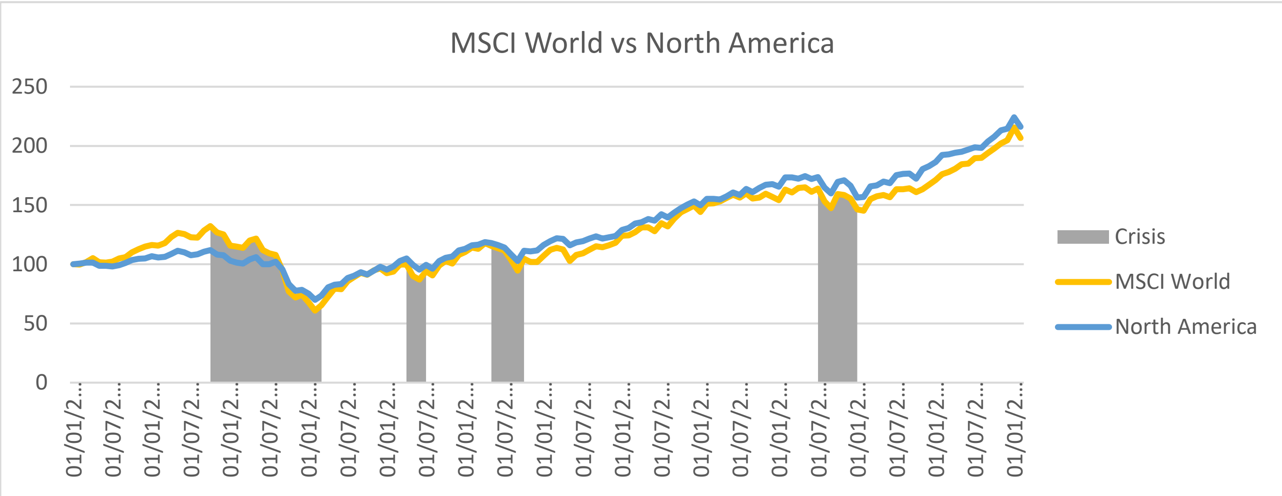


Figure i: Normalized returns of "Europe" plotted against MSCI world with visual identification of crisis periods

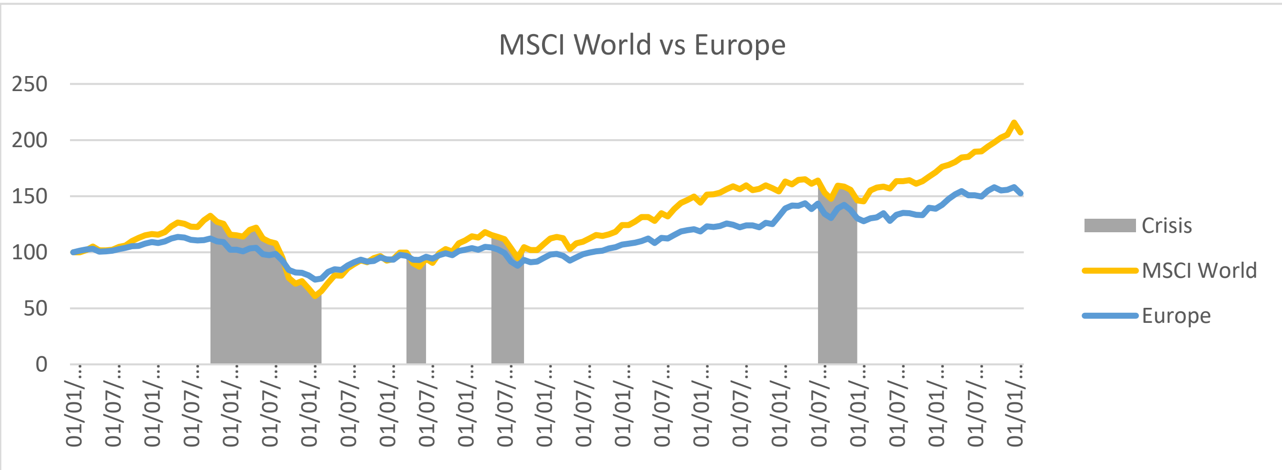


Figure l: Normalized returns of "ESG<55" plotted against MSCI world with visual identification of crisis periods

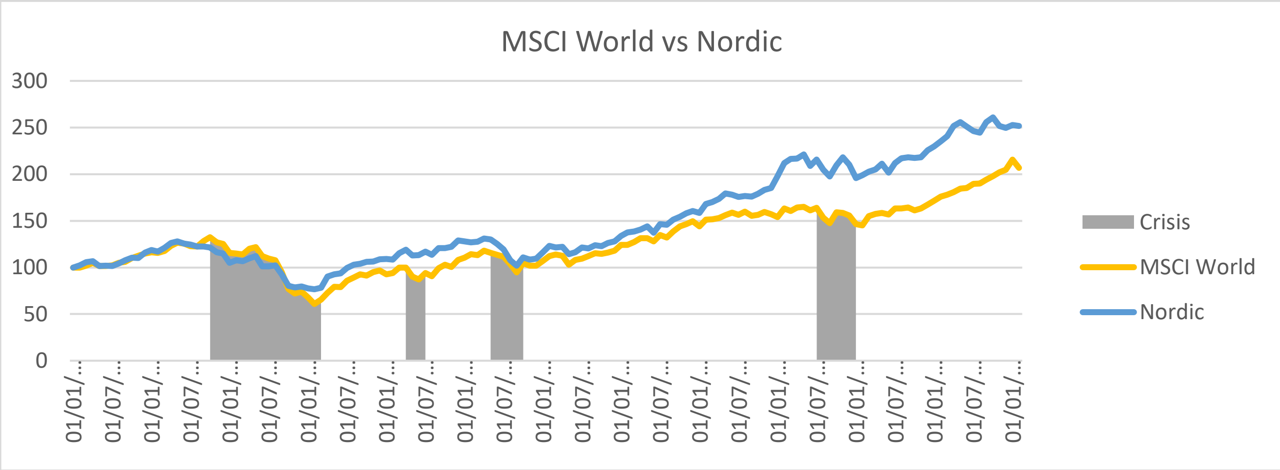


Figure m: Normalized returns of "Nordic" plotted against MSCI Nordic with visual identification of crisis periods

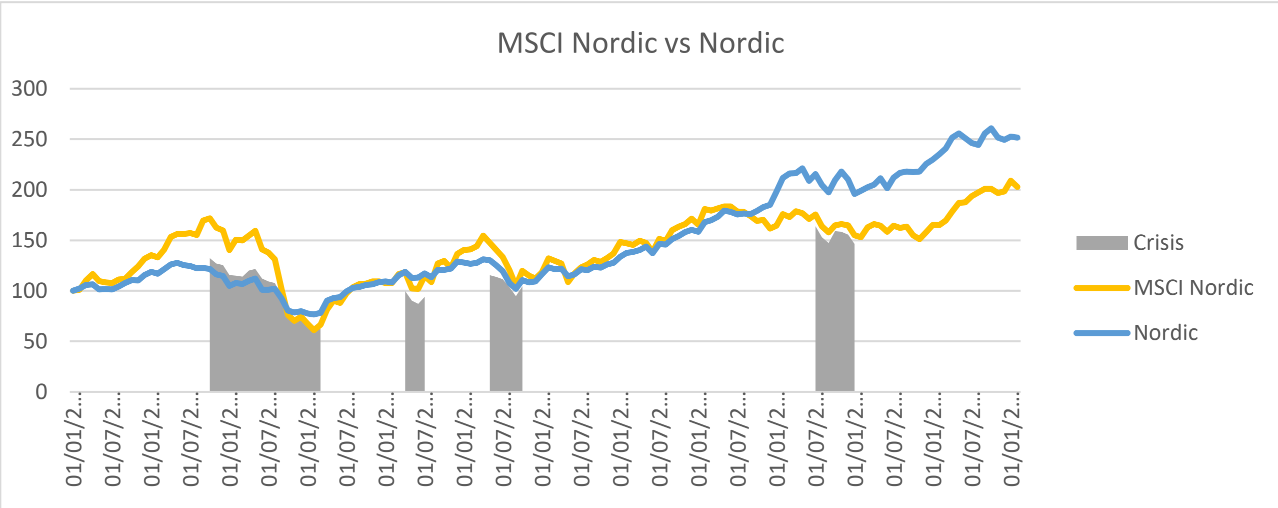


Figure n: Normalized returns of "Europe" plotted against MSCI Europe with visual identification of crisis periods

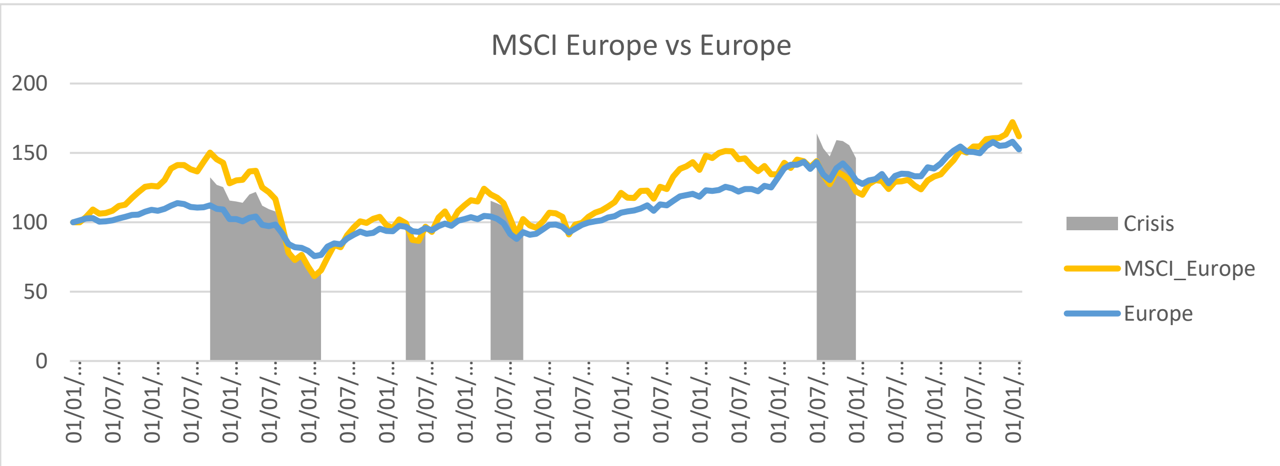
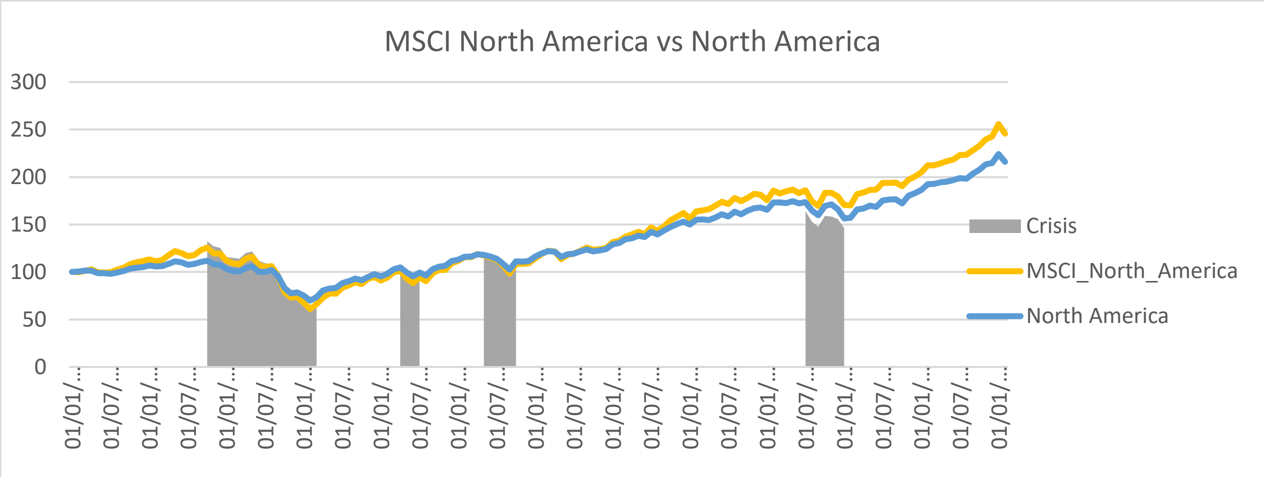


Figure o: Normalized returns of "North America" plotted against MSCI North America with visual identification of crisis periods

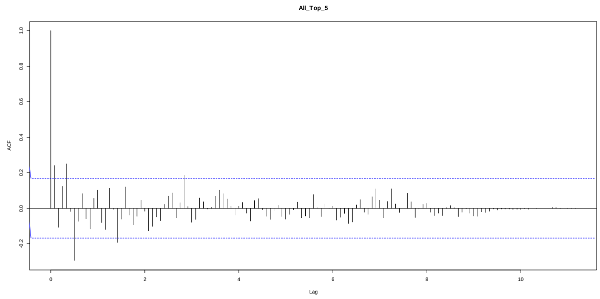


8.3. “Autocorrelation”: Autocorrelation Analysis of the stock groups

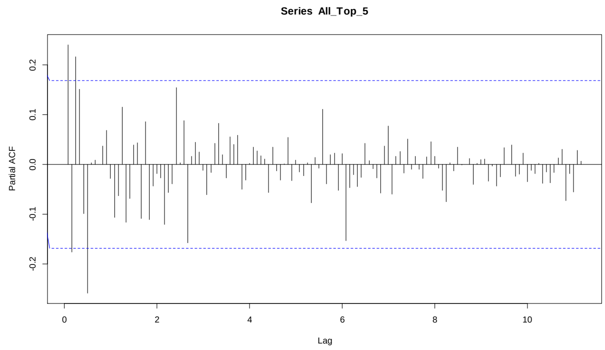
All_Top_5:

Tau3 (Test Value)	-8.5604
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5601
z.lag.1 P(>|t|)	2.88e-14 ***
tt P(>|t|)	0.1330
z.diff.lag P(>|t|)	0.0306 *

Tau3 is lower than its critical value, z.lag.1 is statistically lower than zero, tt and intercept are not statistically different from zero. Given these results we can say the series to be stationary and we can proceed in our analysis.



Total Autocorrelation



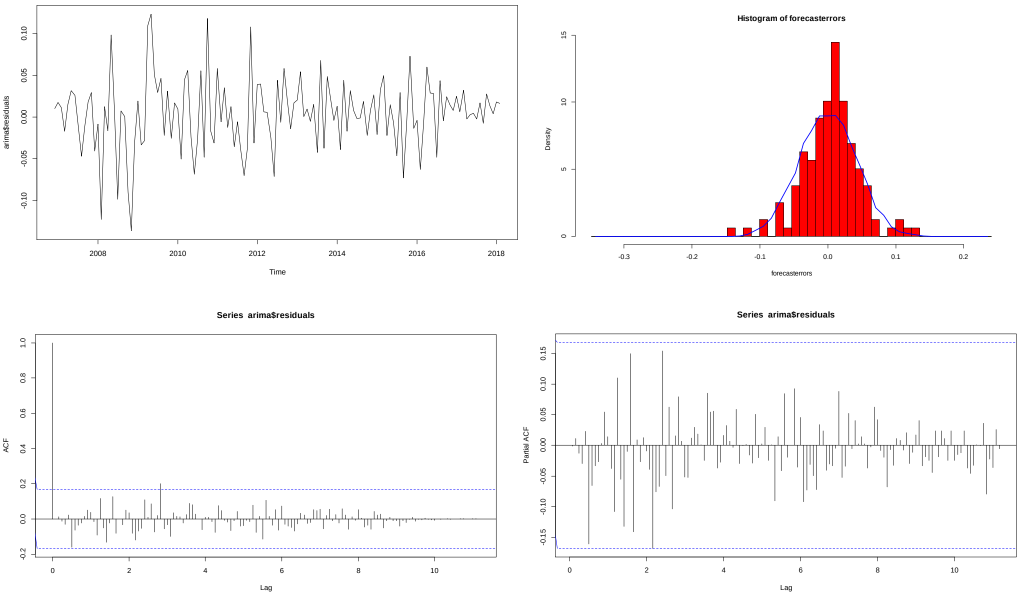
Partial Autocorrelation

The acf shows a cut off after lag=1, while the pacf shows a cut off after lag=3. We can therefore proceed with the autoarima function:

AIC optimization	BIC optimization
Series: All_Top_5	Series: All_Top_5
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7155 -0.7936 0.4252 -0.4282 0.5238	0.3655
s.e. 0.2346 0.1416 0.0806 0.2699 0.1453	s.e. 0.0891
sigma^2 estimated as 0.001863: log likelihood=235	sigma^2 estimated as 0.002031: log likelihood=227.33
AIC=-458 AICc=-457.34 BIC=-440.56	AIC=-450.66 AICc=-450.57 BIC=-444.85

Given the mixed result, we select the model that better reflects the autocorrelation observed in the acf and pacf graphs. Therefore, we select ARIMA(3,0,2). This imply a block length of 3.

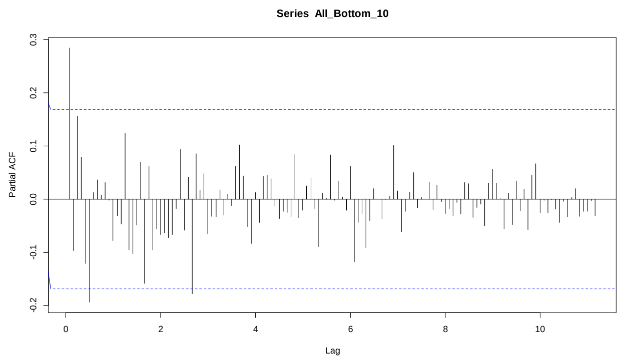
Residuals:



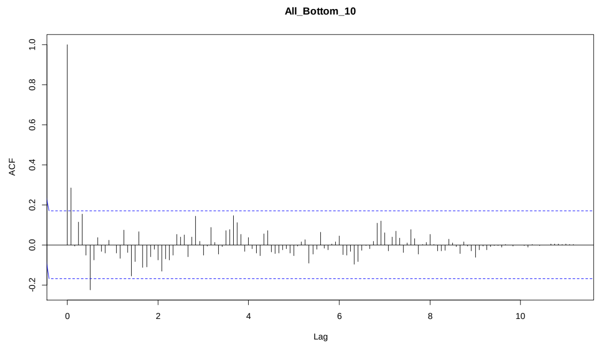
All_Bottom_10:

Tau3 (Test Value)	-7.5873
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.966
z.lag.1 P(>|t|)	5.73e-12 ***
tt P(>|t|)	0.303
z.diff.lag P(>|t|)	0.245

The series is stationary with no drift and no trend.



Partial Autocorrelation



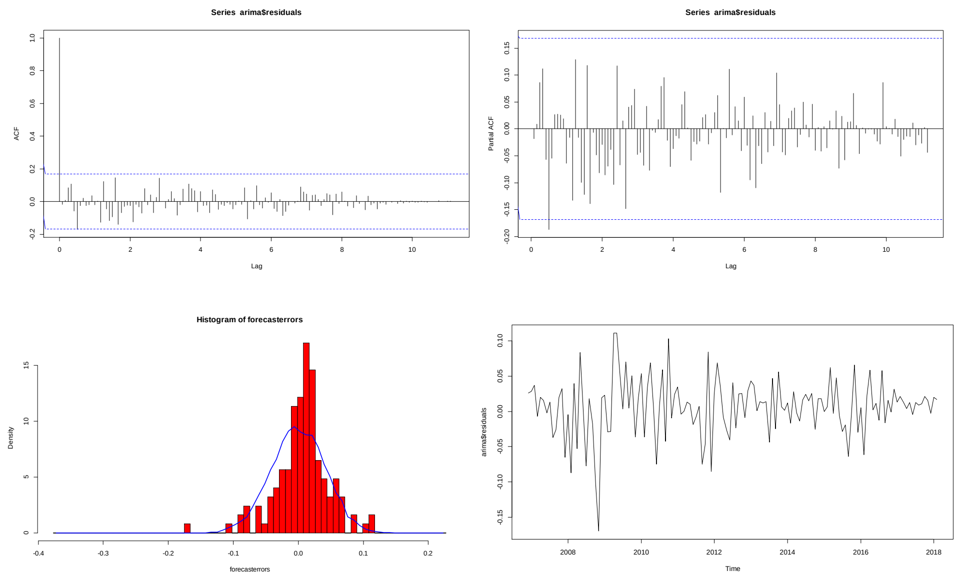
Total Autocorrelation

Both pacf and acf cut off after lag=1. We would therefore expect a serie ARIMA(1,0,1)

AIC optimization	BIC optimization
Series: All_Bottom_5	Series: All_Bottom_5
ARIMA(0,0,1) with non-zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ma1 mean	ma1
0.3548 0.0089	0.3705
s.e. 0.0817 0.0048	s.e. 0.0798
sigma^2 estimated as 0.001721: log likelihood=239.01	sigma^2 estimated as 0.001752: log likelihood=237.31
AIC=-472.01 AICc=-471.83 BIC=-463.3	AIC=-470.62 AICc=-470.53 BIC=-464.81

The difference in these two ARIMA model do not concern the factors (since both have c(0,0,1)) but the presence of a zero mean (i.e. if the intercept of the function should be zero). Since our series is stationary with no drift – and since the BIC or AIC optimization are equivalent – we prefer to keep the zero-mean model. Finally, the block length of our bootstrap process will be 1.

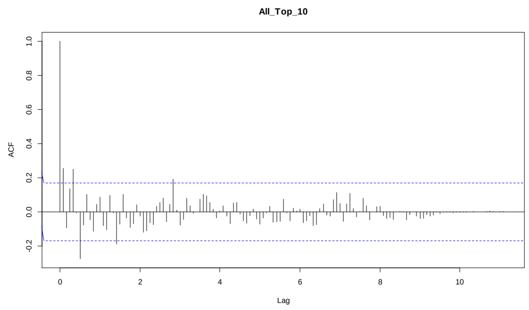
Residuals:



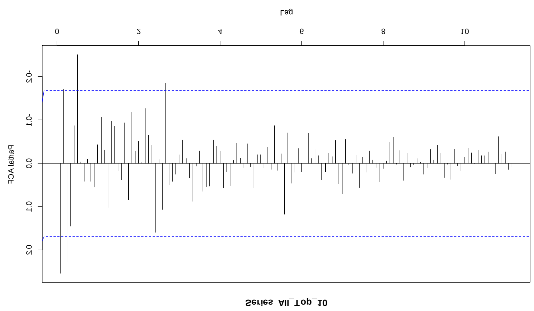
All_Top_10:

Tau3 (Test Value)	-8.4474
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5676
z.lag.1 P(>|t|)	5.38e-14 ***
tt P(>|t|)	0.1378
z.diff.lag P(>|t|)	0.0358 *

The series is stationary with no drift and no trend.



Total Autocorrelation



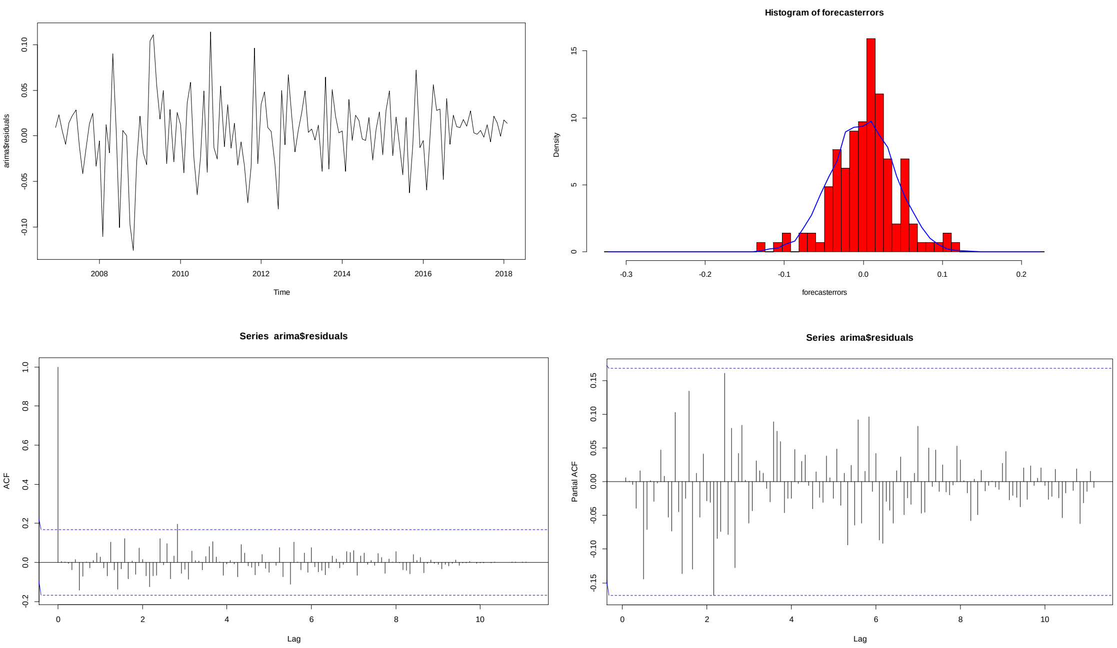
Partial Autocorrelation

The ACF graph cuts off after lag=1, even though it shows relevan autocorrelation also for lag=4 and 6. The pacf cuts off after lag=3, but it is relevant also for lag 6. Therefore, we would expect a ARIMA(3,0,2) model

AIC optimization	BIC optimization
Series: All_Top_10	Series: All_Top_10
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7574 -0.8143 0.4378 -0.4647 0.5415	0.3808
s.e. 0.2299 0.1602 0.0787 0.2688 0.1703	s.e. 0.0880
sigma^2 estimated as 0.001694: log likelihood=241.37	sigma^2 estimated as 0.001848: log likelihood=233.68
AIC=-470.75 AICc=-470.09 BIC=-453.31	AIC=-463.36 AICc=-463.27 BIC=-457.55

Since the ARIMA(3,0,2) is the one that represent better the autocorrelation graphs, we will use it for the determination of the block length (that will be 3).

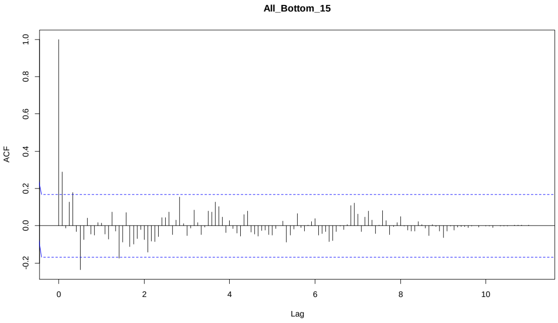
Residuals:



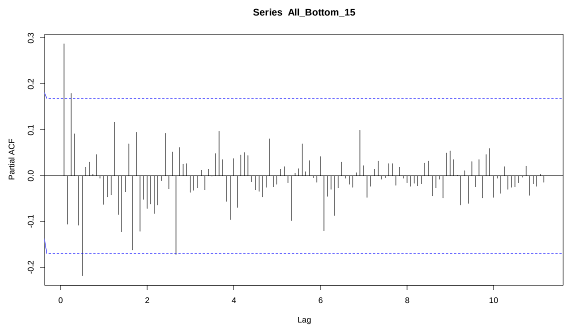
All_Bottom_15

Tau3 (Test Value)	-7.6481
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.937
z.lag.1 P(>|t|)	4.14e-12 ***
tt P(>|t|)	0.293
z.diff.lag P(>|t|)	0.205

The series is stationary and has neither drift nor trend.



Total Autocorrelation



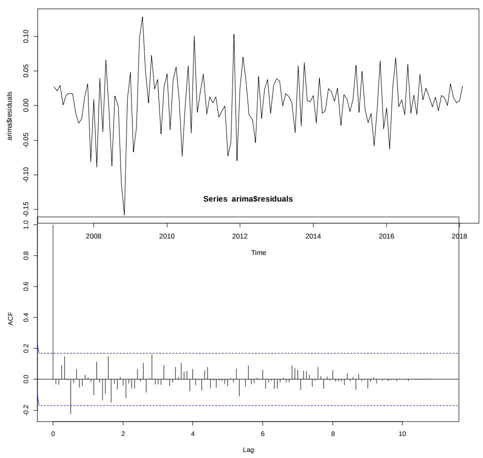
Partial Autocorrelation

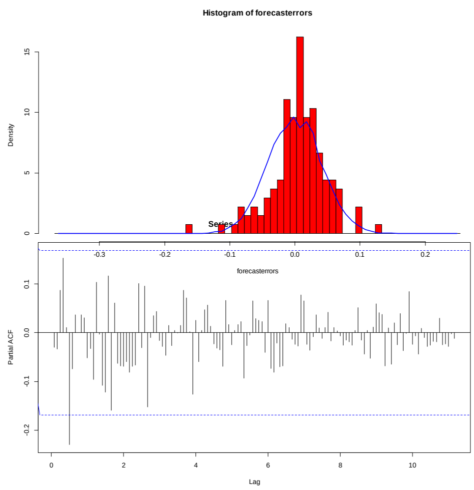
ACF cuts off at lag=1 but it is relevant also for lag=6; pacf cuts off at lag=1 but it is relevant also for lag=3. We will therefore expect one of the following models: ARIMA(0,0,1), ARIMA(1,0,1), ARIMA (3,0,1)

AIC optimization	BIC optimization
Series: All_Bottom_15	Series: All_Bottom_15
ARIMA(3,0,2) with non-zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2 mean	ma1
0.8975 -0.9343 0.3944 -0.6222 0.7138 0.0086	0.3666
s.e. 0.1882 0.2216 0.0783 0.2015 0.2407 0.0059	s.e. 0.0831
sigma^2 estimated as 0.001716: log likelihood=241.1	sigma^2 estimated as 0.001811: log likelihood=235.07
AIC=-468.19 AICc=-467.31 BIC=-447.86	AIC=-466.14 AICc=-466.05 BIC=-460.33

In this case the AIC suggest an ARIMA (3,0,2) with non-zero mean, while the BIC suggest an ARIMA(0,0,1) with zero mean. The presence of a cut off after lag=1 in the pacf graph and the absence of a drift in the series point us to accept the ARIMA(0,0,1). Therefore, the bootstrap block length will be =1.

Residuals:

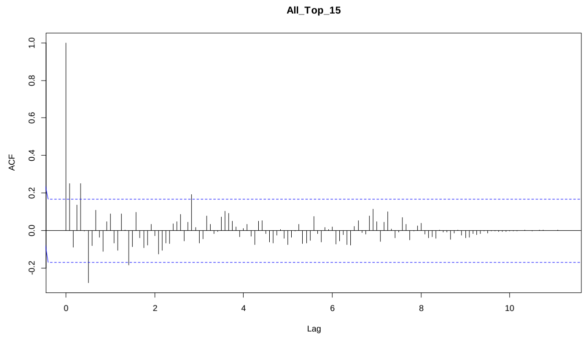




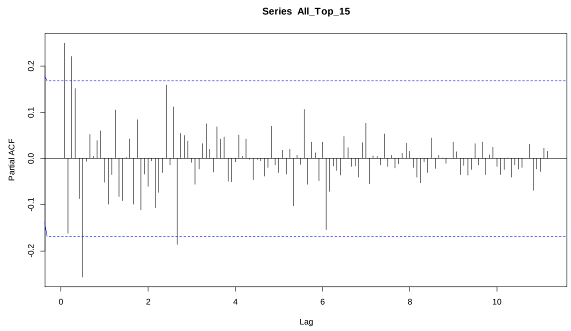
All_Top_15

Tau3 (Test Value)	-8.4036
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5492
z.lag.1 P(>|t|)	6.84e-14 ***
tt P(>|t|)	0.1291
z.diff.lag P(>|t|)	0.0457 *

The series is stationary with no drift nor trend.



Total Autocorrelation



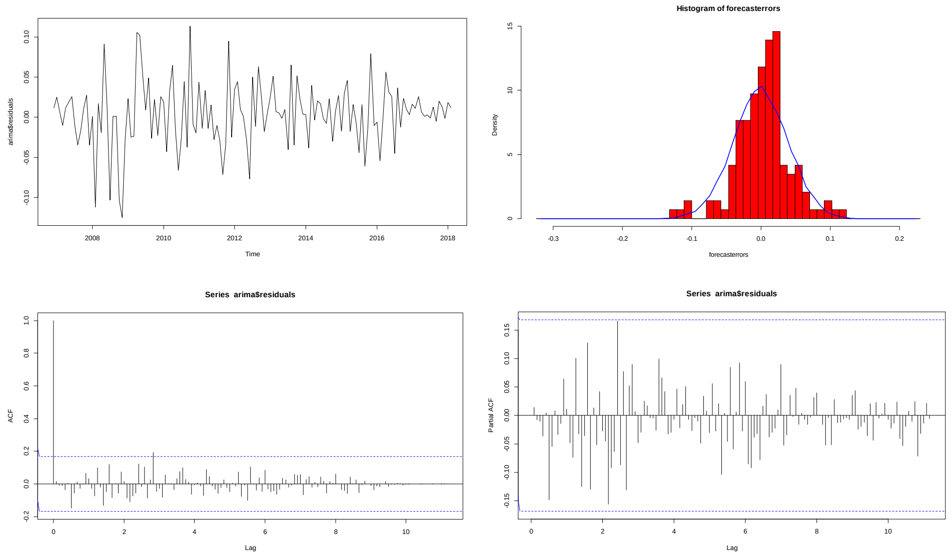
Partial Autocorrelation

The acf shows a cut off after lag 1, while the pacf shows a relevant partial autocorrelation for lags 1,3 and 6.

AIC optimization	BIC optimization
Series: All_Top_15	Series: All_Top_15
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.8147 -0.8557 0.4417 -0.5429 0.5983	0.3679
s.e. 0.2262 0.1857 0.0763 0.2681 0.2150	s.e. 0.0881
sigma^2 estimated as 0.001649: log likelihood=243.17	sigma^2 estimated as 0.001808: log likelihood=235.17
AIC=-474.33 AICc=-473.67 BIC=-456.9	AIC=-466.33 AICc=-466.24 BIC=-460.52

The presence of a relevant total autocorrelation after lag 1 in the acf graph (and also in the pacf) suggest that there could be a loss of information by considering only one level of autocorrelation in the series. Therefore, we accept the ARIMA(3,0,2) model and we will set a bootstrap block length = 3.

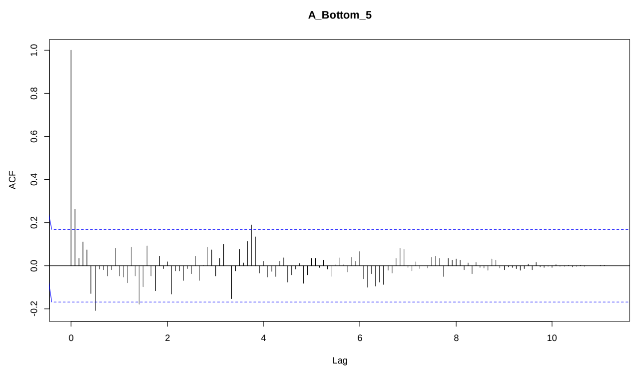
Residuals:



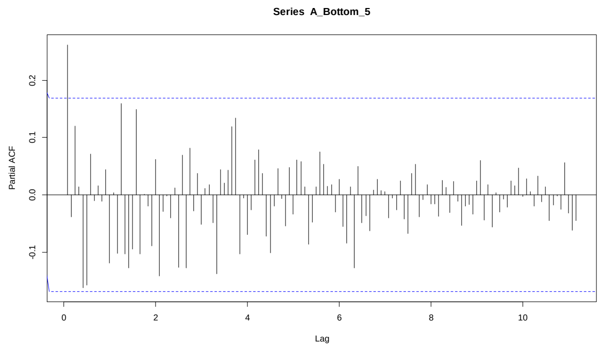
Asia_Bottom_5

Tau3 (Test Value)	-7.2319
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.786
z.lag.1 P(>|t|)	3.75e-11 ***
tt P(>|t|)	0.403
z.diff.lag P(>|t|)	0.632

The series is stationary with no drift and no trend.



Total Autocorrelation



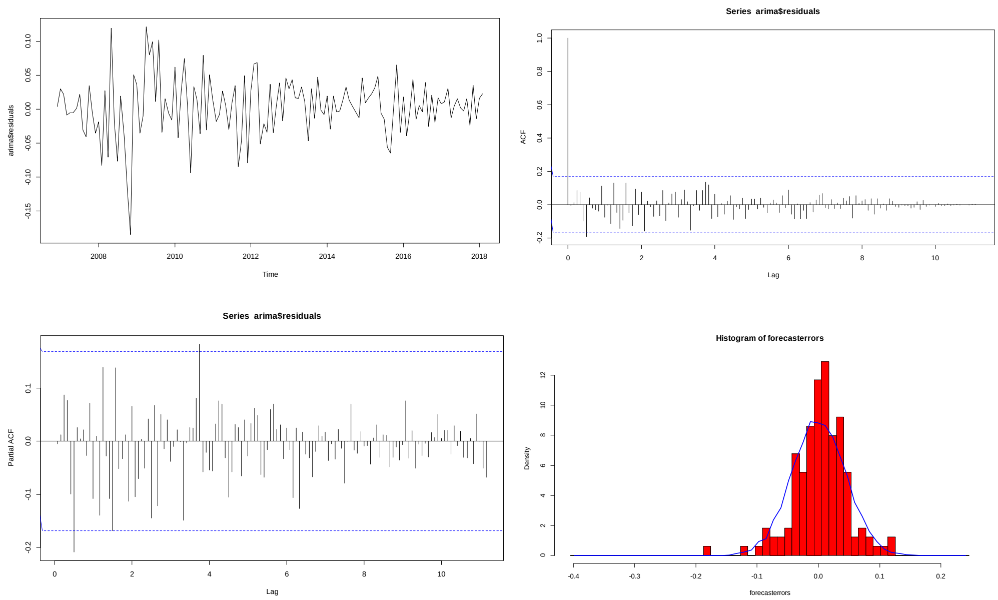
Partial Autocorrelation

Both the acf and the pacf cuts off after lag=1. Therefore, we expect a ARIMA(0,0,1) series.

AIC optimization	BIC optimization
Series: A_Bottom_5	Series: A_Bottom_5
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.2233 -0.8987 0.3382 0.0676 0.887	0.2854
s.e. 0.1351 0.0607 0.0843 0.1267 0.080	s.e. 0.0832
sigma^2 estimated as 0.001871: log likelihood=234.67	sigma^2 estimated as 0.001942: log likelihood=230.37
AIC=-457.33 AICc=-456.68 BIC=-439.9	AIC=-456.74 AICc=-456.65 BIC=-450.93

Given the weak autocorrelation in both the graphs, we decide to accept the ARIMA (0,0,1) model. Also, the ARIMA(0,0,1) has a very similar AIC value, while a way higher BIC. Finally, the bootstrap block length will be =1.

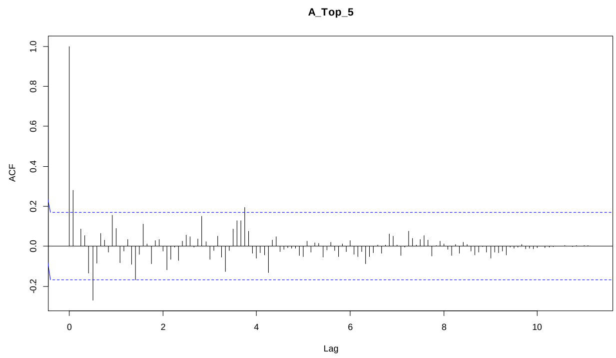
Residuals:



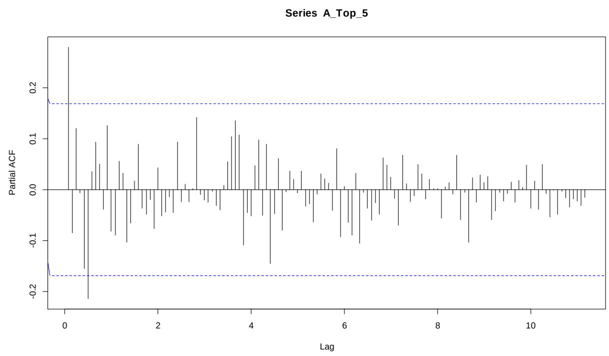
Asia_Top_5

Tau3 (Test Value)	-7.6778
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.319
z.lag.1 P(>|t|)	3.53e-12 ***
tt P(>|t|)	0.105
z.diff.lag P(>|t|)	0.251

The series is stationary with no drift and no trend.



Total Autocorrelation



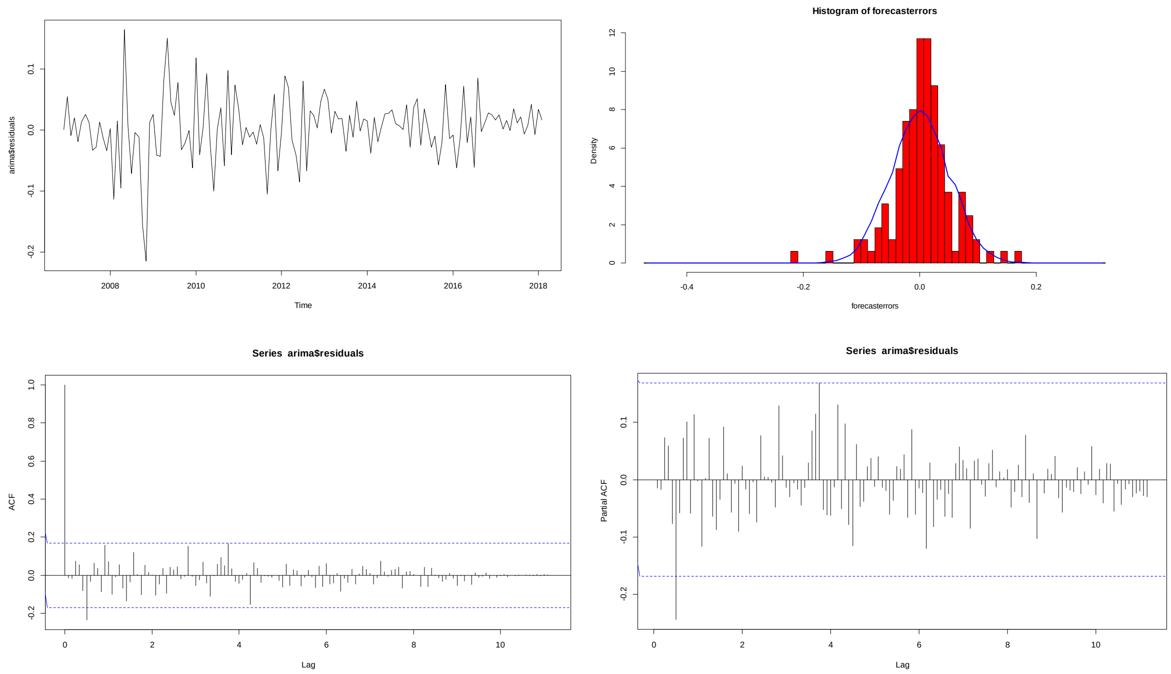
Partial Autocorrelation

Both pacf and acf cut off at lag=1.

AIC optimization	BIC optimization
Series: A_Top_5	Series: A_Top_5
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.1329 -0.9198 0.3103 0.1958 0.9222	0.3301
s.e. 0.1094 0.0368 0.0885 0.0779 0.0667	s.e. 0.0847
sigma^2 estimated as 0.002584: log likelihood=212.7	sigma^2 estimated as 0.002695: log likelihood=208.25
AIC=-413.4 AICc=-412.74 BIC=-395.96	AIC=-412.5 AICc=-412.41 BIC=-406.69

Again, the ARIMA(0,0,1) seems to be the fittest to our graphs, moreover, the difference in the BIC score is strong, while the two AIC scores are very close to each other. Therefore, we will use the AIRMA(0,0,1) and we will set a block length =1.

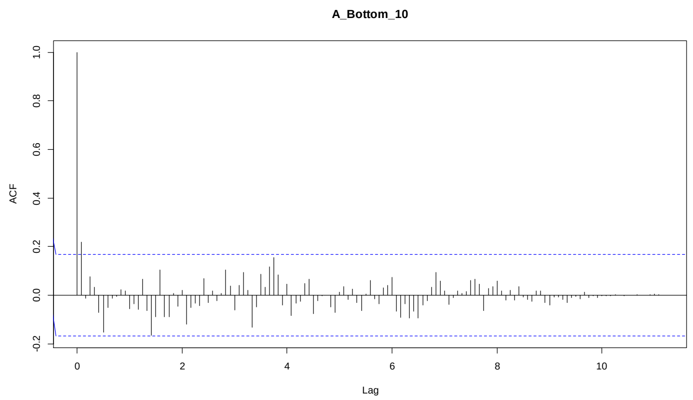
Residuals:



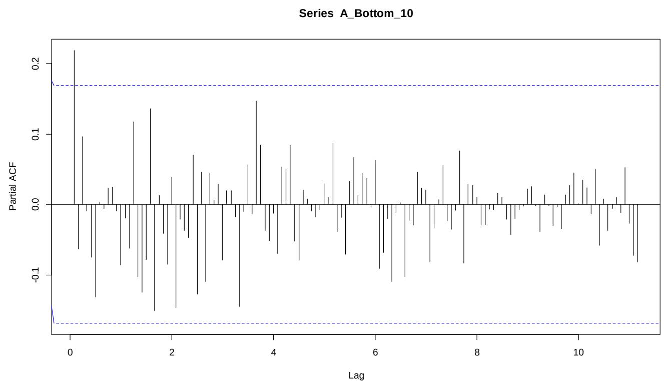
Asia_Bottom_10:

Tau3 (Test Value)	-7.6255
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.833
z.lag.1 P(>|t|)	4.67e-12 ***
tt P(>|t|)	0.389
z.diff.lag P(>|t|)	0.448

The series is stationary with no drift and no trend.



Total Autocorrelation



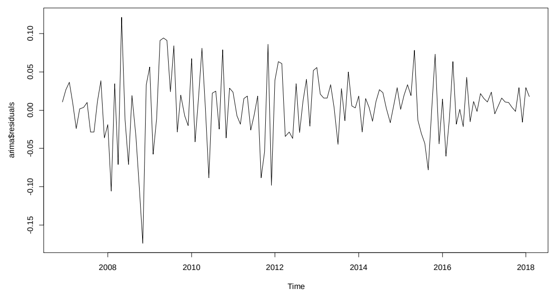
Partial Autocorrelation

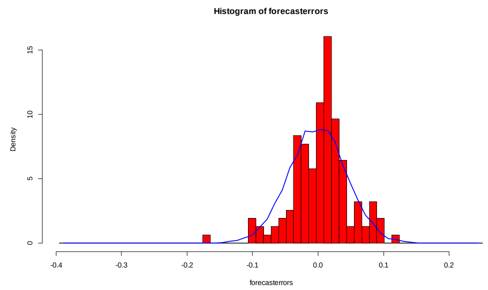
Expected ARIMA(0,0,1)

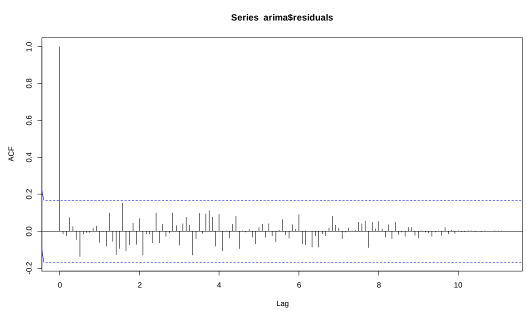
AIC optimization	BIC optimization
Series: A_Bottom_10	Series: A_Bottom_10
ARIMA(0,0,1) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ma1	ma1
0.2559	0.2559
s.e. 0.0873	s.e. 0.0873
sigma^2 estimated as 0.001975: log likelihood=229.26	sigma^2 estimated as 0.001975: log likelihood=229.26
AIC=-454.51 AICc=-454.42 BIC=-448.7	AIC=-454.51 AICc=-454.42 BIC=-448.7

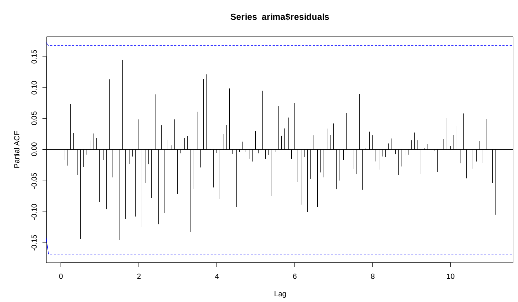
Both BIC and AIC scores indicate ARIMA(0,0,1) as the fittest model. Therefore, we will use a bootstrap block length = 1.

Residuals:





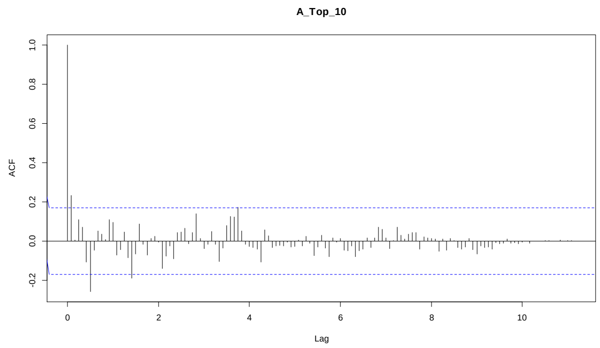




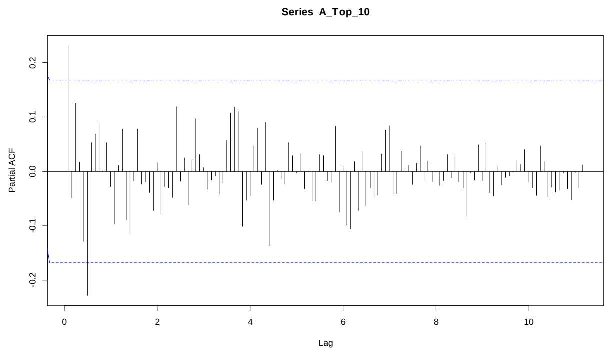
Asia_Top_10:

Tau3 (Test Value)	-7.6605
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.320
z.lag.1 P(>|t|)	3.88e-12 ***
tt P(>|t|)	0.102
z.diff.lag P(>|t|)	0.454

The series is stationary with no drift and no trend.



Total Autocorrelation



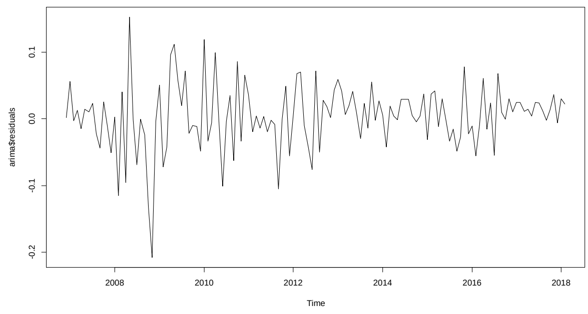
Partial Autocorrelation

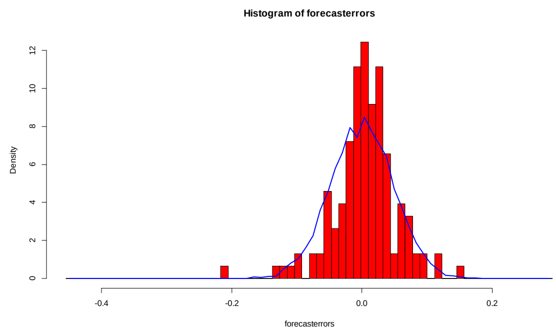
acf and pacf cut off after lag=1. Expect ARIMA(0,0,1)

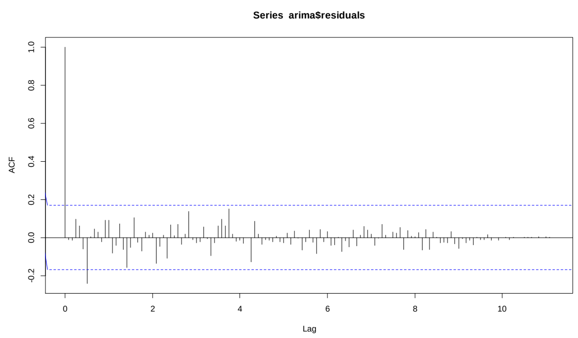
AIC optimization	BIC optimization
Series: A_Top_10	Series: A_Top_10
ARIMA(2,0,3) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ma1 ma2 ma3	ma1
0.6503 -0.6105 -0.430 0.4895 0.3402	0.2642
s.e. 0.2129 0.1535 0.206 0.1419 0.0852	s.e. 0.0868
sigma^2 estimated as 0.002334: log likelihood=219.79	sigma^2 estimated as 0.002433: log likelihood=215.16
AIC=-427.59 AICc=-426.93 BIC=-410.15	AIC=-426.31 AICc=-426.22 BIC=-420.5

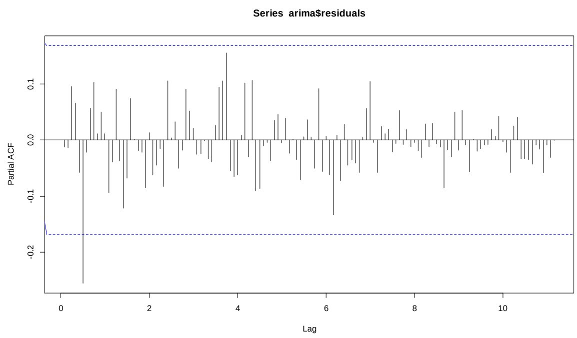
The ARIMA(0,0,1) has lower values of BIC and fit better the autocorrelation graphs. We will use a bootstrap length =1.

Residuals:





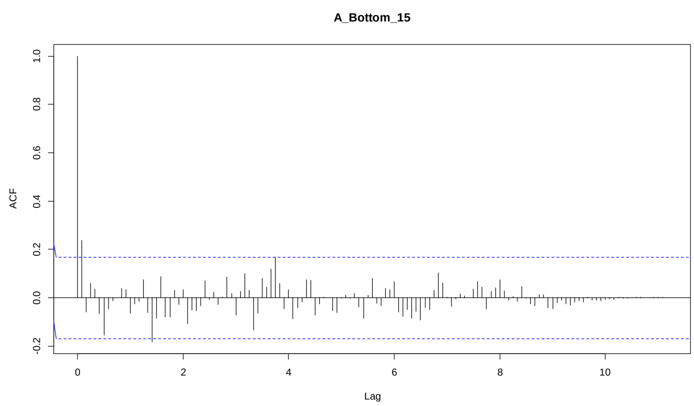




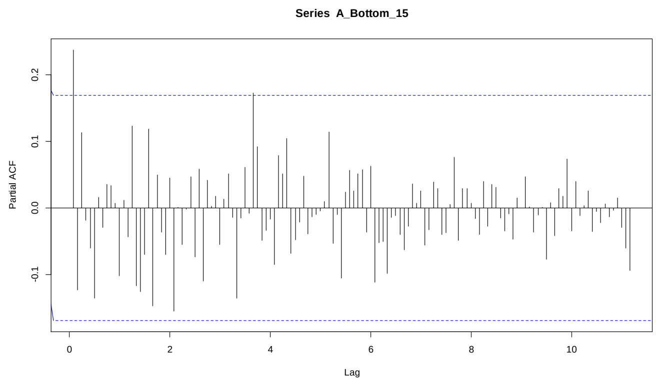
Asia_Bottom_15:

Tau3 (Test Value)	-8.4036
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5492
z.lag.1 P(>|t|)	6.84e-14 ***
tt P(>|t|)	0.1291
z.diff.lag P(>|t|)	0.0457 *

The series is stationary with no drift and no trend.



Total Autocorrelation



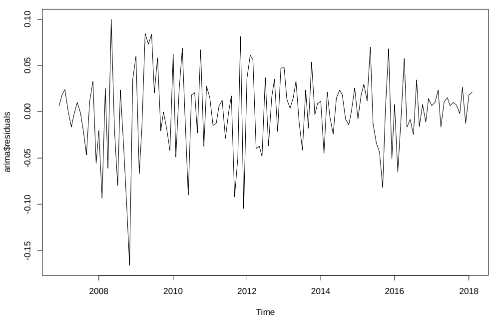
Partial Autocorrelation

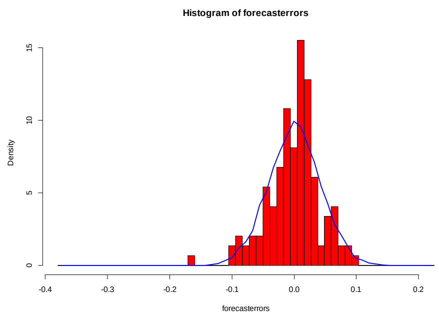
We expect ARIMA(0,0,1)

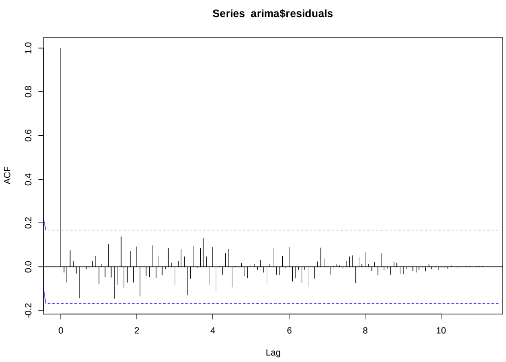
AIC optimization	BIC optimization
Series: A_Bottom_15	Series: A_Bottom_15
ARIMA(0,0,1) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ma1	ma1
0.3140	0.3140
s.e. 0.0904	s.e. 0.0904
sigma^2 estimated as 0.001838: log likelihood=234.07	sigma^2 estimated as 0.001838: log likelihood=234.07
AIC=-464.15 AICc=-464.05 BIC=-458.33	AIC=-464.15 AICc=-464.05 BIC=-458.33

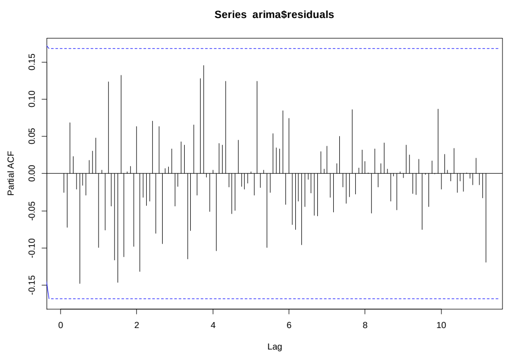
The Arima(0,0,1) has the lowest both AIC and BIC. We will use a bootstrap length=1.

Residuals:





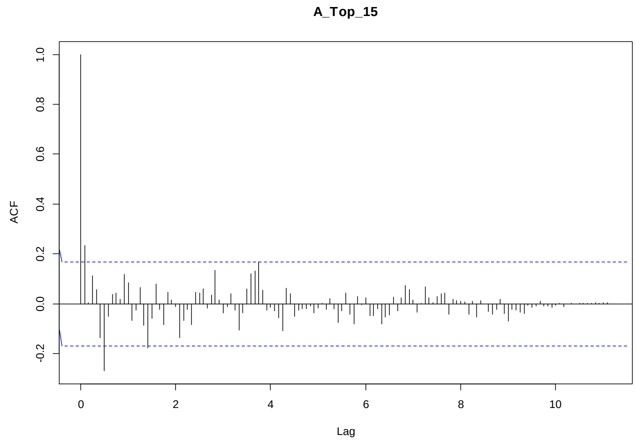




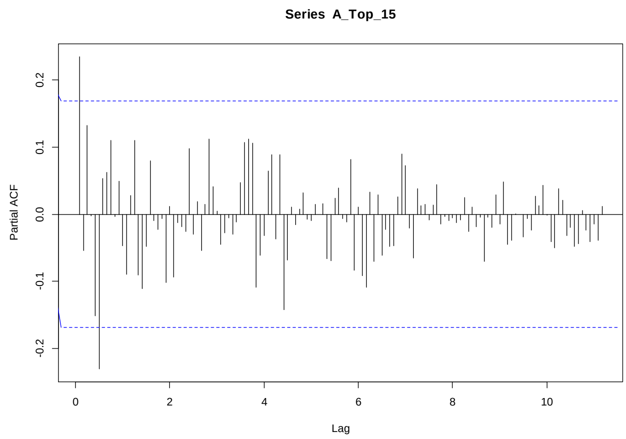
Asia_Top_15:

Tau3 (Test Value)	-7.6657
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.350
z.lag.1 P(>|t|)	3.77e-12 ***
tt P(>|t|)	0.108
z.diff.lag P(>|t|)	0.427

The series is stationary with no drift and no trend.



Total Autocorrelation



Partial Autocorrelation

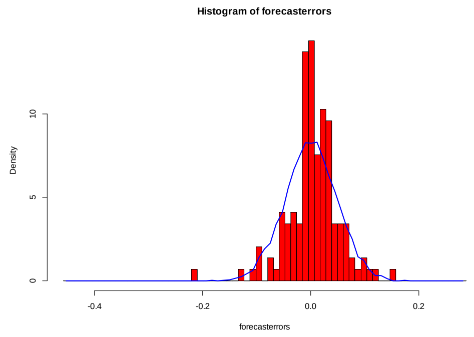
We expect Arima(0,0,1) to be the fittest model

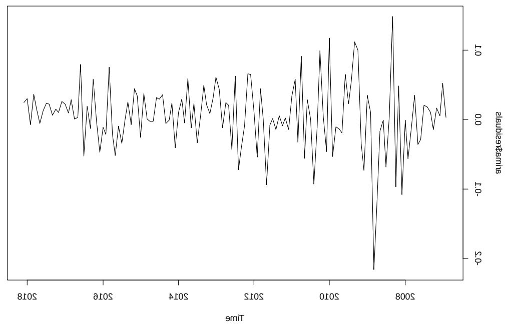
AIC optimization	BIC optimization
Series: A_Top_15	Series: A_Top_15
ARIMA(2,0,3) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ma1 ma2 ma3	ma1
0.7418 -0.6538 -0.5157 0.5295 0.3609	0.2731
s.e. 0.1679 0.1808 0.1718 0.1690 0.0913	s.e. 0.0874
sigma^2 estimated as 0.002208: log likelihood=223.35	sigma^2 estimated as 0.00235: log likelihood=217.52
AIC=-434.71 AICc=-434.05 BIC=-417.27	AIC=-431.03 AICc=-430.94 BIC=-425.22

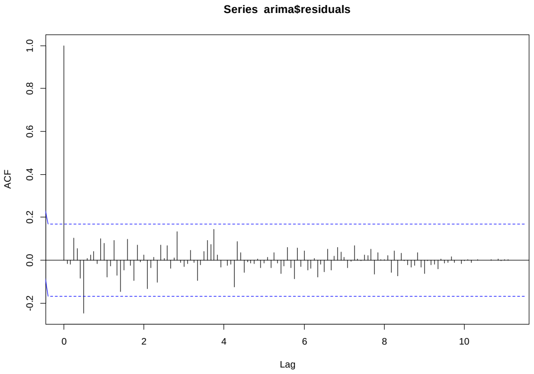
Given the autocorrelation graphs, the ARIMA(0,0,1) is the model that better represent the series.

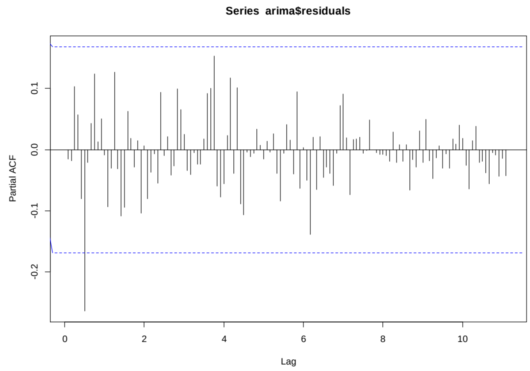
Therefore, we will use a bootstrap length=1.

Residuals:





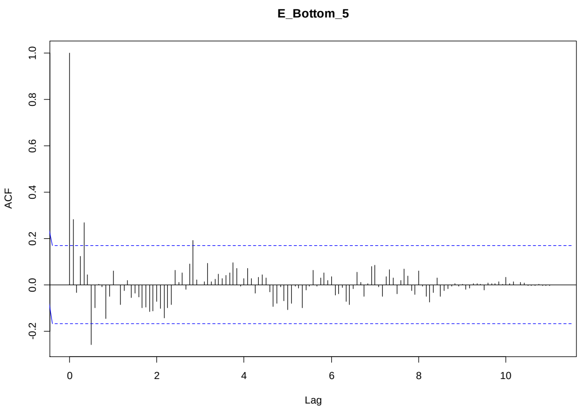




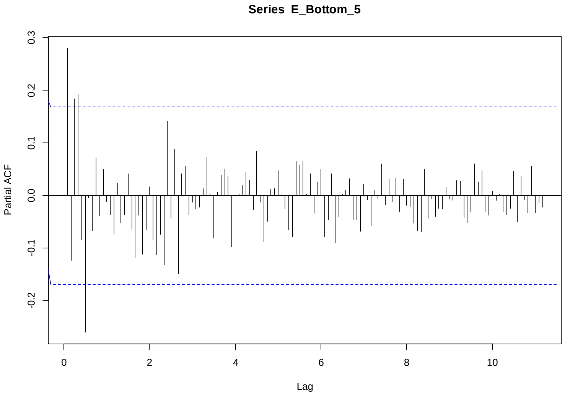
Europe_Bottom_5

Tau3 (Test Value)	-7.8171
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.812
z.lag.1 P(>|t|)	1.67e-12 ***
tt P(>|t|)	0.385
z.diff.lag P(>|t|)	0.149

The series is stationary with no drift and no trend.



Total Autocorrelation

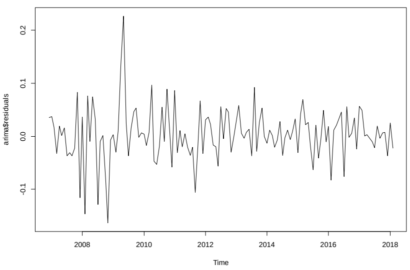


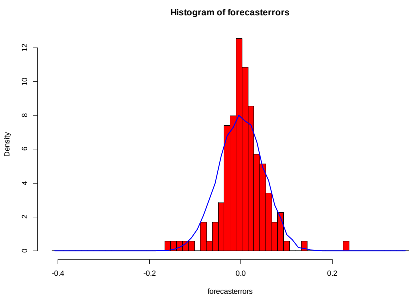
Partial Autocorrelation

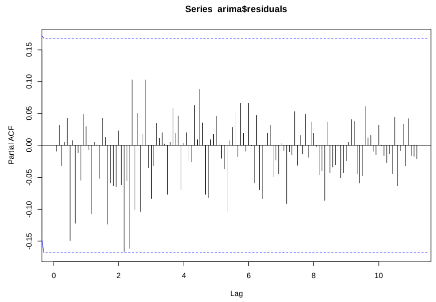
Acf cuts off after lag 1, but it is relevant also for lags 4 and 6. Pacf also cuts off after lag=1, and it is relevant also for lags 3,4 and 6. Therefore, we would expect an Arima(3,0,1), that will reflect both the total autocorrelation at lag 1 and the partial autocorrelation that is relevant until lag 4 (but with a non relevance value for lag=2).

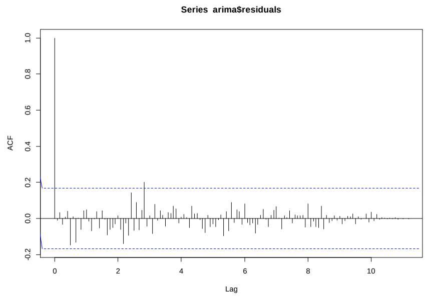
AIC optimization	BIC optimization
Series: E_Bottom_5	Series: E_Bottom_5
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.8204 -0.8671 0.4373 -0.5150 0.5927	0.3524
s.e. 0.1715 0.1345 0.0770 0.1879 0.1409	s.e. 0.0830
sigma^2 estimated as 0.002584: log likelihood=212.89	sigma^2 estimated as 0.002806: log likelihood=205.52
AIC=-413.79 AICc=-413.13 BIC=-396.36	AIC=-407.04 AICc=-406.94 BIC=-401.22

The model ARIMA(3,0,2) is the one that better fits the series. Indeed, given the autocorrelation graphs, it would be reductive to use a ARIMA(0,0,1), that would fail to account for the relevant partial autocorrelation also at lags 3 and 4. Therefore, ARIMA(3,0,2) seems like a good compromise. We will use a bootstrap block length = 3.





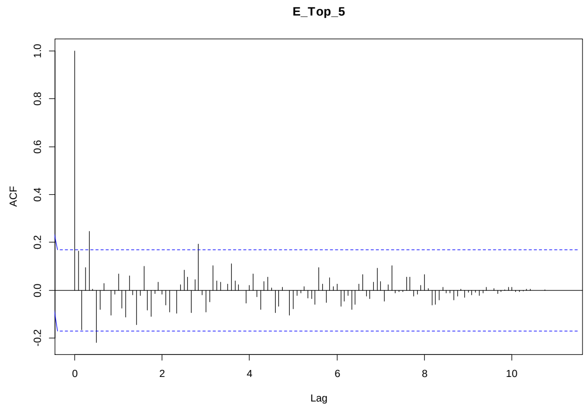




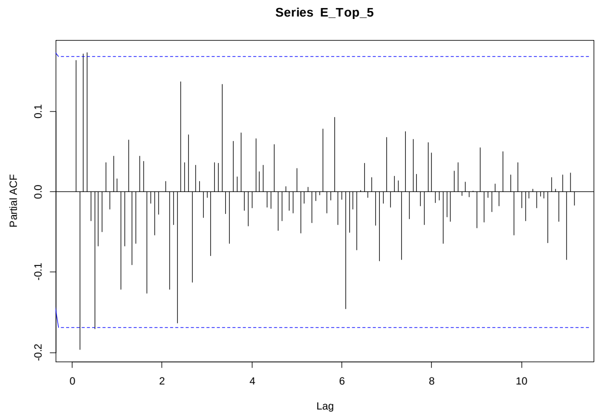
Europe_Top_5

Tau3 (Test Value)	-9.0773
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5698
z.lag.1 P(>|t|)	1.62e-15 ***
tt P(>|t|)	0.2600
z.diff.lag P(>|t|)	0.0194 *

The series is stationary with no drift and no trend.



Total Autocorrelation

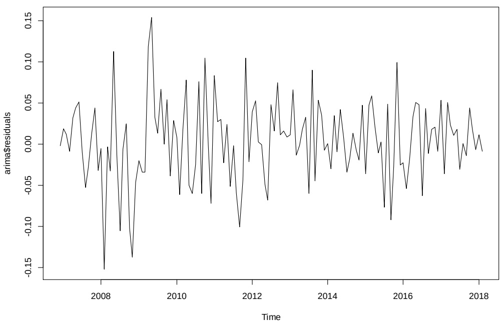


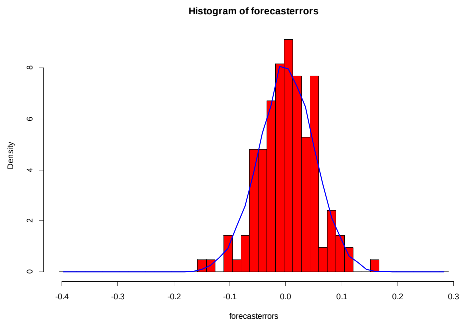
Partial Autocorrelation

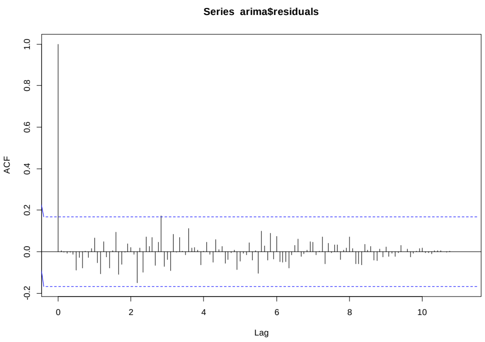
Acf is not relevant in lag=1 and 2 and it is relevant for lags 4 and 6. Pacf is relevant for lags 2 and 3. We expect a model (3,0,2).

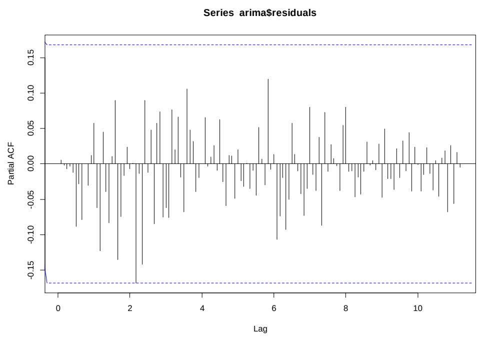
AIC optimization	BIC optimization
Series: E_Top_5	Series: E_Top_5
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7471 -0.7902 0.3559 -0.5590 0.5329	0.2677
s.e. 0.2118 0.1597 0.0824 0.2271 0.1757	s.e. 0.0994
sigma^2 estimated as 0.002628: log likelihood=211.81	sigma^2 estimated as 0.002836: log likelihood=204.83
AIC=-411.63 AICc=-410.97 BIC=-394.2	AIC=-405.66 AICc=-405.57 BIC=-399.85

The Arima (3,0,2) better fits what we observed in the graphs. Block bootstrap length = 3.





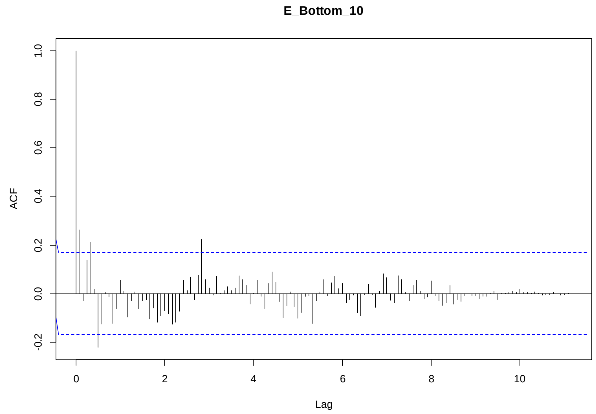




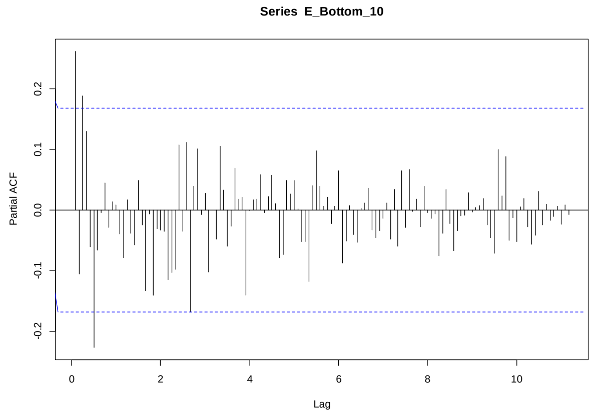
Europe_Bottom_10

Tau3 (Test Value)	-7.825
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.704
z.lag.1 P(>|t|)	1.6e-12 ***
tt P(>|t|)	0.317
z.diff.lag P(>|t|)	0.208

The series is stationary with no drift and no trend.



Total Autocorrelation

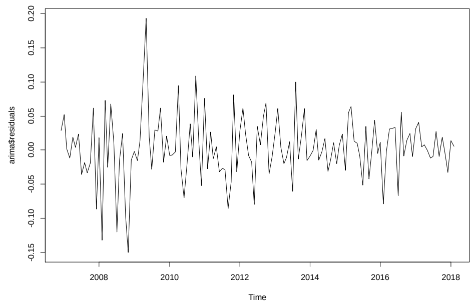


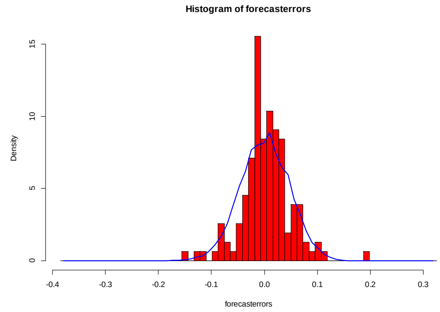
Total Autocorrelation

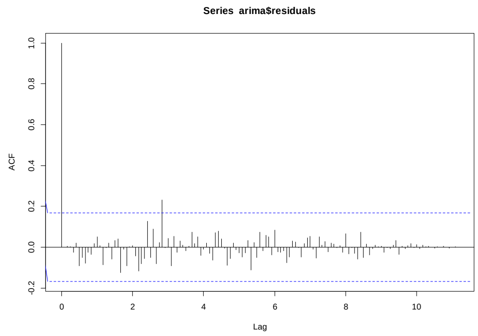
Therefore, we expect ARIMA(3,0,1) to be the fittest model

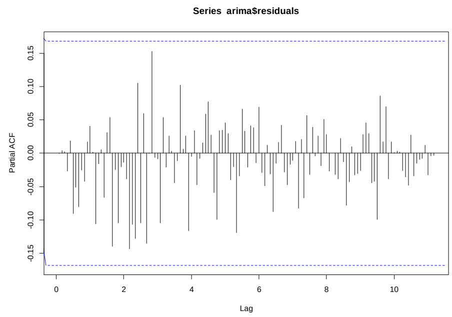
AIC optimization	BIC optimization
Series: E_Bottom_10	Series: E_Bottom_10
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.8052 -0.8586 0.4066 -0.5293 0.6210	0.3336
s.e. 0.1676 0.1572 0.0781 0.1772 0.1658	s.e. 0.0864
sigma^2 estimated as 0.002275: log likelihood=221.52	sigma^2 estimated as 0.00242: log likelihood=215.5
AIC=-431.03 AICc=-430.38 BIC=-413.6	AIC=-427 AICc=-426.91 BIC=-421.19

The Arima(3,0,2) better fits what we observed in the graphs. Block bootstrap length = 3.





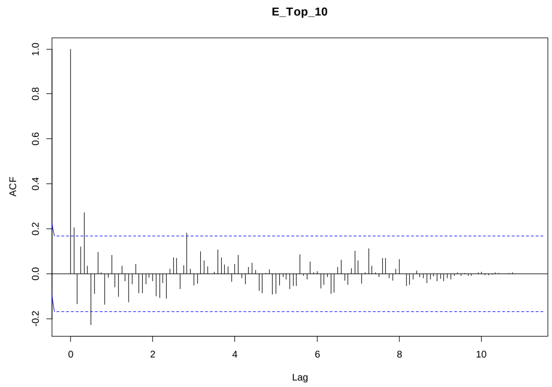




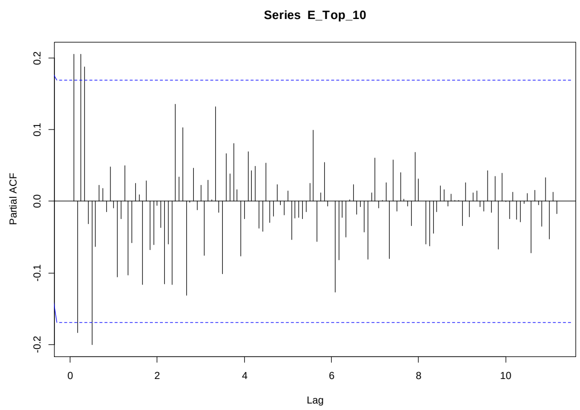
Europe_Top_10

Tau3 (Test Value)	-8.7815
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.505
z.lag.1 P(>|t|)	8.45e-15 ***
tt P(>|t|)	0.183
z.diff.lag P(>|t|)	0.027 *

The series is stationary with no drift and no trend.



Total Autocorrelation



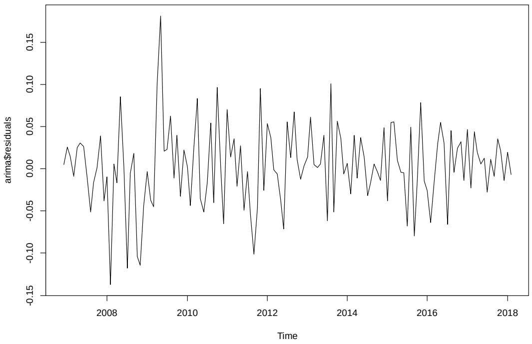
Partial Autocorrelation

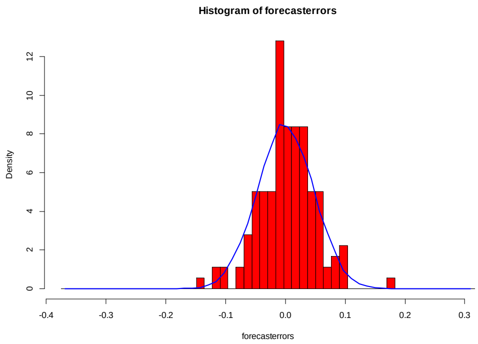
Therefore we expect Arima(3,0,1) to be the fittest model

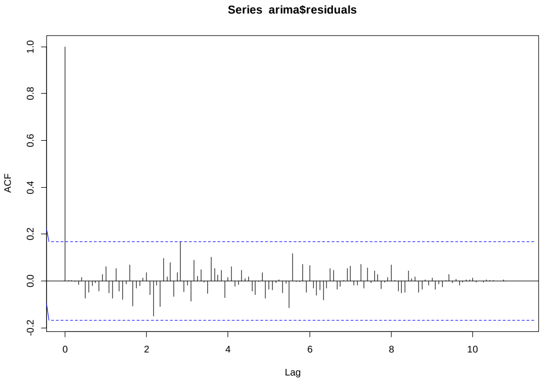
AIC optimization	BIC optimization
Series: E_Top_10	Series: E_Top_10
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7699 -0.8344 0.405 -0.5379 0.5685	0.3195
s.e. 0.1704 0.1415 0.079 0.1840 0.1584	s.e. 0.0936
sigma^2 estimated as 0.002342: log likelihood=219.52	sigma^2 estimated as 0.002573: log likelihood=211.37
AIC=-427.04 AICc=-426.39 BIC=-409.61	AIC=-418.74 AICc=-418.65 BIC=-412.93

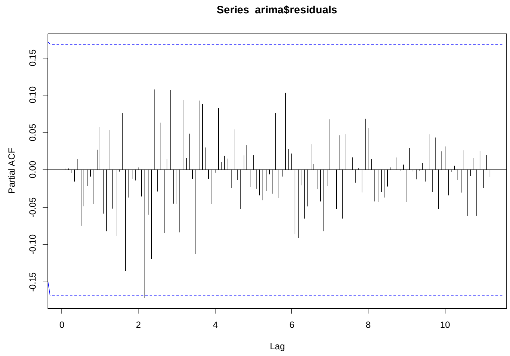
Therefore we choose Arima(3,0,2). Bootstrap length =3.

Residuals:





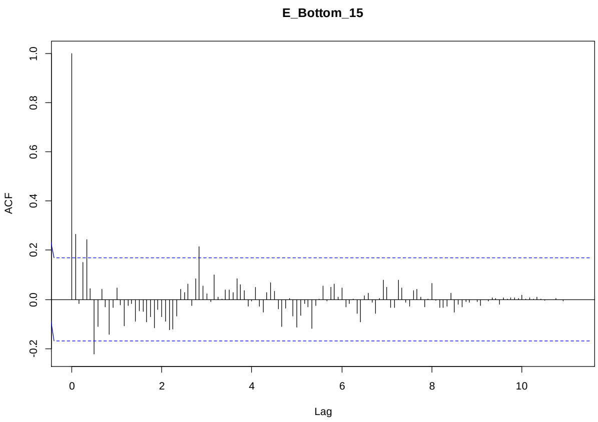




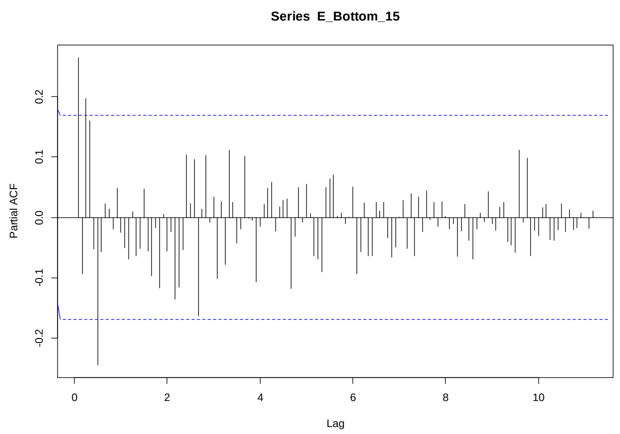
Europe_Bottom_15:

Tau3 (Test Value)	-7.7149
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.704
z.lag.1 P(>|t|)	2.9e-12 ***
tt P(>|t|)	0.310
z.diff.lag P(>|t|)	0.262

The series is stationary with no drift and no trend.



Total Autocorrelation

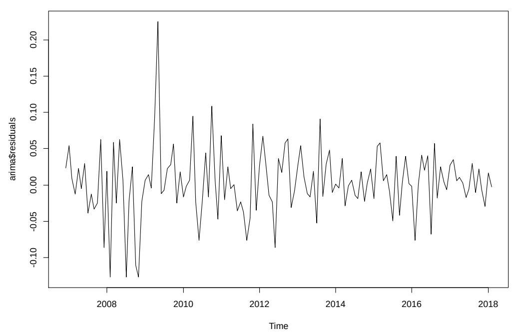


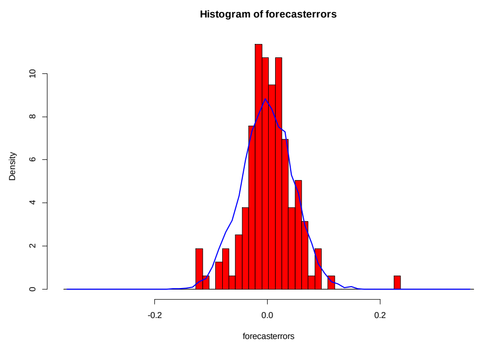
Partial Autocorrelation

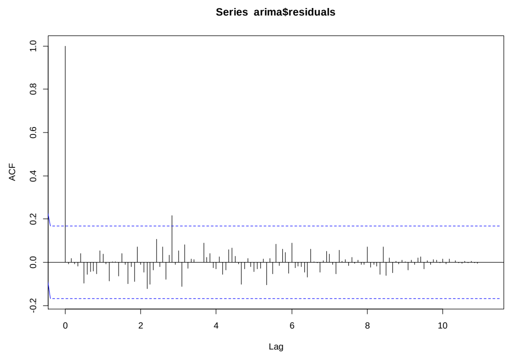
AIC optimization	BIC optimization
Series: E_Bottom_15	Series: E_Bottom_15
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.8206 -0.8452 0.4204 -0.5475 0.6071	0.3283
s.e. 0.1641 0.1473 0.0773 0.1754 0.1515	s.e. 0.0855
sigma^2 estimated as 0.002288: log likelihood=221.11	sigma^2 estimated as 0.002471: log likelihood=214.11
AIC=-430.22 AICc=-429.56 BIC=-412.78	AIC=-424.21 AICc=-424.12 BIC=-418.4

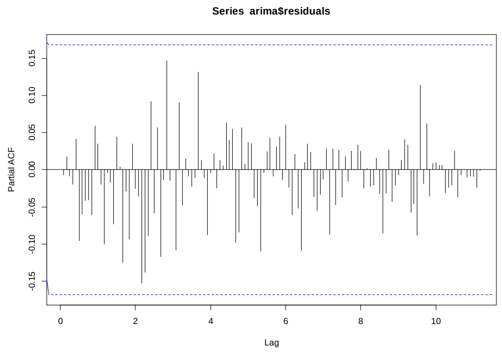
We accept ARIMA(3,0,2). Bootstrap length = 3.

Residuals:





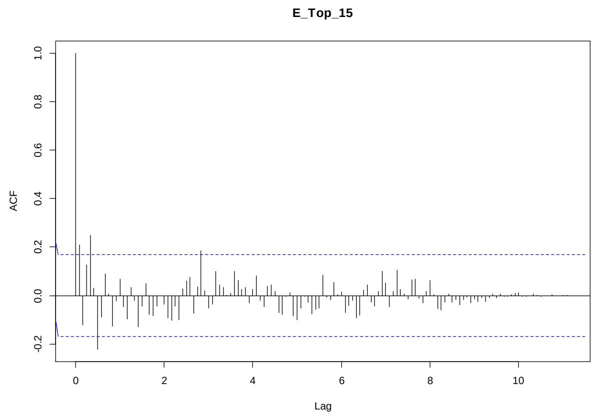




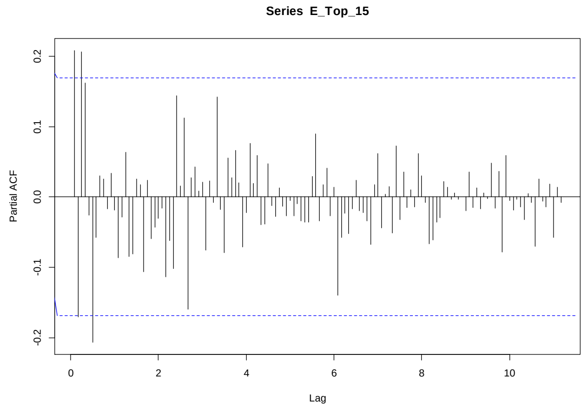
Europe_Top_15:

Tau3 (Test Value)	-8.6577
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.5729
z.lag.1 P(>|t|)	1.68e-14 ***
tt P(>|t|)	0.1906
z.diff.lag P(>|t|)	0.0388 *

The series is stationary with no drift and no trend.



Total Autocorrelation



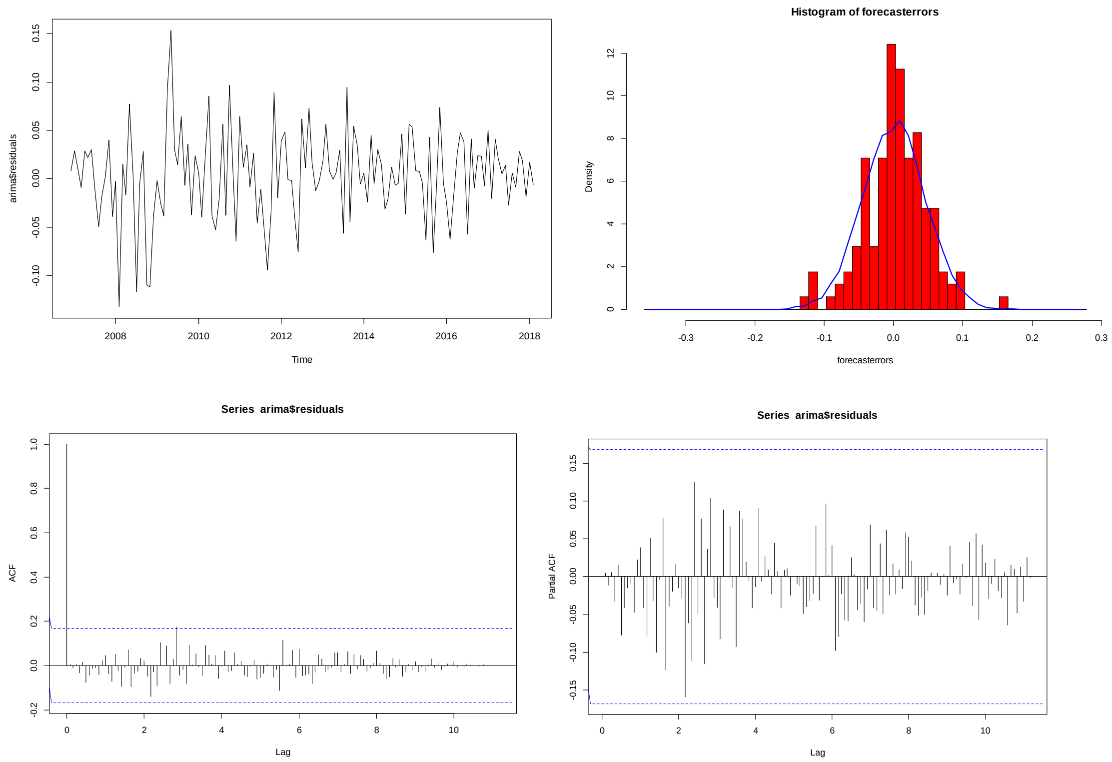
Partial Autocorrelation

Acf cuts off at lag=1, while pacf at lag=3.

AIC optimization	BIC optimization
Series: E_Top_15	Series: E_Top_15
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7805 -0.8501 0.4012 -0.5500 0.6008	0.3204
s.e. 0.1657 0.1544 0.0785 0.1781 0.1763	s.e. 0.0940
sigma^2 estimated as 0.002145: log likelihood=225.47	sigma^2 estimated as 0.002336: log likelihood=217.89
AIC=-438.93 AICc=-438.27 BIC=-421.5	AIC=-431.78 AICc=-431.69 BIC=-425.97

ARIMA(3,0,2) is the model that we will use. Bootstrap block length = 3.

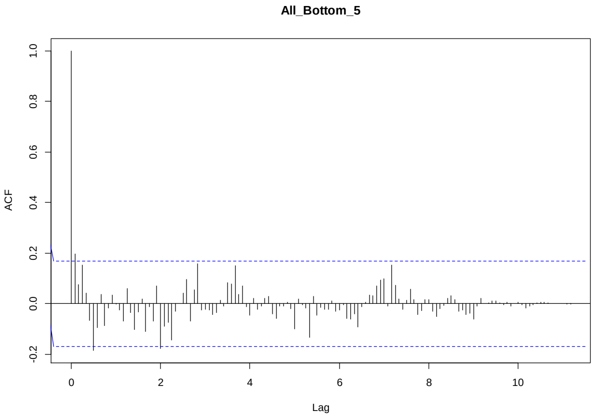
Residuals:



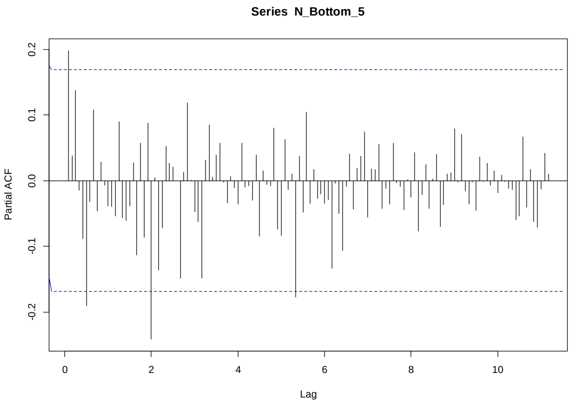
North_America_Bottom_5:

Tau3 (Test Value)	-6.9658
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.550
z.lag.1 P(>|t|)	1.5e-10 ***
tt P(>|t|)	0.554
z.diff.lag P(>|t|)	0.688

The series is stationary with no drift and no trend.



Total Autocorrelation



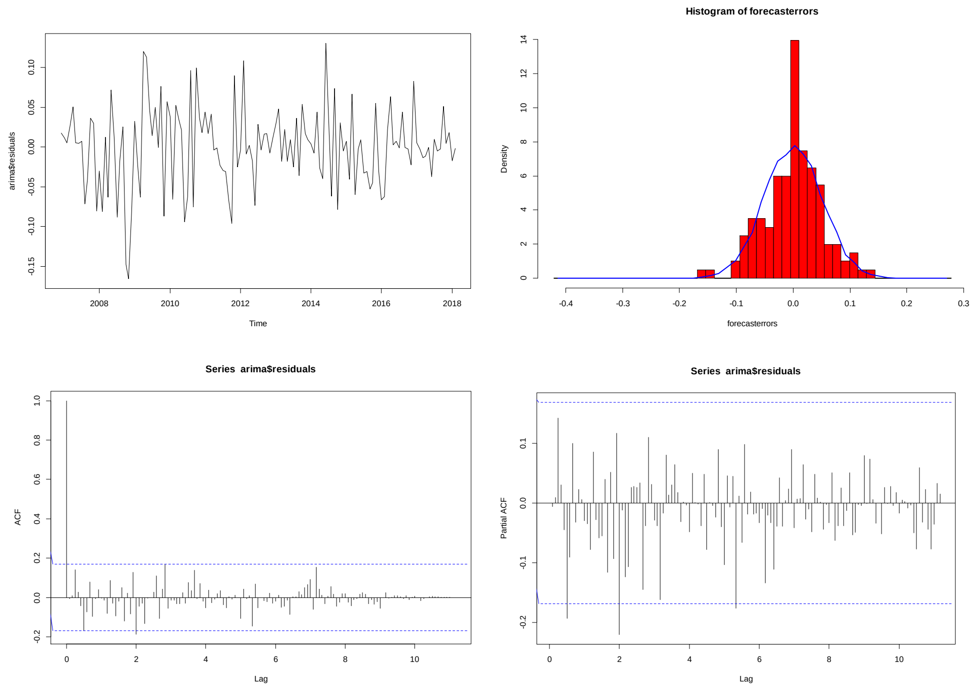
Partial Autocorrelation

Both Acf and pacf cut off after lag=1. We expect an ARIMA(0,0,1) model or AIMRA(1,0,0) or ARIMA(1,0,1)

AIC optimization	BIC optimization
Series: N_Bottom_5	Series: N_Bottom_5
ARIMA(1,0,0) with non-zero mean	ARIMA(1,0,0) with non-zero mean
Coefficients:	Coefficients:
ar1 mean	ar1 mean
0.1968 0.0135	0.1968 0.0135
s.e. 0.0840 0.0054	s.e. 0.0840 0.0054
sigma^2 estimated as 0.002627: log likelihood=210.5	sigma^2 estimated as 0.002627: log likelihood=210.5
AIC=-415 AICc=-414.82 BIC=-406.29	AIC=-415 AICc=-414.82 BIC=-406.29

Both optimization processes identify in ARIMA(1,0,0) the fittest model for the time series. We therefore will use a bootstrap block length =1.

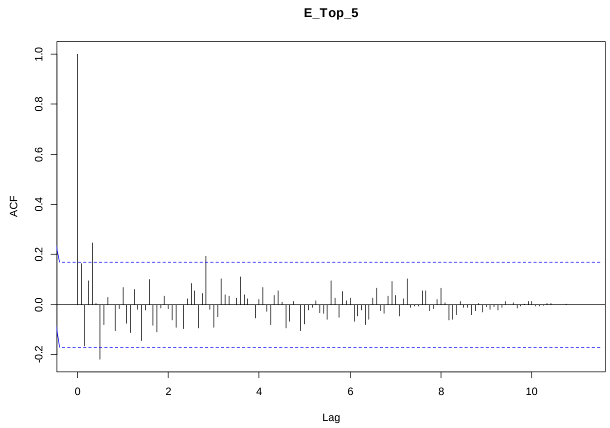
Residuals:



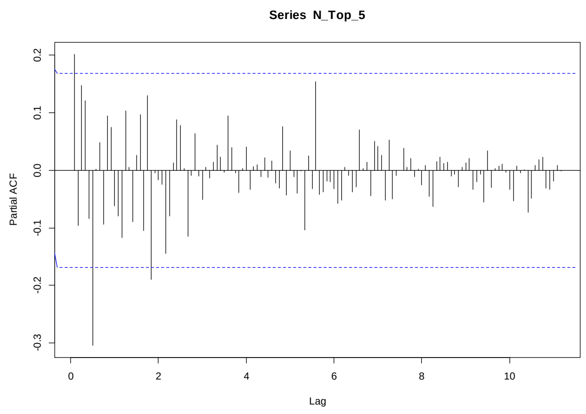
North_America_Top_5:

Tau3 (Test Value)	-8.0664
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.822
z.lag.1 P(>|t|)	4.34e-13 ***
tt P(>|t|)	0.164
z.diff.lag P(>|t|)	0.219

The series is stationary with no drift and no trend.



Total Autocorrelation

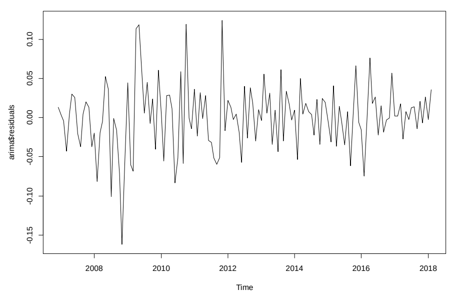


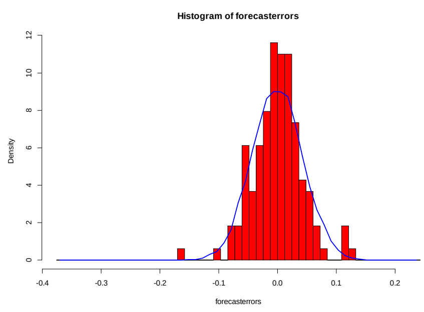
Partial Autocorrelation

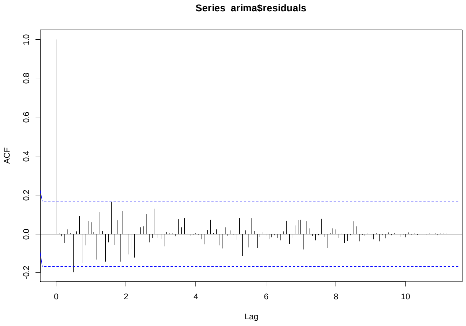
The acf shows a non relevant autocorrelation until lag=3, while relevant in lag=4. The pacf shows a cut off after lag=4, with a non relevant autocorrelation at lag=1.

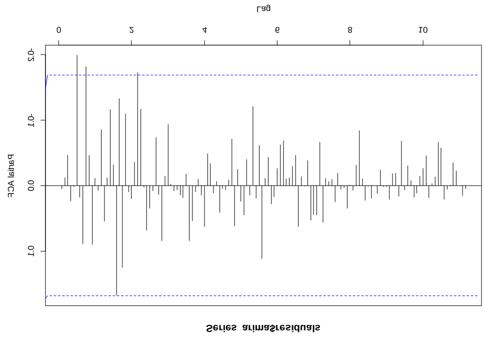
AIC optimization	BIC optimization
Series: N_Top_5	Series: N_Top_5
ARIMA(2,0,3) with non-zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ma1 ma2 ma3 mean	ma1
0.8240 -0.6099 -0.6368 0.4278 0.3488 0.0091	0.2721
s.e. 0.1727 0.1761 0.1702 0.1733 0.0853 0.0053	s.e. 0.0886
sigma^2 estimated as 0.001898: log likelihood=234.21	sigma^2 estimated as 0.002046: log likelihood=226.86
AIC=-454.41 AICc=-453.53 BIC=-434.07	AIC=-449.72 AICc=-449.63 BIC=-443.91

Given the presence of autocorrelations statistically relevant up to lag=4 in both partial and total autocorrelation models, we decide to utilize ARIMA(2,0,3), that consider an higher number of autocorrelated lags. Block length = 3.





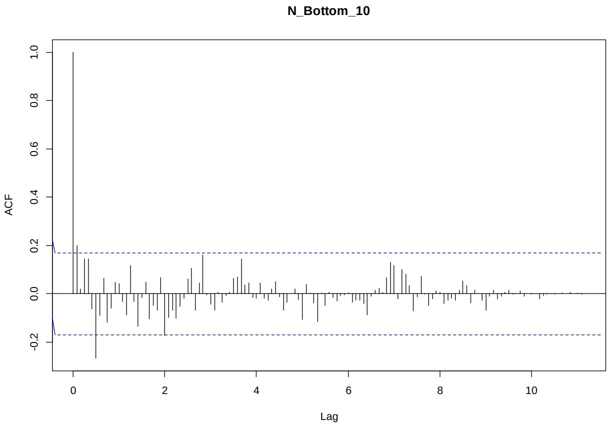




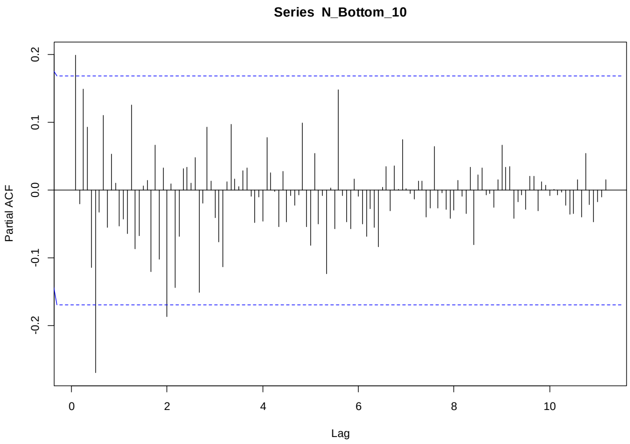
North_America_Bottom_10:

Tau3 (Test Value)	-7.3646
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.588
z.lag.1 P(>|t|)	1.87e-11 ***
tt P(>|t|)	0.524
z.diff.lag P(>|t|)	0.791

The series is stationary with no drift and no trend.



Total Autocorrelation



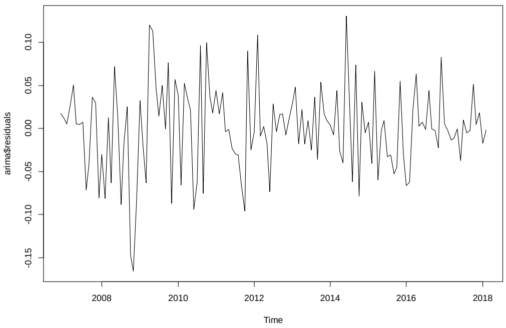
Partial Autocorrelation

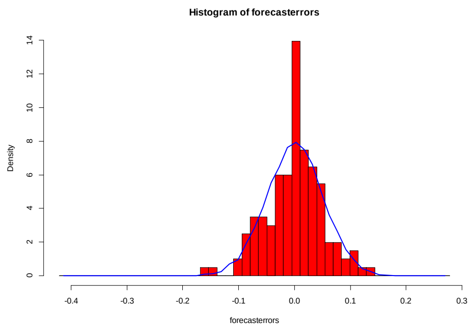
Both acf and pacf cuts off after lag=1.

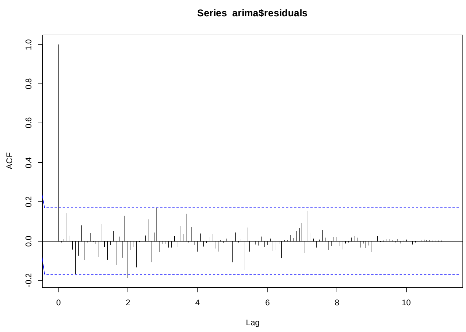
AIC optimization	BIC optimization
Series: N_Bottom_5	Series: N_Bottom_5
ARIMA(1,0,0) with non-zero mean	ARIMA(1,0,0) with non-zero mean
Coefficients:	Coefficients:
ar1 mean	ar1 mean
0.1968 0.0135	0.1968 0.0135
s.e. 0.0840 0.0054	s.e. 0.0840 0.0054
sigma^2 estimated as 0.002627: log likelihood=210.5	sigma^2 estimated as 0.002627: log likelihood=210.5
AIC=-415 AICc=-414.82 BIC=-406.29	AIC=-415 AICc=-414.82 BIC=-406.29

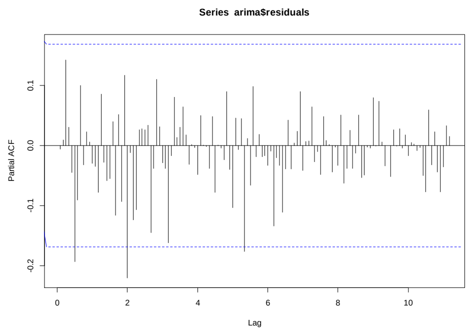
Therefore, we will use the ARIMA(1,0,0) model. Bootrasp block length = 1.

Residuals:





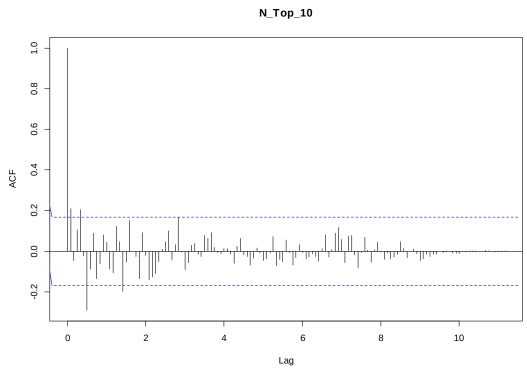




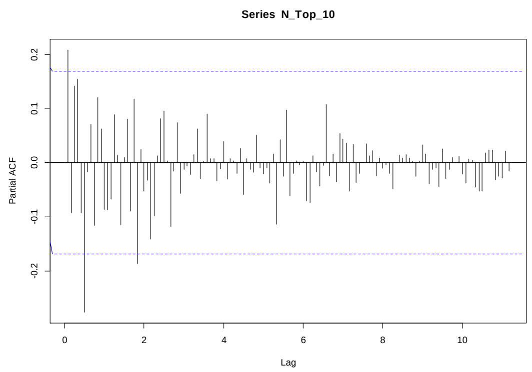
North_America_Top_10:

Tau3 (Test Value)	-7.9606
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.824
z.lag.1 P(>|t|)	7.7e-13 ***
tt P(>|t|)	0.226
z.diff.lag P(>|t|)	0.249

The series is stationary with no drift and no trend.



Total Autocorrelation



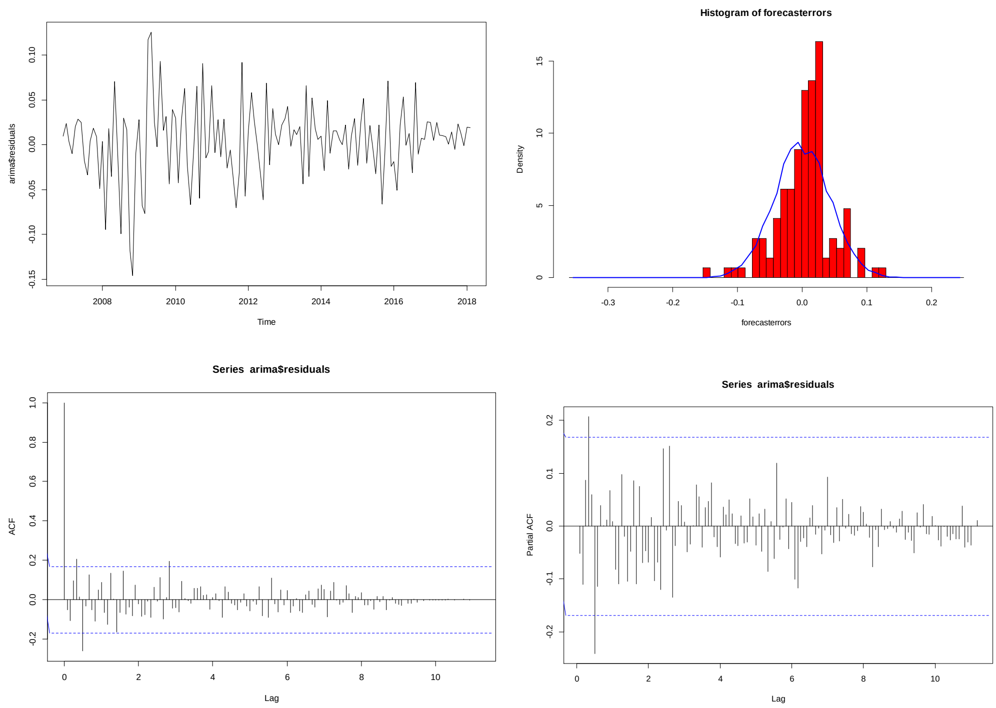
Partial Autocorrelation

Both Pacf and Acf cut off after lag = 1

AIC optimization	BIC optimization
Series: All_Top_10	Series: All_Top_10
ARIMA(3,0,2) with zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2	ma1
0.7574 -0.8143 0.4378 -0.4647 0.5415	0.3808
s.e. 0.2299 0.1602 0.0787 0.2688 0.1703	s.e. 0.0880
sigma^2 estimated as 0.001694: log likelihood=241.37	sigma^2 estimated as 0.001848: log likelihood=233.68
AIC=-470.75 AICc=-470.09 BIC=-453.31	AIC=-463.36 AICc=-463.27 BIC=-457.55

Given the weak autocorrelation shown by the graphs, we decide to use ARIMA(0,0,1). Block bootstrap length=1.

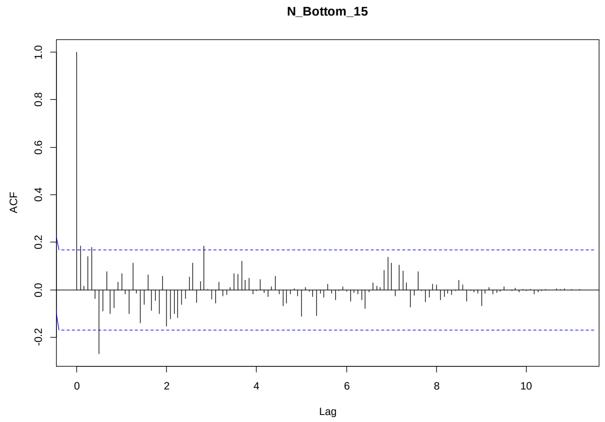
Residuals:



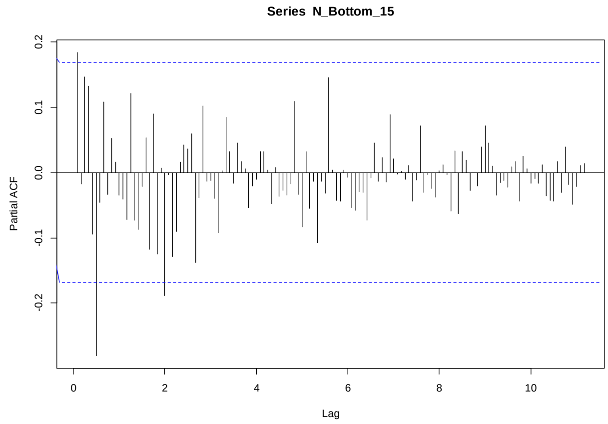
North_America_Bottom_15:

Tau3 (Test Value)	-7.4381
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.698
z.lag.1 P(>|t|)	1.27e-11 ***
tt P(>|t|)	0.440
z.diff.lag P(>|t|)	0.801

The series is stationary with no drift and no trend.



Total Autocorrelation



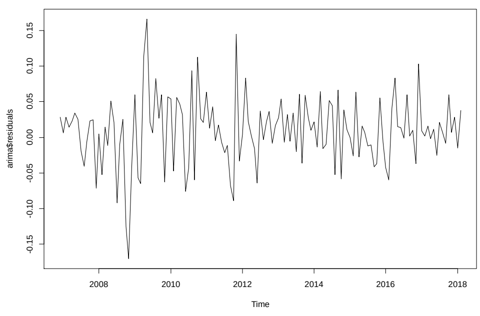
Partial Autocorrelation

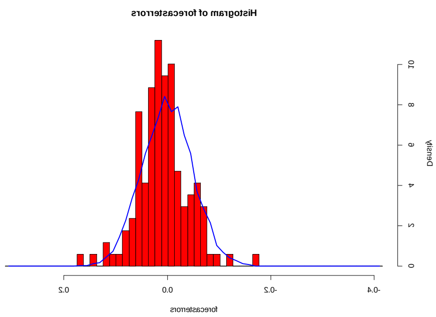
Both acf and pacf cuts off after lag =1.

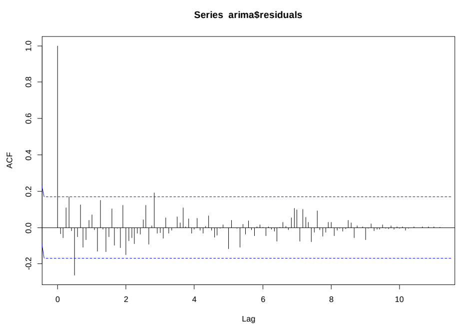
AIC optimization	BIC optimization
Series: N_Bottom_15	Series: N_Bottom_15
ARIMA(3,0,2) with non-zero mean	ARIMA(1,0,0) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2 mean	ar1
0.8328 -0.8362 0.3241 -0.6873 0.7446 0.0113	0.2225
s.e. 0.1423 0.1459 0.0839 0.1367 0.1302 0.0062	s.e. 0.0838
sigma^2 estimated as 0.002259: log likelihood=222.49	sigma^2 estimated as 0.002493: log likelihood=213.53
AIC=-430.98 AICc=-430.1 BIC=-410.64	AIC=-423.06 AICc=-422.97 BIC=-417.25

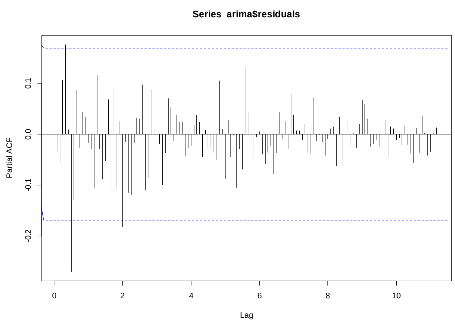
Given the weak autocorrelation shown by the graphs, we decide to use ARIMA(1,0,0). Block bootstrap length=1.

Residuals:





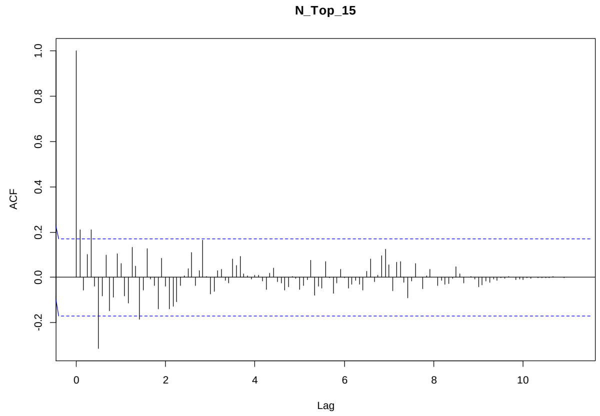




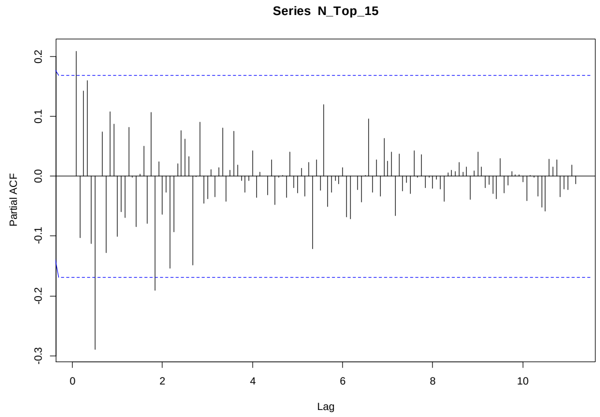
North_America_Top_15:

Tau3 (Test Value)	-8.0198
Tau3 (Critical Value 1%)	-3.99
Intercept P(>|t|)	0.950
z.lag.1 P(>|t|)	5 .58e-13 ***
tt P(>|t|)	0.275
z.diff.lag P(>|t|)	0.210

The series is stationary with no drift and no trend.



Total Autocorrelation

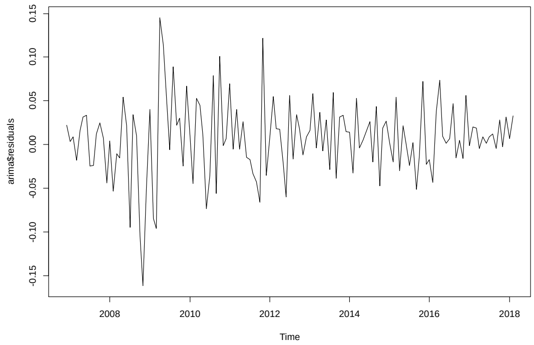


Partial Autocorrelation

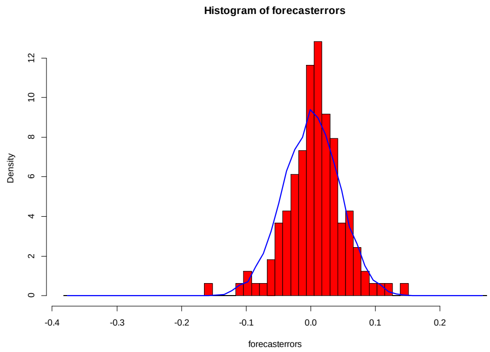
AIC optimization	BIC optimization
Series: N_Top_15	Series: N_Top_15
ARIMA(3,0,2) with non-zero mean	ARIMA(0,0,1) with zero mean
Coefficients:	Coefficients:
ar1 ar2 ar3 ma1 ma2 mean	ma1
0.0845 -0.9034 0.2240 0.1833 0.8328 0.0083	0.2733
s.e. 0.2286 0.0524 0.1261 0.2180 0.1123 0.0045	s.e. 0.0867
sigma^2 estimated as 0.001815: log likelihood=237.21	sigma^2 estimated as 0.001994: log likelihood=228.58
AIC=-460.42 AICc=-459.54 BIC=-440.0	AIC=-453.16 AICc=-453.07 BIC=-447.35

Bothy acf and pacf cut off at lag=1. Therefore, we choose ARIMA(0,0,1) and set bootstrap block length =1.

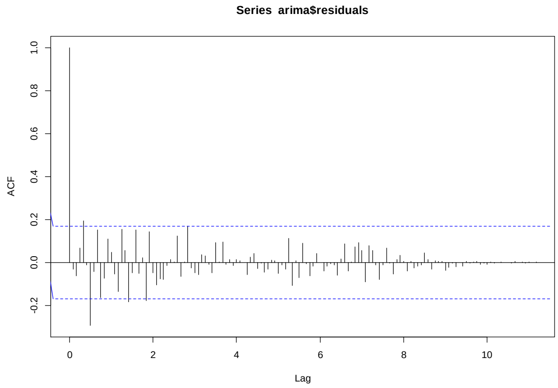
Residuals:



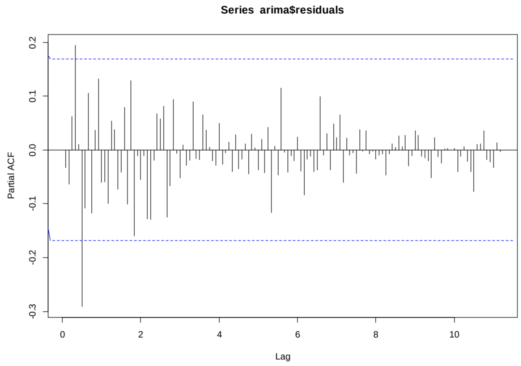
arima$residuals
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
2008
2010
2012
2014
2016
2018
Time



Histogram of forecasterrors
Density
0
2
4
6
8
10
12
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
forecasterrors



Series arima$residuals
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
2
4
6
8
10
Lag



Series arima$residuals
Partial ACF
0.2
0.1
0.0
-0.1
-0.2
-0.3
0
2
4
6
8
10
Lag