J Bus Ethics (2018) 151:665–692
https://doi.org/10.1007/s10551-016-3261-0

On the Price of Morals in Markets: An Empirical Study of the Swedish AP-Funds and the Norwegian Government Pension Fund

Andreas G. F. Hoepner[1,2] · **Lisa Schopohl**[1]

Received: 12 January 2016 / Accepted: 3 July 2016 / Published online: 19 July 2016

Abstract This study empirically analyses the exclusion of companies from investors' investment universe due to a company's business model (sector-based exclusion) or due to a company's violations of international norms (norm-based exclusion). We conduct a time-series analysis of the performance implications of the exclusion decisions of two leading Nordic investors, Norway's Government Pension Fund-Global (GPFG) and Sweden's AP-funds. We find that their portfolios of excluded companies do not generate an abnormal return relative to the funds' benchmark index. While the exclusion portfolios show higher risk than the respective benchmark, this difference is only statistically significant for the case of GPFG. These findings suggest that the exclusion of the companies generally does not harm funds' performance. We interpret these findings as indicative that with exclusionary screening, as practiced by the sample funds, asset owners can meet the ethical objectives of their beneficiaries without compromising financial returns.

Keywords ESG · GPFG · AP-funds · Exclusions · Exclusionary screening · Divestment · Negative screening · Norm-based exclusions · Sector-based exclusions · SRI

✉ Lisa Schopohl
l.schopohl@icmacentre.ac.uk

Andreas G. F. Hoepner
a.hoepner@icmacentre.ac.uk

[1] Henley Business School, ICMA Centre, University of Reading, Reading RG6 6BA, UK

[2] Mistra Financial Systems, Stockholm School of Economics, Box 6501, Stockholm SE-113 83, Sweden

Introduction

Over the last few decades, the general public has increasingly become aware of the social, environmental and ethical impacts of the investment and financing decisions of large financial institutions. Through movements like Occupy Wall Street, the public is gradually calling into question the ability of these players to serve the economy and society as well as to act in the best interests of their ultimate beneficiaries (Blanc and Cozic 2012). This development coincides with the emergence of the idea that investors are indirectly responsible for the corporate misconduct of the companies they hold. Especially public pension funds and other large public asset owners, such as Sovereign Wealth Funds (SWF), have been openly accused of complicity when financing companies that are involved in unethical behaviour, including violations of human rights and labour rights, gross corruption and environmental pollution. This investor group is especially susceptible to public scrutiny as it invests large sums of state-owned assets for the benefit of the general public, the funds' ultimate beneficiaries (Richardson 2011). On a global scale, this scrutiny has strongly increased since the outbreak of the financial crisis in 2008—although it is not a new development in the Nordic countries. Investments made by Norway's Government Pension Fund-Global (GPFG) and the Swedish AP-funds regularly make the headlines in the media.[1] The GPFG is the SWF of Norway,

[1] For example, the Swedish AP-funds have been heavily criticised for their holdings in Total AS, after the company was incriminated of corruptive practices and collaboration with the dictatorship in Burma (Bengtsson 2008b). Similarly, the Norwegian GPFG has come under attack for owing shares in the mining company POSCO after allegations against the firm emerged regarding its involvement in human rights violations in India (Meller and Husson-Trarore 2013).

established to invest the revenues from Norway's oil and gas exploration with the objective to ensure the long-term wealth of current and future generations of Norwegians (Richardson 2011; Jensen 2016b). With assets worth almost USD 900 billion, it is one of the largest SWFs in the world.[2] Although slightly smaller in size, the Swedish AP-funds which constitute the national pension system of Sweden also rank among the largest global asset owners (Severinson and Stewart 2012).

One reaction of these investors to the increased scrutiny is divesting from companies associated with unethical behaviour. For example, following several instances where the Norwegian GPFG attracted attention for holding companies involved in the production of controversial weapons and tobacco, the Norwegian Government devised ethical guidelines to ban these investments, together with investments in companies that contribute to serious human rights violations, severe environmental damage, gross corruption and other particularly serious violations of fundamental ethical norms.[3] The Swedish AP-funds have similar guidelines that require them to consider the ethical and environmental implications of their investments (Sandberg et al. 2014; Du Rietz 2016).

The growing popularity of exclusionary screening by large institutional investors appears to be in contrast to the general consent in the academic literature on socially responsible investment (SRI) which positions that exclusionary screening is an outdated approach.[4] This literature argues that SRI has moved on to more sophisticated strategies, such as active ownership and engagement as well as positive screening and best-in-class investing (e.g. Sparkes and Cowton 2004). In addition, a large part of the literature concludes that exclusionary screening and especially screening on industries that offer products and services considered as sinful and/or unethical financially hurts investors as these "sin" stocks tend to offer superior financial performance (e.g. Fabozzi et al. 2008; Adler and Kritzman 2008; Hong and Kacperczyk 2009). By excluding these firms from their investment universe, asset owners might forgo profitable investment opportunities "and thereby sacrifice vast sums of wealth through time" (Adler and Kritzman 2008, p. 55). This finding poses a potential conflict between the ethical and financial objectives of these funds, given that their financial objective is traditionally interpreted as the duty to maximise beneficiaries' long-term wealth.

This study attempts to address the question whether a conflict truly exists between the ethical and financial expectations faced by these asset owners. In other words, can the funds incorporate the ethical views of their beneficiaries without sacrificing financial returns? To answer this question, we focus on one particular SRI approach that is aimed at reducing investor's exposure to unethical business practices: exclusionary screening. In particular, we analyse the performance implications of the exclusion decisions by the Norwegian GPFG and the Swedish AP-funds. These funds exclude companies either due to the unethical nature of the sector that the company operates in (sector-based exclusions) or due to the company's involvement in violations of ethical standards and norms (norm-based exclusions). Our results suggest that the excluded companies neither significantly under- nor outperform relative to the funds' performance benchmarks. These findings hold for the entire portfolio of excluded companies and when separating the performance effect by reason for exclusion. We interpret these findings as evidence that by using specific forms of sector-based and norm-based screens asset owners can meet both, their beneficiaries' ethical and financial objectives.

Our study makes several important contributions to the academic literature on exclusionary screening. To the best of our knowledge, we are the first study to systematically analyse the performance effect of exclusionary screening by two of the leading institutional investor groups, i.e. public pension funds and SWFs. So far, the literature has either constructed theoretical portfolios by applying exclusionary criteria to a predefined investment universe (e.g. Adler and Kritzman 2008; Fabozzi et al. 2008; Hong and Kacperczyk 2009; Durand et al. 2013b; Salaber 2013; Trinks and Scholtens 2015) or it has analysed the performance of SRI mutual funds that apply exclusionary screening (e.g. Barnett and Salomon 2006; Renneboog et al. 2008b; Lobe and Walkshäusl 2011; Humphrey and Lee 2011; Capelle-Blancard and Monjon 2014; Humphrey and Tan 2014). While our finding of an insignificant performance effect is generally in line with findings derived from the SRI mutual fund literature, we contribute to this literature in several ways.

Firstly, as pointed out by Sparkes and Cowton (2004, p. 50), "the rapid growth in pension funds [and SWFs] that have adopted socially responsible criteria means that such

[2] The most recent market values of GPFG's assets can be obtained via the following homepage: http://www.nbim.no/.

[3] In early 2016, two new criteria have been included in GPFG's guidelines. One criterion targets conduct resulting in unacceptable greenhouse gas emissions at the aggregate company level. The other criterion, a sector-based screen, focuses on mining companies and energy producers with 30 % or more of revenues from thermal coal (Norwegian Ministry of Finance 2016). As these exclusion policies were introduced after the end of our sample period, we do not include them in our empirical analysis. The latest version of the guidelines for exclusion of companies from GPFG's portfolio can be found here: http://etikkradet.no/en/guidelines/.

[4] For the purpose of this study, we apply the broad definition of SRI used in Renneboog et al. (2008a, b) and Scholtens and Sievänen (2013). They define SRI as a way by which investors account for environmental, social, governance (ESG) and ethical issues in the investment process.

research can no longer be regarded as representative". Secondly, public asset owners have a considerably different relation to their beneficiaries than mutual funds (Richardson 2011). Not only do they invest on behalf of a far larger stakeholder group with non-uniform interests and ethical standards (Bengtsson 2008a, b; Richardson 2011), the ultimate beneficiaries of these funds also do not have the option to "exit" the fund, in case they do not agree with the fund's investment objectives and/or are not willing to bear potential costs of applying ethical standards (Clark 2004; Sandberg et al. 2014). As a consequence, the public scrutiny and societal pressures on these public asset owners are higher than for the average mutual fund (Blanc and Cozic 2012; Hawley 2016). Finally, the exclusions of GPFG and the Swedish AP-funds have a strong signalling effect on other global asset owners with many investors following their exclusion decisions (Bengtsson 2008a; Scholtens and Sievänen 2013; Jensen 2016b; Du Rietz 2016). Such domino effects of exclusion decisions are hardly observed for SRI mutual funds, rendering the exclusions of the investors studied in our sample of greater importance to the overall financial markets as well as to the corporations that are being excluded.

Our study also contributes to the emerging literature on norm-based screening. This practice of divesting from companies based on the company's association to violations of international norms is said to have originated in Scandinavia but it increasingly gains momentum among other large asset owners (Blanc and Cozic 2012; Du Rietz 2016). Currently three studies explicitly address norm-based screening. Capelle-Blancard and Monjon (2014) study French SRI mutual funds and contrast the performance differences between funds applying sector-related screens and norm-based screens. The studies by Blanc and Cozic (2012) and Meller and Husson-Traore (2013) compare the application of norm-based screening across European asset owners, however, without addressing the performance effects of such exclusions. Thus, we are the first to study the performance impact of norm-based screening by large public asset owners.

Besides these conceptual contributions, we also address some of the methodological concerns of previous studies on exclusionary screening. Previous research on exclusionary screening has either been criticised for neglecting real-world investment restrictions (see the criticism by Adamsson and Hoepner 2015, and Hoepner and Zeume 2014) or the inability to disentangle the performance effect of the exclusionary screening from other fund-specific factors such as manager skill (see Humphrey and Tan 2014). In comparison, by looking at the exclusion lists of GPFG and the AP-funds we are able to exactly identify the excluded companies, together with the reason and time of exclusion, thus enabling us to abstract from confounding fund-specific factors such as manager skill. At the same time, we automatically account for real-world investment restrictions by focussing on the funds' actual divestments.

The remainder of the study is structured as follows. "Literature Review" section provides an overview of the literature on the special role of the GPFG and the AP-funds in promoting ethical standards as well as on the performance effects of exclusionary screening. In "Research Questions and Hypotheses Development" section, we formulate the research questions and develop testable hypotheses. "Data and Methodology" section introduces the data and methodology used for testing the performance implications of the exclusion decisions of the GPFG and the AP-funds, while "Results" section presents the results of the empirical analysis and a discussion on the performance impact of exclusionary screening. We test the robustness of our findings in "Robustness Tests" section. "Conclusion" section draws the main conclusions based on the findings and discusses the implications of our findings.

Literature Review

The GPFG and the AP-funds: Balancing Ethical and Financial Objectives

Compared to other major financial markets such as the U.S. or the U.K., relatively little research exists on the Scandinavian SRI market and its major players. Notable exceptions include the studies by Bengtsson (2008a, b) and Scholtens and Sievänen (2013) which analyse the historical development of SRI and its drivers in the Scandinavian market. More closely related to our study, Sandberg et al. (2014) compare the legal environment regarding SRI in Sweden with the fiduciary duty concept in Anglo-American countries and particularly focus on the conflicting expectations faced by the Swedish AP-funds regarding their beneficiaries' financial and ethical interests, while Richardson (2011) discusses the tension between financial and ethical demands for the GPFG. In addition, Jensen (2016a, b) and Du Rietz (2016) provide overviews on the current state of the SRI development in Scandinavia as a whole, and in Norway and in Sweden in particular. Besides, several studies review the investment framework and policy guidelines of the GPFG (e.g. Clark and Monk 2010; Myklebust 2010; Chambers et al. 2012; Dimson et al. 2013) and the AP-funds (Severinson and Stewart 2012), touching on topics of SRI and the funds' particular duties as public asset owners. Yet, no study explicitly analyses the performance implications of the SRI approaches adopted by the GPFG and the AP-funds, especially regarding their most prominent feature, their exclusion

policies.[5] The following section reviews the above studies while focusing on the funds' special relation to their beneficiaries which distinguish public asset owners from other market participants such as SRI mutual funds. We also show that the demands from beneficiaries have been the primary driver to adapt exclusionary screening.

As highlighted in Richardson (2011, p. 22f.), "SWFs [such as the GPFG and other large public asset owners like the AP-funds] resemble institutional chameleons in the conflicting expectations they face. They operate like private investment vehicles for maximising shareholder value, while encumbered with public responsibilities to fulfil the ethical policies of their state". In terms of their financial objectives, both funds are expected to maximise long-term financial returns. The GPFG is required by the Norwegian Government to achieve a high return for the benefit of future generations, which is widely interpreted as the duty to maximise financial returns, within acceptable risk limits (Bengtsson, 2008a, b; Richardson 2011; Chambers et al. 2012; Dimson et al. 2013). In fact, GPFG has achieved an absolute return of 5.27 % per annum, i.e. a return of 0.51 % per annum in excess of its benchmark index, on its equity investments since its inception in 1998, which indicates that GFPG has been reasonably successful in achieving its financial objective.[6] Similarly, in case of the Swedish AP-funds, the National Pension Insurance Funds Act requires them to "manage fund assets in such a manner so as to achieve the greatest possible return" (cited according to Sandberg et al. 2014).[7] As such, the financial objectives of these funds are not different to those faced by most private market actors. However, due to their status as public asset owners, these funds are also obliged to fulfil the ethical standards expected from them by the general public. In case of the Swedish AP-funds, a legal requirement was introduced in 2001 that obliges the funds to consider ethical and environmental aspects in their investment policies and led the funds to establish a new investment policy that involves the exclusion of companies that are not in line with universally agreed ethical and environmental standards (Bengtsson 2008b; Sandberg et al. 2014).[8] GPFG's turn towards ethics started in 2002 with its first ethically motivated divestment and resulted in GPFG's implementation of a range of detailed ethical guidelines (Bengtsson 2008b; Jensen 2016b). The current version of the ethical guidelines restricts the fund from investing in companies that contribute to serious human rights violations, severe environmental damage, gross corruption and other particularly serious violations of fundamental ethical norms as well as in companies related to the production of tobacco and controversial weapons.

While the direct reason for the funds' move towards ethical exclusionary screening relates to legal changes, the governments themselves were responding to pressures from the public that did not want to see state assets invested in unethical business practices and thus act as accomplices to gross, systematic breaches of ethical norms (Bengtsson 2008b). In fact, both the Swedish AP-funds and Norway's GPFG named the avoidance of complicity and the appeal to public trust as main drivers for establishing their ethical investment policies of exclusionary screening (see Sandberg et al. 2014, for the AP-funds, and Richardson 2011, for the GPFG). In contrast to mutual fund investors, the beneficiaries of the GPFG and the AP-funds do not have the option to exit the funds, in case that they do not agree with the funds' investment objectives and/or are not willing to incur potential costs of applying ethical standards (Clark 2004; Sandberg et al. 2014).[9] They are rather "locked in" the funds and thus, they inevitably bear any potential costs of ethically motivated exclusionary screening. As the ultimate beneficiaries of these funds comprise both the state's current population as well as future generations (Bengtsson 2008a, b), reaching a consensus on one ethical perspective shared by all beneficiaries is rendered difficult,

[5] The only exception is an internal study by one of the Swedish national pension funds themselves. As stated in Sandberg et al. (2014, footnote on page 66), AP7 conducted an internal inquiry into the performance implications of its exclusionary screening practices. The results of this analysis suggest that the screened fund carried a marginally higher risk than a hypothetical unscreened portfolio, but did not show any significant difference in returns.

[6] Information on the (relative) equity performance of GPFG can be obtained via the following homepage: https://www.nbim.no/en/transparency/reports/2015/performance-and-risk-2015/.

[7] The first four AP-funds do not provide return figures on their equity performance relative to the benchmark. The absolute equity returns for AP1 over 2011–2015 is 6.9 % per annum, for AP2 11.1 % per annum, for AP3 13 % per annum and for AP4 10.7 % per annum, respectively. Compared to the 5-year return on the MSCI All Country World index which amounts to 7.3 % per annum, all but the AP1 fund outperformed this benchmark. AP7's equity portfolio earned an absolute return of 17.3 % per annum over the years 2010–2014, which represents an average return of −0.2 % in excess of its benchmark. Information on the AP-funds' equity performance is taken from the funds' annual reports which are available via the following homepages: for AP1 http://www.ap1.se/en/Financial-information-and-press/Reports/ , for AP2 http://www.ap2.se/en/Financial-information/financial-reports/ , for AP3 http://www.ap3.se/sites/english/financial_reports/Pages/default.aspx , for AP4 http://www.ap4.se/en/financial-reports-and-press/reports/ , for AP7 https://www.ap7.se/globalassets/kiidar/kiid-ap7-aktiefond-2015-06-23.pdf.

[8] While AP7 focuses only on exclusionary screening, AP1, AP2, AP3 and AP4 combine exclusionary screening with engagement and only exclude a company after engagement has proven unsuccessful.

[9] This point is also explicitly highlighted by the Graver Committee, an expert committee that has been appointed by the Norwegian Government to define ethical guidelines for GPFG, as "a defining characteristic of the Fund…that a substantial proportion of those on whose behalf the Fund is managed cannot choose its manager". The English version of the Report from the Graver Committee is available online: https://www.regjeringen.no/en/dokumenter/Report-on-ethical-guidelines/id420232/.

if not impossible.[10] To overcome this challenge and to assure a broad basis of support for their SRI decisions, both the Norwegian GPFG and the Swedish AP-funds decided to rely on national law and international standards to set out a minimum of ethical norms that they expect all the companies that they hold to abide to. The latter standards comprise the UN Global Compact, the OECD Guidelines for Corporate Governance and for Multinational Enterprises, labour standards set out by the International Labour Organization, as well as conventions that ban particular controversial weapons (Richardson 2011; Sandberg et al. 2014; Norwegian Ministry of Finance 2015).[11] Using this principle of finding the lowest common ethical factor, funds sought to account for their ethical obligation as public asset owners while at the same time minimising the financial impact to the beneficiaries of applying these ethical standards (Sandberg et al. 2014).

Performance Effects of Exclusionary Screening

Besides the literature on Scandinavian public asset owners, our study also contributes to the vast literature on the performance impact of exclusionary screening. Arguably, the most prominent study in this stream of the literature is by Hong and Kacperczyk (2009). In their study, the authors find that investing in 156 U.S. companies that operate in sectors related to alcohol, gambling and tobacco—the so-called triumvirate of sin—over the period 1965–2006 leads to a positive abnormal return relative to industry-comparable stocks. Many studies have since attempted to confirm or disprove the original results by Hong and Kacperczyk (2009) and have extended the original set of screens to reflect a broader range of societal norms. For instance, studies by Adler and Kritzman (2008), Durand et al. (2013a, b) and Trinks and Scholtens (2015) find support for an outperformance of sin stocks in the U.S. markets, Salaber (2013) for a European stock universe, Visaltanachoti et al. (2009) for China and Hong Kong, and Fabozzi et al. (2008) for a set of 21 global equity markets, respectively. However, there is also a considerable body of research that finds no or only an insignificant outperformance of sin stocks. For instance, Kempf and Osthoff (2007) and Statman and Glushkov (2009) find a positive but insignificant abnormal return, when applying six common sin screens to a U.S. stock universe over a 14-year and 16-year period, respectively. Similarly, Lobe and Walkshäusl (2011) and Adamsson and Hoepner (2015), looking at a global and U.S. set of sin companies, conclude that the performance of these stocks does not significantly differ from benchmark returns. In addition, several studies find that the extent to which investors shun sin stocks significantly varies across markets and that markets with more restrictive social norms show a stronger "sin" effect (e.g. Salaber 2013; Fauver and McDonald 2014; Liu et al. 2014; Adamsson and Hoepner 2015).[12] One aspect that the above studies have in common is that they test the performance implications of exclusionary screening by applying screening criteria (e.g. based on industry classifications) to a predefined investment universe. Thus, they construct theoretical, and in a sense "fictive", portfolios of excluded companies. "Fictive" as it is not clear whether any real-world investor actually applies these exact screens. While this approach allows dissecting the "sin" impact on performance, it has been criticised for neglecting real-world investment restrictions. In particular, Adamsson and Hoepner (2015) and Hoepner and Zeume (2014) argue that the significant outperformance of "sin" stocks found in large parts of the literature may disappear, once restricting the investment universe to stocks that are liquid and large enough to qualify as suitable investments for institutional investors.

A stream of the literature that overcomes this criticism comprises studies that analyse the performance of SRI mutual funds that apply exclusionary screens (Barnett and Salomon 2006; Renneboog et al. 2008b; Lee et al. 2010; Renneboog et al. 2011; Humphrey and Lee 2011; Capelle-Blancard and Monjon 2014; Humphrey and Tan 2014). In

[10] In the Report from the Graver Committee it explicitly says that "there is no consensus on one particular uniform ethical perspective". See also Richardson (2011) for a further discussion of this issue.

[11] See for example the statement by the Ethical Council of AP1, AP2, AP3 and AP4: "The Swedish Government's core values…find expression in those international conventions that Sweden has signed, which include conventions on the environment, human rights, labour law, corruption and inhumane weapons, as well as through the support given to initiatives such as the United Nations Global Compact and OECD guidelines for multinational companies, in addition to Sweden's own stance on international public law issues. In tandem with the Swedish Government's value system, international conventions constitute essential instruments for the Ethical Council in its efforts to ensure the AP Funds satisfactorily take into account environmental issues and ethical dimensions in their work". The statement is available online: http://etikradet.se/etikradets-arbete/vardegrund/?lang=en. Similar references for the GPFG can be found in the Graver Report on Ethical Guidelines and in the Report to the Storting (Norwegian Ministry of Finance 2016).

[12] Durand et al. (2013b) find that the institutional investors have lower holdings in sin stocks in markets that are culturally closer to the U.S. (i.e. Australia, New Zealand) than they do in markets with a larger cultural distance to the U.S. (i.e. Japan, South Korea). Salaber (2013) concludes that sin stocks earn higher returns in markets with a higher share of Protestant population, compared to Catholic orientations, while Liu et al. (2014) find a lower institutional ownership of sin stocks in regions with a higher share of Protestants. According to Fauver and McDonald (2014), sin stocks have a lower equity valuation in markets that consider these stocks more controversial, and vice versa. Adamsson and Hoepner (2015) map countries according to different cultural dimensions and find returns for sin stocks to differ across markets, although most of these differences are statistically insignificant.

contrast to the "sin" studies, the mutual funds literature does not analyse the performance of the excluded companies but it instead looks at the returns of the funds applying the exclusionary screens. While a large part of the literature concludes that screening mutual funds do not generally perform differently from their conventional peers (e.g. Lee et al. 2010; Humphrey and Lee 2011; Humphrey and Tan 2014), several studies show that the relation between screening and performance might be more complex and depends on several fund-specific factors. For instance, Barnett and Salomon (2006), Renneboog et al. (2008b), Lee et al. (2010) and Capelle-Blancard and Monjon (2014) find that the screening-performance relation depends both on the type of screens and the fund's screening intensity, as measured by the number of screens applied. In addition, Humphrey and Lee (2011) find that exclusionary screening can impact the risk characteristics of the funds. However, these studies come with their own methodological restrictions. As Humphrey and Tan (2014) point out, SRI mutual funds are very heterogeneous and might apply other SRI approaches or forms of active management. Thus, studying returns at the fund level does not allow distinguishing the performance contribution of the ethical screens from other fund-specific effects such as managerial skill.

While the review of the considerable body of literature on exclusionary screening might suggest that the performance impact of exclusionary screening is already well understood, we argue that the literature has predominantly focused on certain aspects of this problem while leaving others still mainly unexplored. To illustrate, when categorising the above studies based on the type of exclusionary screens, we find that most studies cover sector-based exclusions (e.g. Adler and Kritzman 2008; Fabozzi et al. 2008; Hong and Kacperczyk 2009; Trinks and Scholtens 2015; Salaber 2013; Humphrey and Tan 2014; Adamsson and Hoepner 2015), while currently three studies explicitly address norm-based screening (Blanc and Cozic 2012; Meller and Husson-Traore 2013; Capelle-Blancard and Monjon 2014). Thus, there is a clear need for further research on the performance impact of norm-based screening. In addition, none of the studies focuses on the performance implications of exclusionary screening by investors other than mutual funds, although the previous section has established that public asset owners are especially susceptible to public pressures to balance their ethical and financial objectives.

Research Questions and Hypotheses Development

The literature review highlights the ambiguous findings of the prior literature regarding the performance effects of exclusionary screening as well as the lack of research on exclusionary screening by public asset owners in general and on norm-based screening in particular. Given the special role of these funds within their state's society, shedding light on these unexplored topics is not only of relevance to the funds themselves but also to other global market participants, policy makers and the Norwegian and Swedish society. In our study we aim to fill these gaps by asking:

RQ1 What are the performance implications of exclusionary screening by the GPFG and the AP-funds?

Turning to the previous studies, we may generally expect three performance effects of applying these screens.

H1a The exclusion portfolios outperform the market.

The hypothesis of a significant outperformance of excluded "unethical" companies is mainly promoted by the early parts of the literature, especially the "sin stock" studies (Adler and Kritzman 2008; Fabozzi et al. 2008; Hong and Kacperczyk 2009). Relying on Merton's (1987) incomplete information model and related arguments of segmented capital markets (Derwall et al. 2011), these studies argue that norm-constrained investors such as pension funds and university endowment funds shun controversial stocks. This leads to limited risk sharing among those investors that hold the controversial companies and as a consequence, investors require higher returns for holding the stock. In addition, Fabozzi et al. (2008) argue that it is costly to implement and uphold social and environmental standards and hence compliance with these norms should decrease firm's profits. Especially if the cost of complying with the norms is higher than the costs of breaking the standards (e.g. litigation risks from being caught, reputational costs), non-compliant companies are expected to show higher future profits and cash flows. The asset pricing implications of these effects are formalised in Heinkel et al. (2001) who develop a theoretical model of the impact of exclusionary ethical investing on corporate behaviour in a risk-averse equilibrium setting. The authors conclude that the shunned firms should earn a positive abnormal return relative to the market, while "acceptable" firms are expected to underperform. However, it is important to note that this argument is based on the idea of a temporary undervaluation of the shunned stocks which is eventually corrected and thereby generates a positive abnormal return for investors holding the stocks.

H1b The exclusion portfolios underperform the market.

In contrast, the proponents of an underperformance effect of exclusion portfolios argue that the unethical companies are overvalued. They postulate that the market does not fully incorporate the risks that are associated with unethical corporate practices and breaches of international norms. For instance, Barnett and Salomon (2006) and Petersen and Vredenburg (2009) point out that these firms

are exposed to risks of negative government and/or social actions such as litigation risk, penalties and increased opposition from communities and local authorities regarding future investment projects. In addition, unethical companies could face reputational costs that might lead to a loss in customer and client loyalty and thus lower revenues or to higher employee turnover and a loss in competitiveness in corporate hiring (Barnett and Salomon 2006). Finally, the involvement in scandals could also signal bad managerial talent, exposing investors to greater management risk (Renneboog et al. 2008b). However, the GPFG and AP-funds only exclude a company after the breach has occurred and/or after the involvement in the unethical business practice has become public knowledge. Thus, for these risks to affect the funds' portfolio performance two potential channels are possible. On the one hand, the potential risks associated with the unethical business practices are not being properly priced in the market at the time of divestment. Thus, even if the funds only divest from the company after the incident has occurred they might still avoid some of the stock price decline as the market slowly learns about the true costs of the unethical practices. On the other hand, investors could regard past breaches of norms as a predictor of future incidents. Again assuming that the market does not account for this increased risk exposure, divestment could shield the funds from the negative financial consequences of future incidents. Considering that the GPFG and the AP-funds aim to only exclude companies that have a high risk of future breaches and that show no willingness to change their corporate practices (Richardson 2011; Sandberg et al. 2014), the latter channel may explain a potential underperformance of their excluded companies.

H1c The exclusion portfolios do not show significant performance differences compared to the market.

Finally, one might expect no significant performance effect of exclusionary screening (see e.g. the assessment of Kurtz 2005, based on a review of the long-term performance of social indices and SRI mutual funds). For one thing, the two previous hypotheses rely on the assumption of (partial) market inefficiency. However, if the market was efficient, it would instantaneously and correctly adjust the market price of stocks to reflect all material risks upon disclosure of the incident. Thus, divesting from the company after the incident has occurred should not lead to any abnormal performance difference relative to the market. In addition, one could expect an insignificant performance impact of exclusions if the funds consciously balance the financial and ethical expectations of their beneficiaries by only excluding companies if the exclusion does not harm fund performance. However, this line of argument relies on several critical assumptions. First, the funds would need to select exclusion targets from a set of unethical companies. This assumption is quite realistic as time and resource constraints provide a natural limit to the number of companies that the fund can investigate and engage with (Clark and Monk 2010). In addition, the argument assumes an implicit prioritisation of the financial objectives over the ethical objectives, which, as will be discussed in the Conclusion, cannot be regarded as given. And finally, the argument implies that the funds are able to correctly evaluate the future performance effect of their exclusion decisions.

So far, we have regarded the excluded companies as one homogeneous group. However, prior research indicates that performance effects may differ depending on the nature of the exclusionary screen (Barnett and Salomon 2006; Renneboog et al. 2008b; Capelle-Blancard and Monjon 2014; Trinks and Scholtens 2015). Looking at the case of the GPFG and the AP-funds, we can differentiate between sector-based exclusions and norm-based exclusions. Based on these differences, we pose a subordinate research question:

RQ2 Do the performance implications of exclusionary screening differ across different types of screens, especially regarding sector-based versus norm-based screens?

Norm-based exclusions are naturally not restricted to a certain business sector but theoretically apply to all companies in the portfolio. In addition, the latter practices can be changed by the company without changing the nature of the operations whereas a company had to sell (part of) its operations to remove the basis for a sector-based exclusion. These differences have the effect that companies excluded due to norm-based screens are exposed to the previously discussed sources of risks to varying degrees. For instance, it can be argued that investors applying sector-based exclusions are more strongly exposed to the limited-risk-sharing-problem due to market segmentation and less exposed to companies' "hidden" risks. As limited risk sharing is associated with an outperformance of excluded companies due to limited diversification opportunities across investors, we expect sector-based exclusions to generate superior performance.

H2a Exclusion portfolios based on sector-based screens outperform the market.

H2b Exclusion portfolios based on norm-based screens underperform the market.

To illustrate, for market participants it is easier to identify what operations a company runs as to assess the way that the business is operated. This makes market segmentation based on sectors more feasible than based on business practices. In addition, while for most norm-based exclusions a comparable substitute from the same industry is available, adequate substitution is often not possible

when excluding an entire business sector. Finally, while the business sector is a more permanent feature of a company, the way that the company runs its business, i.e. in a responsible or irresponsible manner, can be altered more easily. Hence, in conclusion, the risk from market segmentation and limited risk sharing is more likely to materialise for sector-based exclusions while it is more easily diversifiable and thus less likely to be compensated in case of norm-based exclusions.

On the other hand, companies that are excluded due to norm-based screens are more likely to bear "hidden" risks that are not correctly priced by the market than those excluded due to sector-based screens. These "hidden" or mispriced risks associated with unethical behaviour imply that norm-based exclusions are more likely to generate inferior financial performance.

For one thing, breaches of norms and unethical business practices are less visible to the market, especially since the company has a high incentive to obscure the true extent of the incident. This is evidenced by the literature that assesses the impact of announcements of negative human and labour rights and environmental incidents on firm value (e.g. Kappel et al. 2009, for human rights issues; Klassen and McLaughlin 1996; Dasgupta and Laplante 2001; Gupta and Goldar 2005; Konar and Cohen 1997; Flammer 2013, for environmental violations; Hirsh and Cha 2015, for labour rights issues; and Amer 2015, for issues related to non-conformity with the UN Global Compact). These studies predominantly find a loss in firm value around the announcement date indicating that the market has previously mispriced the risk of the company. Depending on the timeliness of the divestment and the speed of market adjustment, a divestment from these companies could protect the GPFG and the AP-funds at least partially against the downward price adjustment caused by the incident or, alternatively, safeguard the funds against the negative return consequences of potential future breaches of norms. In comparison, such misevaluations of the risk involved with operating in a particular sector are less likely, given the often long history of operations of these sectors and the fact that the sector is not an unexpected element of a company. Thus, to conclude market segmentation risks that could result in a temporary undervaluation are more likely to be found for sector-based exclusions while companies excluded due to norm-based screens are more prone to overvaluation related to hidden risks.

Data and Methodology

The following sections introduce the data and methodology used to test the performance implications of the exclusion decisions of the GPFG and the AP-funds.

Data and Portfolio Construction

Our main data source are the exclusion lists published by the GPFG and the AP-funds. The exclusion decisions of these funds are the outcome of a systematic review of companies accused of serious norm violations and other business practices that are in conflict with the ethical standards set out by the funds. These reviews resemble a "quasi-legal" process that assesses the seriousness and extent of the violation as well as the willingness of the company to change its practices. It also allows the companies to respond to the allegations made against them before any exclusion takes place (Richardson 2011). Regarding the scope of the exclusions and in particular the asset classes involved, the AP-funds and the GPFG are generally required to divest from any form of investment in the unethical company, including listed equities, fixed income and other forms of investment such as real-estate.[13] In line with the previous literature, this study particularly focuses on the effect of divestment from listed equities.[14] In the case of the GPFG, a separate body, the Council for Ethics, reviews the allegations made against companies and issues recommendations regarding the exclusion, or otherwise, of a company. Up until the end of 2014 the Ministry of Finance made the final decision on a case-by-case basis, while from 2015 onwards the Norges Bank has been assigned the task of decision making on the observation and exclusion of companies (Norwegian Ministry of Finance 2016). The Swedish AP-funds consist of the five separate funds AP1, AP2, AP3, AP4 and AP6—which represent the income-based pension—and the fund AP7—

[13] While the funds do not clearly state the scope of their exclusion decisions there is considerable evidence that leads us to believe that the entire portfolio is concerned. For the case of GPFG, in the ethical guidelines that stipulate the fund's exclusion criteria it says that "the guidelines cover investments in the Fund's equity and fixed-income portfolios, as well as instruments in the Fund's real-estate portfolio issued by companies listed on a regulated market". In addition, GPFG specifies in its latest RI report for the Principles for Responsible Investment (PRI) that its SRI policies, including its screening/exclusion policy, cover all assets under management (AUM). The AP-funds state in their RI Reports that their SRI policies including screening/exclusion policy comprise their entire AUM (for the case of AP1, AP2, AP7) or the majority of their AUM (for AP3 and AP4). The RI reports for the funds can be obtained online: https://www.unpri.org/signatory-directory/ Furthermore, the ethical council that recommends exclusions for AP1–AP4 states on its homepage that "the recommendation for exclusion also applies to listed subsidiaries" http://etikradet.se/etikradets-arbete/reaktiva-dialoger-vid-krankning/rekommenderade-uteslutningar/?lang=en.

[14] To the best of our knowledge, no empirical study has, so far, analysed the performance implications of exclusionary screening for asset classes other than public equity. While representing an interesting avenue for future research, an extension of the analysis of the performance implications of GPFG's and the AP-funds' exclusions beyond public equity is empirically beyond the scope of this study.

which serves as the government default fund for the premium reserve system.[15] AP1, AP2, AP3 and AP4 follow a similar exclusion process as GPFG in terms of the exclusion process and prior engagement with the company. At the beginning of 2007 the four funds established a joint Ethical Council to coordinate the analysis of the environmental and ethical compliance of their holdings. The purpose of this collaboration is to combine the four funds' resources and votes for greater leverage in influencing companies and to increase the efficiency of the engagements. Although the Ethical Council only issues recommendations and the four funds have the final say regarding the exclusion decisions, AP1, AP2, AP3 and AP4 have all been following the Council's recommendations. Due to their identical exclusions and exclusion policy, we regard AP1, AP2, AP3 and AP4 as one joint fund for the sake of this study, though we acknowledge that they might deviate from each other in terms of investment strategy in other respects.[16] Unlike GFPG and AP1-4, AP7 does not individually disclose each exclusion decision, but it provides a list of its current exclusions in its annual reports (Du Rietz 2016; Bengtsson 2008b). In addition, while it states the reason for exclusion, AP7 does not provide the exact exclusion date. Another difference between AP7's and the other funds' exclusion approach is that it does not rely on prior engagement with the accused company but proceeds straight to exclusion. AP6 does not publish any exclusion list and is thus not considered in this study.

For our study, we collect the entire history of the divestments, including the company name, the reason for exclusion and, if available, the exact date of exclusion, for GPFG, AP7 and the joint exclusions of AP1, AP2, AP3 and AP4. Such detailed information on funds' exclusion decisions is hardly available for other (private) market participants and thus allows us to gain unique insights into the trends in exclusionary screening over time. For instance, studies analysing exclusionary screening by SRI mutual funds do not have information on the excluded companies or on the precise reason for exclusion. For GPFG and the joint exclusions of AP1, AP2, AP3 and AP4, we start from the most recent exclusion list, published on the funds' websites, and reconstruct the lists back in time based on the funds' announcements of past exclusions and re-inclusions. In the few cases, where no precise exclusion date is provided, we use the announcement date of the exclusion instead. For AP7 we rely on the list of excluded companies published in its past annual reports. Our sample starts at the end of 2001 when AP7 publishes its first exclusion list in its annual report. GPFG undertook its first divestment in 2002, while for the case of AP1 to AP4 we document the first exclusion in 2006. We account for all subsequent exclusions and re-inclusions until the end of 2015.

In a next step, we construct portfolios that contain the companies that are being excluded by the funds at any point in time. To do so, we match the exclusion lists published by the funds with the stock price data of the excluded stocks. We obtain monthly stock price data for the excluded companies from Datastream. In line with related studies by Fabozzi et al. (2008), Lobe and Walkshäusl (2011), Salaber (2013) and Trinks and Scholtens (2015), we use Datastream's Total Return Index which reflects a stock's theoretical growth in value assuming all dividends are re-invested.[17] For GPFG and AP1-4, we add a company to the portfolio of excluded companies, based on the stated date of exclusion from the fund's portfolio. We remove a company from the portfolio of excluded companies, once the re-inclusion is announced. Lacking the exact date of AP7's exclusions, we assume that the exclusion list at the end of the year forms the basis for AP7's exclusion portfolio of the following year. We update AP7's portfolio on a year-by-year basis, using the latest annual report. We require a company to appear on AP7's exclusion list in two consecutive years as we must assume that a company which appears on one year's exclusion list but is absent from next year's list could have been re-included by AP7 at any point in time throughout the consecutive year.

For each of the three fund groups, we construct monthly continuously compounded returns for both equal- and value-weighted portfolios. The equal-weighted portfolios assign equal weight to each company so that the return of the portfolio represents the simple average of the individual stock returns. The equal-weighted return is calculated as the natural logarithm of the average return of all companies excluded at the end of a particular month, which can be expressed in the following way:

$$r_{\mathrm{ew},t} = \ln\left[\frac{1}{k}\sum_{i=1}^{k}\frac{P_{i,t}}{P_{i,t-1}}\right], \tag{1}$$

where $r_{\mathrm{ew},t}$ is the equal-weighted, continuously compounded portfolio return over month t, $P_{i,t}$ is the stock price of company i at the end of month t, $P_{i,t-1}$ is that company's

[15] The premium reserve system relates to the part of the public pension money that savers can invest themselves.

[16] The only deviation in the four funds' exclusion portfolios relates to a case prior to the establishment of the joint Ethical Council. In particular, AP2 excluded Wal-Mart at an earlier time than AP1, AP3 and AP4. When constructing the portfolio of excluded companies we use the time of AP2's exclusion as the exclusion date.

[17] Datastream is a database for financial and economic research data from Thomson Reuters. It is widely used in empirical studies of socially responsible investment and in the fields of empirical finance more generally. While the use of this database is not undisputed (see e.g. the analysis of Ince and Porter 2006, regarding potential data errors in Datastream), the data issues pointed out in the prior literature are unlikely to affect our results.

stock price at the end of the previous month $t-1$, and the total number of companies in the portfolio equals k.

In comparison, value-weighted returns account for the weight of a company in the equity market by attaching a higher (lower) weight to companies that represent a larger (smaller) share of the overall equity market. They are computed in a similar fashion to equal-weighted returns but instead of giving each company the same weight in the portfolio, a company's return is weighted by its market capitalisation at the end of the previous month:

$$r_{\mathrm{vw},t} = \ln\left[\sum_{i=1}^{k}\left(\frac{P_{i,t}}{P_{i,t-1}} * \frac{\mathrm{MCap}_{i,t-1}}{\sum_{i=1}^{k}\mathrm{MCap}_{i,t-1}}\right)\right], \quad (2)$$

where $r_{\mathrm{ew},t}$ is the value-weighted, continuously compounded portfolio return over month t and $\mathrm{MCap}_{i,t-1}$ is the market capitalisation of company i at the end of month $t-1$.

Using value-weighted portfolio returns is not only in line with the related literature (e.g. Statman and Glushkov 2009; Lobe and Walkshäusl 2011; Salaber 2013; Adamsson and Hoepner 2015; Trinks and Scholtens 2015), it also better reflects the investment realities at the funds we study. For one thing, these investors are mainly passive investors and thus the weights of the companies in their portfolios closely follow the market weights (e.g. Chambers et al. 2012). In addition, their performance is usually benchmarked against (value-weighted) market indices, as we will discuss in more detail in the following section. Although less practically relevant, equal-weighted portfolio returns have been employed as the sole return measure in the early literature (e.g. Fabozzi et al. 2008; Hong and Kacperczyk 2009) and using them allows us to compare our results to these early findings.

Methodology

To test the performance implications of applying exclusionary screens, we employ two standard asset pricing models. Firstly, we estimate a Capital Asset Pricing Model (CAPM) with the market risk premium corresponding to the excess return of the fund's performance benchmark. Secondly, we test the performance effects in the framework of a Four-Factor model, where we add a size, value and momentum factor to the market factor (Fama and French 1993; Carhart 1997). Using these models is not only standard in the literature and in line with related studies (e.g. Statman and Glushkov 2009; Humphrey and Lee 2011; Humphrey and Tan 2014; Trinks and Scholtens 2015; Adamsson and Hoepner 2015), it also corresponds to the way that these funds are managed. For instance, Chambers et al. (2012) point out that GPFG almost exclusively relies on publicly traded securities, while being constrained to very low deviations from the benchmark portfolio (see also Hoepner et al. 2013, discussing this issue for pension funds in general). Thus, models like the CAPM and the extended factor models which measure performance relative to a benchmark, best capture this management style.

As the funds invest in a global, well-diversified portfolio, the market benchmark used in the models needs to be a global, diversified index. The MSCI All Country World index reflects these features and consequently it is widely used in academic research (e.g. Trinks and Scholtens 2015). In addition, AP7 explicitly employs the index as its benchmark for global equities.[18] The CAPM model can be expressed in the following way:

$$r_{p,t} - r_{f,t-1} = \propto_p + \beta_p\left(r_{m,t} - r_{f,t-1}\right) + u_{p,t}, \quad (3)$$

where $r_{p,t}$ is the continuously compounded return on either the equal-weighted or value-weighted exclusion portfolio p over month t, $r_{f,t-1}$ is the continuously compounded 3-month U.S. Treasury bill rate at the end of month $t-1$ which serves as a proxy for the risk-free rate applicable for month t,[19] $r_{m,t}$ is the continuously compounded return on the MSCI All Country World index which represents the market benchmark portfolio, $\propto_p$ is Jensen's alpha measuring the abnormal return of portfolio p relative to the market, β_p is the market beta of portfolio p capturing the systematic risk exposure of the portfolio and $u_{p,t}$ is the independent disturbance term.

The CAPM model assumes that the only priced risk is a security's exposure to the systematic market risk. However, since its development numerous studies have found that other factors besides the market risk are priced in the cross-section of returns. Among the well-documented factors are the premium for small stocks and value stocks, i.e. stocks with high book-to-market ratios, (e.g. Fama and French 1993) and the outperformance of past winning stocks over past losing stocks, called the momentum effect (Carhart

[18] For further information on AP7's evaluation of equity investments see the fund's homepage: http://www.ap7.se/en/About-AP7/About-us/Our-approach/Equity-investments/ . AP1–4 do not state any official benchmark index. Their performance seems to be mainly measured in relation to each other and based on the level of their annual operating costs (Severinson and Stewart 2012). GPFG uses an individually customised equity index which can be downloaded from GPFG's homepage. However, the MSCI All Country World index and the returns of GPFG's strategic index have a correlation of 99.49 % and replacing the MSCI index with GPFG's strategic index does not significantly change our results. The results of the latter analysis are available from the authors upon request.

[19] We transform the 3-month U.S. Treasury bill rate into a continuously compounded risk-free rate using the following formula: $R_{f,t,1m} = \ln\left(1 + SR_{f,t,3m} * \frac{3}{12}\right)^{\frac{1}{3}}$, where $R_{f,t,1m}$ is the continuously compounded 1-month rate and $SR_{f,t,3m}$ is the stated 3-month Treasury rate.

1997). Previous literature has found that companies that act in a socially responsible manner show a different exposure to these size, value and momentum factors than socially irresponsible firms (e.g. Bauer et al. 2005; Galema et al. 2008; Statman and Glushkov 2009). Thus, to make sure that any performance difference between the excluded companies and the benchmark is not purely driven by different loadings on these risk factors, we add these three factors to our market model, which can now be expressed in the following way:

$$r_{i,t} - r_{f,t-1} = \propto_i + \beta_i\left(r_{m,t} - r_{f,t-1}\right) + \gamma_i \mathrm{SMB}_t + \delta_i \mathrm{HML}_t + \varphi_i \mathrm{WML}_t + u_{i,t}, \quad (4)$$

where SMB_t (small minus big) is the global size factor calculated as the difference in return of the stocks in the lower half of a market capitalisation ranked global stock universe and the stocks in the upper half of the same universe, HML_t (high minus low) is the global value factor calculated as the return difference of the top 30 % of global stocks ranked by book-to-market ratio and the bottom 30 % of these stocks ranked by book-to-market ratio, and the WML_t (winner minus loser) is the global momentum factor calculated as the return difference between the top 30 % and the bottom 30 % of stocks ranked by previous 12 months returns.[20]

Results

This section presents and discusses the results on the performance impact of exclusionary screening. We first describe the composition of the exclusion lists of the funds in our sample, before we present the estimation results of the factor models.[21]

Descriptive Statistics

Panel A of Table 1 provides an overview of the cross-sectional characteristics of the exclusion lists of the GPFG, AP1-4 and AP7, based on the composition of their exclusion lists at the end of a year. Several interesting differences across the funds can be observed. Firstly, the three fund groups seem to differ with respect to the extent to which they apply exclusionary screens. While AP1 to AP4 only exclude a total of 20 companies over the sample period with an average of just over 14 exclusions per year, the GPFG's exclusion lists comprise an average of about 49 companies per year representing 74 different firms. Although AP7's annual exclusion list, on average, only consists of less than 43 companies, the fund has excluded a total of 152 different companies over the entire sample period. Comparing the number of exclusions to the total number of companies that these funds invest in, the extent of exclusionary screening appears small. To illustrate, the GPFG currently holds around 9000 companies while AP7's equity investment universe spans around 2500 different companies.[22] Thus, the excluded companies only make out around 0.7 % (i.e. 63/9000) of the total number of holdings for GPFG and 1.8 % (i.e. 46/2500) for AP7, respectively. Similarly, the share of excluded companies to total number of holdings is about 0.8 % (i.e. 20/2500) for AP2, 0.7 % (i.e. 20/3000) for AP3, and 1.2 % (i.e. 20/1700) for AP4.[23] The only exception is AP1. From 2014 onwards, AP1 has been shifting its equity strategy from holding a broad universe of global and domestic stocks to a strategy of concentrated ownership and has reduced its equity holdings from about 3000 to 600 companies.[24] However, the 20 companies excluded in 2015 still only represent a small fraction of the total number of holdings of just over 6 %. These figures are also in line with the number of excluded companies typically found in the mutual fund industry. For instance, Humphrey and Tan (2014) simulate exclusion portfolios of a typical mutual fund and their portfolios comprise an average of 60 exclusions. In addition, Blanc and Cozic (2012) reviewing the norm-based exclusions of 32 European asset owners and asset managers find that these investors exclude on average 26 companies based on violations of international norms and association to controversial weapons.

Secondly, the funds in our sample do not only seem to differ in their tendency to exclude companies but also in their likelihood to re-include companies. Re-inclusions are cases where the fund revokes its exclusion decision and the company re-enters the fund's investment portfolio. These re-inclusions are usually a result of the periodic reviews undertaken by the funds to check whether the reason for exclusion still exists. For GPFG, we document a total of eight re-inclusions, while we find no re-inclusion

[20] The data for the global size, value and momentum factors are obtained from Kenneth French's online data library: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[21] All estimations and empirical tests that are presented in this study have been performed using the statistical software package Stata.

[22] Details about GPFG's assets and holdings can be obtained from its homepage: http://www.nbim.no/en/the-fund/. AP7 states that its equity investments follow the MSCI All Country World Index which currently covers around 2500 companies: https://www.ap7.se/en/About-AP7/About-us/Our-approach/Equity-investments/.

[23] The most recent equity portfolios of AP2, AP3 and AP4 can be obtained on the following websites: AP2 http://www.ap2.se/en/Portfolio/portfolio/; AP3 http://www.ap3.se/sites/english/portfolio/totalportfolio/Pages/Securitiesholdingsnew.aspx; AP4 http://www.ap4.se/en/financial-reports-and-press/reports/.

[24] See AP1's 2014 Ownership report on page 10, which is available online: http://www.ap1.se/en/Financial-information-and-press/Topical/.

Table 1 Summary statistics of the exclusion lists

Total exclusions		Total re-inclusions		Number of excluded companies (per year)							Average time of exclusion (in years)					
				Mean			Min				Mean			Min		
Panel A: cross-sectional characteristics																
GPFG	74	8		49.03			1				63			7.93		
AP1–4	20	–		13.59			1				20			7.05		
AP7	152	99		42.53			19				54			5.71		
		2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Panel B: exclusion lists at year end																
Exclusions by all funds		26	40	35	20	40	45	56	97	109	115	115	118	121	125	129
Thereof excluded by																
GPFG		–	1	1	1	13	19	26	30	48	49	53	54	57	59	63
AP1–4		–	–	–	–	–	2	3	13	13	14	14	14	17	19	20
AP7		26	39	34	19	27	24	27	54	48	52	48	50	47	47	46
Thereof excluded due to																
Environmental issues		8	10	7	9	8	4	12	20	17	17	15	15	19	20	24
Human rights issues		14	17	17	7	15	15	17	21	18	26	25	29	25	30	30
Labour rights issues		9	18	15	6	8	8	7	8	8	7	8	7	7	6	7
Controversial weapons		1	2	3	2	13	20	23	52	52	53	53	53	52	51	50
Tobacco		–	–	–	–	–	–	–	–	17	17	19	19	21	21	21
Thereof located in the following regions																
North America		14	13	8	7	22	25	24	43	47	49	46	47	53	56	56
Europe		5	9	7	4	9	12	10	20	23	24	26	26	23	22	22
Asia		5	15	16	7	8	4	16	26	31	35	34	34	33	35	39
Australia		1	2	2	1	1	2	2	6	6	4	4	4	5	5	5
Africa		–	–	–	–	–	–	2	1	–	–	–	–	–	–	–
South America		1	1	2	1	–	2	2	2	3	4	6	8	8	8	8
Thereof operating in the following industries																
Aerospace and defence		–	2	3	2	12	20	21	40	40	40	39	39	37	37	37
Chemicals related		2	3	2	2	3	1	–	2	2	4	4	4	7	9	9
Construction and materials		–	–	–	–	–	–	–	1	1	3	5	7	9	10	11
General retailers		3	3	2	1	1	5	5	6	6	6	6	6	6	6	6
Industrial metals and mining		2	4	2	1	1	4	8	13	13	11	11	13	14	15	16
Oil and gas related		8	7	5	5	10	8	6	7	3	5	4	5	4	4	4
Tobacco		–	–	–	–	–	–	–	–	17	17	17	17	18	18	18
Other industries		11	21	21	9	13	7	16	28	27	29	29	27	26	26	28

Panel A provides summary statistics on the cross-sectional characteristics of the exclusion lists of the three fund groups—GPFG, AP1–AP4, and AP7. The first two columns provide the number of different companies that are excluded from and re-included to the funds over the sample period 2001–2015. Note that we do not count companies double. The third column provides summary statistics on the number of excluded companies per fund. The fourth column states the average time (in years) that a company has been on a fund's exclusion list. Panel B reports the number of companies that are excluded from the funds at year end. The first row lists the number of exclusions by all funds. The following sets of rows report the number of exclusions (a) by each fund, (b) by the reason for exclusion (as stated by the fund), (c) by the location of the excluded company (company's headquarter), and (d) by the company's main industry. Note that companies can be excluded due to more than one reason. All figures are based on the comparison of end-of-year exclusion lists

announcements for AP1-4. AP7 seems to frequently re-include companies with a total of 99 cases of re-inclusions between 2001 and 2015.[25] In general, AP7's exclusion lists show far higher variation across years, while GPFG's and

[25] We interpret every case when a company appears on last year's exclusion list but is absent on the exclusion list in the consecutive year as a re-inclusion of that company, unless the company disappeared from the investment universe e.g. due to a merger, bankruptcy or privatisation.

AP1-4's exclusion lists appear more constant over time. This pattern is confirmed when comparing the average duration of a company on the funds' exclusion lists which is 8 years for GPFG and 7 years for AP1-4, respectively, while it is less than 6 years for the case of AP7. However, when comparing these figures across countries one has to consider that AP1-4 have started their exclusionary screening considerably later than the other two funds which biases the average duration of companies on AP1-4's exclusion list downwards.

Panel B of Table 1 provides a comparison of the exclusion lists over time and thus enables us to identify some interesting patterns in the exclusionary approaches adopted by the funds. For one thing, we document a gradually increasing trend in the number of exclusions, both when aggregating across funds and for each fund individually. In addition, we find further support that AP7's approach towards exclusionary screening differs from that of the other funds. AP7's exclusion list already comprises a comparably high number of 26 companies right from the beginning of its exclusionary screening in year 2001. In comparison, GPFG and AP1-4 start off with singular exclusions of one and two companies, respectively. AP7 also almost gradually increases the number of excluded companies over time, whereas the exclusion lists of GPFG and AP1-4 experience wave-like rises in the number of excluded companies. Although the reasons for these differences are unknown, they might relate to AP7 having a less formalised exclusion process than the other two funds, such as no separate ethical council, no public justification of the reasons for exclusion and no prior engagement with the companies, allowing it greater flexibility in the exclusion decisions.

Panel B of Table 1 also offers a break-down of the exclusions by reason for exclusion.[26] Overall, the companies in our sample are either excluded due to environmental, human rights or labour rights issues or because they are associated to the production of controversial weapons or tobacco. Interestingly, the funds do not exclude companies due to other reasons frequently studied in the academic literature such as alcohol, gambling and adult entertainment (see e.g. the early studies by Adler and Kritzman 2008; Fabozzi et al. 2008; Hong and Kacperczyk 2009) as well as fossil fuel companies which have recently become a popular target of divestment campaigns. Thus, our study contributes to the literature by shedding light on less well researched areas of exclusionary screening. Looking at the trends over time, we find that human rights issues and labour rights issues have been the most frequently applied screens in the early part of the sample, while controversial weapons and tobacco gained importance in the later years. In fact, screening for controversial weapons has been the most frequently applied screen since 2008. Tobacco companies entered the exclusion list in 2009 when GPFG added tobacco to its exclusion criteria. In contrast, the Swedish funds do not exclude tobacco stocks, arguing that the manufacture, sale and use of tobacco is not illegal in Sweden so that tobacco divestment does not have a legal basis.[27]

Finally, Panel B of Table 1 allows insights into the geographical and sectoral distribution of exclusions. Most of the excluded companies appear to be located in North America, followed by Asia and Europe. Only few excluded companies are located in South America, Australia and Africa. However, while this finding does not imply that the corporate misconduct must have been committed e.g. in North America—it can relate to unethical behaviour in other parts of the world committed by companies headquartered in North America—it suggests that unethical business practices and violations of international norms are not restricted to the corporate sector of emerging and developing markets. In fact, they are most frequently committed by companies from regions which rank highly on rankings of the quality of governance and the legal system.[28]

Regarding the industries that the excluded companies operate in, the majority of exclusions comprise aerospace and defence companies, reflecting the popularity of the controversial weapons screen. The same holds for tobacco companies which constitute a considerable share of the exclusion portfolio due to GPFG's tobacco divestment. In addition, companies operating in the sectors of construction & materials as well as industrial metals & mining appear frequently on the exclusion lists. This finding is in line with the results obtained by Blanc and Cozic (2012) based on a comparison of 32 European investors and, according to the authors, relates to the higher exposure of these sectors to environmental, social and governance risks. Besides, we do not find a strong dominance of other sectors. Interestingly, and in line with Blanc and Cozic (2012), with the exception of Wal-Mart, the lists do not feature companies from the mass retailing industry, such as popular warehouse chains, e-retailers and the food processing industry which have

[26] Some companies are excluded due to more than one reason, for example based on violations of environmental norms and due to human rights issues. Thus, the individual columns of Table 1 may not always sum up to the overall number of excluded companies.

[27] A position statement of why the funds AP1, AP2, AP3 and AP4 do not divest from tobacco companies is provided on the homepage of their ethical council, available online: http://etikradet.se/etikradets-arbete/positioner/tobak/?lang=en.

[28] See for example the latest Rule of Law ranking by the World Justice Project, available online: http://worldjusticeproject.org/rule-law-around-world.

been involved in several corporate scandals over the last years.

Table 2 provides descriptive statistics on the returns of the exclusion portfolios across funds and thus allows a preliminary assessment of the performance implications of the exclusions. Panel A focuses on the entire set of excluded companies, while Panel B compares returns on the excluded companies sorted by the different types of exclusionary screens. Overall, the average returns on the exclusion portfolios are relatively low and mostly positive. The highest average monthly return amounts to 1.3 % and is documented for GPFG's value-weighted tobacco exclusions. Only three portfolios yield negative average returns, namely AP1-4's equal- and value-weighted environmental exclusions with monthly returns of −4 % and −4.1 %, respectively, and GPFG's value-weighted environmental exclusions with a return of −0.6 %. In addition, we find that in the majority of cases the equal-weighted portfolios have slightly higher returns than their value-weighted equivalents. This finding is in line with the widely documented "size" effect in stock returns and reflects the empirical observation that smaller stocks tend to exhibit higher than average returns (e.g. Fama and French 1993). As the equal-weighted portfolios give greater weight to the smaller stocks than value-weighted portfolios, the higher returns are likely to reflect the different loadings on the size factor. This finding highlights the importance of using value-weighted portfolio returns as well as the need to explicitly control for the size effect in the later estimations.

When comparing the portfolio returns across the type of exclusionary screens (Table 2, Panel B), we find greater differences in average portfolio returns than on the aggregate level. This finding provides initial evidence that the performance implications of ethical screening might differ across screens and in this sense is in line with the existing literature (e.g. Renneboog et al. 2008b; Capelle-Blancard and Monjon 2014; Trinks and Scholtens 2015).

Main Portfolio Performance Results

While the descriptive statistics allow a first assessment of the performance of the different exclusion portfolios, they do not account for different exposures to risk. This section presents the results of measuring the risk-adjusted performance of the exclusion portfolios using the CAPM and the Four-Factor models. We are particularly interested in the alpha estimates from these regressions as a positive (negative) and significant alpha estimate indicates that the exclusion portfolio outperforms (underperforms) relative to the market. Thus, excluding these companies from the funds' investment universe financially hurts (benefits) the fund. In comparison, if we find no significant performance difference we conclude that these funds can meet their ethical standards without sacrificing returns.

Panel A of Table 3 presents the estimation results based on the CAPM model which accounts for the systematic market risk of a portfolio. Overall, we only find very weak evidence of any significant performance effect of applying exclusionary screens. Out of the six exclusion portfolios, three portfolios exhibit a positive and significant alpha, of which two are only significant at the 10 % level. These include AP7's equal-weighted and value-weighted exclusion portfolios and GPFG's equal-weighted exclusions. On an annualised basis, the abnormal returns on AP7's exclusions amount to 5.4 % for the equal-weighted portfolio and 3.4 % for the value-weighted portfolio. GPFG's equal-weighted exclusion portfolio generates an annual return of 4.4 %. However, due to the low statistical significance, especially of the practically more relevant value-weighted portfolios, it is highly doubtful whether investing in the excluded companies would have yielded a measurable abnormal return. Finally, the exclusion portfolios of AP1-4 neither out- nor underperform in the CAPM-framework, independent of the weighting scheme.

The results of the Four-Factor model are presented in Panel B of Table 3. Having added the additional global risk factors, we find that only two portfolios significantly outperform the benchmark model. AP7's equal-weighted portfolio generates a positive and significant abnormal return of 4.3 % per annum, while AP1-4's equal-weighted exclusion portfolio outperforms the benchmark by 6.2 % per annum. However, in both cases the results are only weakly statistically significant and the significance is lost when applying value-weighting to the returns. This finding is in line with results presented in Statman and Glushkov (2009) and Adamsson and Hoepner (2015) who find that the outperformance of shunned stocks is only statistically significant for equal-weighted portfolios, while the effect becomes statistically insignificant and economically smaller for value-weighted portfolios.

To conclude, the majority of the results suggests that the funds are neither significantly hurt nor do they financially benefit from excluding the stocks from their portfolios which supports hypothesis *H1c* of an insignificant performance effect of applying exclusionary screening. While we find no support for *H1b* and thus a performance enhancing effect of exclusionary screening, we find very limited evidence that the excluded companies outperform the benchmark which is in line with *H1a*. Overall, our results confirm findings of the literature on SRI mutual funds, though using a different methodological approach by focusing on the returns of the excluded companies instead of the returns of the screening fund (e.g. Lobe and Walkshäusl 2011; Humphrey and Lee 2011; Humphrey and Tan 2014). In comparison, our findings are in contrast to the

Table 2 Descriptive statistics on portfolio returns

	AP7		AP1–4		GPFG	
	Equal weighted	Value weighted	Equal weighted	Value weighted	Equal weighted	Value weighted
Panel A: all excluded companies						
Mean	0.009	0.006	0.007	0.005	0.008	0.006
SD	0.055	0.040	0.062	0.040	0.056	0.045
Min	−0.219	−0.116	−0.285	−0.131	−0.274	−0.173
Max	0.195	0.101	0.225	0.099	0.182	0.108
Count	166	166	109	109	162	162
Panel B: mean returns by exclusionary screen						
Human rights issues	0.004	0.002	0.002	0.000	0.011	0.006
Labour rights issues	0.009	0.003	0.005	0.004	–	–
Environmental issues	0.007	0.005	−0.040	−0.041	0.000	−0.006
Controversial weapons	0.004	0.002	0.006	0.004	0.008	0.008
Tobacco	–	–	–	–	0.012	0.013

This table reports descriptive statistics for the equal-weighted and value-weighted continuously compounded returns of the portfolios of excluded companies for the AP7, the AP1–4 and the GPFG funds. Panel A focuses on the portfolios consisting of all companies excluded by a fund during a particular time. Panel B reports average continuously compounded portfolio returns sorted by the reason for exclusion. For the value-weighting, companies are weighted by their market capitalisation of the previous month in order to prevent any look-ahead bias

studies by Adler and Kritzman (2008), Fabozzi et al. (2008), Hong and Kacperczyk (2009), Durand et al. (2013b), Salaber (2013), and Trinks and Scholtens (2015) which, based on an analysis of theoretical portfolios of unethical companies, conclude that these companies generate superior financial performance and that, as a consequence, exclusionary screening has a negative performance impact.

However, our analysis considerably differs from the above studies in several ways. Firstly, while the exclusionary screens studied in the previous literature mainly comprise the traditional sin screens with several additions of other sector-based screens, the exclusions by the GPFG and the AP-funds mainly reflect norm-based screening (with the exception of tobacco for GPFG). As holding companies that violate international norms may expose investors to different risks than holding companies that operate in "sin" sectors, we should not expect that the results of the previous literature can simply be extended to all forms of exclusionary screens. We will explore this aspect in more detail in the following section. Secondly, our analysis differs from the above studies because we rely on actual exclusions of real-world investors and analyse the performance effect at the company-level. This way we prevent our results from being driven by confounding factors such as manager skill. And finally, we put greater emphasis on value-weighted returns as these are practically more relevant – an aspect that is neglected by several studies including Fabozzi et al. (2008) and Hong and Kacperczyk (2009).

Turning to the coefficient estimates on the four risk factors, it appears that the size of the estimates on the MSCI market returns are only slightly affected by the inclusion of the additional risk factors and all maintain their high statistical significance. The estimates on the size factor are intuitive. They are positive and significant for all equal-weighted portfolios due to the overexposure to small-capitalisation stocks induced by the weighting scheme and they turn negative when value-weighting the returns. The latter indicates that the excluded companies tend to be larger than the average company in the MSCI universe, after accounting for the companies' market capitalisation. This is in line with anecdotal evidence that the GPFG and the AP-funds rather focus on large and more publicly visible companies when it comes to their divestment decisions (e.g. Clark and Monk 2010). Apart from AP7, none of the funds' exclusion portfolios has a significant exposure to value or growth stocks as shown by the insignificant coefficient estimates on the HML factor. The momentum factors show weak significance in explaining the portfolios' return variation, with only two cases of statistically significant factor exposure (i.e. AP7's value-weighted and AP1-4's equal-weighted exclusion portfolio). Thus, contrary to previous literature (e.g. Bauer et al. 2005; Galema et al. 2008; Statman and Glushkov 2009), we do not find strong evidence that unethical companies load significantly differently on the standard risk factors, with the exception of the size factor.

Performance Results by Screen

Previous research suggests that the performance impact of exclusions is conditional on the reason for exclusion (e.g.

Table 3 Main performance results

	AP7		AP1–4		GPFG	
	Equal weighted (1)	Value weighted (2)	Equal weighted (3)	Value weighted (4)	Equal weighted (5)	Value weighted (6)
Panel A: CAPM model						
Alpha	0.00450*** (2.632)	0.00286* (1.833)	0.00373 (1.192)	0.00346 (1.127)	0.00425* (1.654)	0.00364 (1.618)
MSCI	1.067*** (27.642)	0.755*** (17.640)	1.044*** (11.652)	0.495*** (7.592)	0.962*** (12.445)	0.737*** (14.479)
Observations	166	166	109	109	162	162
R^2	0.835	0.766	0.731	0.393	0.667	0.595
Adj. R^2	0.834	0.765	0.728	0.387	0.665	0.593
	AP7		AP1–4		GPFG	
	Equal weighted (7)	Value weighted (8)	Equal weighted (9)	Value weighted (10)	Equal weighted (11)	Value weighted (12)
Panel B: four-factor model						
Alpha	0.00361** (2.040)	0.00204 (1.299)	0.00519* (1.707)	0.00345 (1.194)	0.00373 (1.277)	0.00292 (1.168)
MSCI	1.040*** (26.396)	0.791*** (18.656)	0.950*** (11.700)	0.489*** (6.367)	0.956*** (14.269)	0.780*** (15.631)
SMB	0.393*** (3.119)	−0.216** (−2.160)	0.436** (2.396)	−0.766*** (−3.755)	0.419*** (2.686)	−0.0906 (−0.652)
HML	0.232* (1.917)	0.172* (1.882)	0.341 (1.386)	0.270 (1.350)	−0.0106 (−0.056)	0.0428 (0.301)
WML	−0.0416 (−0.568)	0.144** (2.130)	−0.208* (−1.793)	0.0458 (0.677)	−0.0239 (−0.180)	0.142 (1.439)
Observations	166	166	109	109	162	162
R^2	0.855	0.788	0.771	0.491	0.682	0.608
Adj. R^2	0.851	0.783	0.763	0.471	0.674	0.598

This table presents the results of the performance analysis of the portfolios comprising all exclusions (independent of the reason for exclusion). Performance is measured according to two market models. Panel A presents the estimates from a CAPM where the MSCI All Country World index serves as the market factor. Panel B reports results from a global four-factor model where we add the global size factor (SMB), value factor (HML) and momentum factor (WML) to the market factor (MSCI). The dependent variables are the continuously compounded excess returns on the equal- or value-weighted exclusion portfolios of one of the three funds. Robust *t* ratios are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively

Barnett and Salomon 2006; Capelle-Blancard and Monjon 2014; Trinks and Scholtens 2015) and the results in Table 2 draw a similar picture. Thus, in this section, we re-run the performance analyses based on portfolios sorted by different exclusionary screens. The results are presented in Table 4. To save space, we report only the alpha estimates, adjusted R^2 values and the number of observations for each specification.

Overall, we do not find a systematic pattern of abnormal returns based on a specific type of exclusionary screen. From the 24 CAPM specifications presented in Panel A, only six exclusion portfolios generate significant abnormal returns, of which four positively outperform the benchmark and two significantly underperform. For the statistically more accurate Four-Factor model presented in Panel B (see Adamsson and Hoepner 2015), two portfolios generate a positive abnormal return. However, these cases of abnormal performance seem to be rather related to the particular fund or weighting scheme and are only of weak statistical significance. This finding indicates that the performance effect is not systematically linked to the unethical behaviour of the portfolio companies but rather a result of the portfolio construction process. The only possible exception is the outperformance of tobacco stocks, which remains significant in three out of four cases. However, we are cautious in drawing too strong conclusions from this finding. While it might appear as a confirmation of the previous literature (e.g. Hong and Kacperczyk 2009; Trinks and Scholtens 2015), which finds tobacco stocks to outperform the market, tobacco stocks do not outperform in the most relevant of these four specifications—value-weighted portfolio returns in a Four-Factor model. Hence, Adamsson and Hoepner's (2015) thesis that the previous literature only found a small stocks effect among sin stocks instead of a true tobacco-related effect remains valid, since equal-weighted portfolios overemphasise small stocks and resemble real-world investors much less than value-weighted portfolios. In any case, since the tobacco screen is the only purely sector-based screen analysed in this study the pattern of results observed in Table 4 is in line with our hypothesis *H2a*. In comparison, the finding of an insignificant performance effect for the norm-based screening does not support our hypothesis *H2b* which predicts companies excluded due to violations of

Table 4 Performance results by type of screen

	Panel A: CAPM model					Panel B: four-factor model				
	Equal weighted		Value weighted			Equal weighted		Value weighted		
	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.
	AP7									
Human rights issues	0.00002 (0.009)	0.735	−0.00092 (−0.405)	0.583	166	−0.00069 (−0.291)	0.756	−0.00215 (−0.939)	0.618	166
Labour rights issues	**0.00598**** (**2.265**)	0.596	0.00093 (0.384)	0.392	166	**0.00562**** (**2.044**)	0.597	0.00160 (0.632)	0.424	166
Environmental issues	0.00286 (0.932)	0.590	0.00174 (0.626)	0.548	166	−0.00069 (−0.290)	0.756	−0.00215 (−0.939)	0.618	166
Controversial weapons	−0.00177 (−0.288)	0.447	−0.00355 (−0.578)	0.466	166	0.00202 (0.385)	0.490	−0.00039 (−0.072)	0.491	166
Tobacco	–	–	–	–	–	–	–	–	–	–
	AP1–4									
Human rights issues	−0.00145 (−0.287)	0.487	−0.00297 (−0.572)	0.473	85	−0.00119 (−0.248)	0.498	−0.00236 (−0.477)	0.483	85
Labour rights issues	0.00339 (0.788)	0.340	0.00341 (0.853)	0.156	109	0.00360 (0.849)	0.368	0.00288 (0.783)	0.292	109
Environmental issues	**−0.0500*** (**−1.976**)	0.249	**−0.0506*** (**−1.992**)	0.250	25	−0.00119 (−0.248)	0.498	−0.00236 (−0.477)	0.483	25
Controversial weapons	0.00485 (1.336)	0.770	0.00287 (0.761)	0.645	94	0.00525 (1.535)	0.791	0.00318 (0.839)	0.636	94
Tobacco	–	–	–	–	–	–	–	–	–	–
	GPFG									
Human rights issues	0.00721 (1.544)	0.453	0.00444 (1.070)	0.165	124	0.00787 (1.633)	0.456	0.00342 (0.836)	0.216	124
Labour rights issues	–	–	–	–	–	–	–	–	–	–
Environmental issues	−0.00373 (−0.533)	0.579	−0.00972 (−1.298)	0.537	113	0.00787 (1.633)	0.456	0.00342 (0.836)	0.216	113
Controversial weapons	0.00436 (1.603)	0.600	**0.00517*** (**1.846**)	0.537	162	0.00327 (1.038)	0.613	0.00427 (1.351)	0.538	162
Tobacco	**0.00718**** (**2.309**)	0.591	**0.00833**** (**2.169**)	0.461	70	**0.00746*** (**1.948**)	0.608	0.00679 (1.598)	0.553	70

This table presents the results of the performance analysis of the exclusion portfolios sorted by the reason for exclusion. Panel A reports the estimates from a CAPM with the MSCI All Country World index as the market factor. Panel B shows results from a global four-factor model where we add the global size factor (SMB), value factor (HML) and momentum factor (WML) to the market factor (MSCI). The dependent variables are the continuously compounded excess returns on the equal- or value-weighted exclusion portfolios of one of the three funds. The table only reports the alpha estimates representing the abnormal return of a portfolio with respect to the benchmark model, the adjusted R^2 values of the model as well as the number of observations per model. The coefficient estimates on the risk factors are omitted to preserve space, but are available upon request. Robust t-ratios are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively. Alpha estimates that are significant at the 10, 5 or 1 % level are highlighted in bold

international norms to underperform relative to the market. However, we acknowledge that a thorough analysis of the performance differences between norm-based and sector-based screening would require a more comprehensive set of sector-based screens. Our findings confirm those of Capelle-Blancard and Monjon (2014) who compare the performance of 116 French SRI mutual funds that perform either sectoral or norm-based screens. While arriving at the same conclusion, our analysis differs from that of Capelle-Blancard and Monjon (2014) in several ways. Firstly, we focus on a completely different investor class that is subject to different tensions between the ethical and financial demands of their beneficiaries. Secondly, Capelle-Blancard and Monjon (2014) can only observe performance at the fund level. Given the high heterogeneity across SRI mutual funds, they cannot clearly disentangle the performance impact of the exclusionary screening from that of other fund-related factors such as managerial skill (e.g. Humphrey and Tan, 2014). And finally, since they do not know what companies are excluded by the different funds they cannot validate whether the funds truly perform the exclusionary screens that they state.

Robustness Tests

Long-Short Portfolios

In this section, we test the robustness of our results. First, we re-visit the question of the effect of different screens on performance. In particular, we look at the differential impact of screens for norm-based and sector-based exclusions. We argue that while most of the screens do not significantly impact returns when analysed individually, they might show a significant performance difference when comparing them in relation to one another. To filter out these relative performance effects, we construct long-short portfolios within the categories of norm-based screening and sector-based screening. Long-short portfolios invest a certain amount of money in one set of companies (long portfolio), while at the same time short selling a different set of companies (short portfolio) matching the investment in the long portfolio. A special feature of long-short portfolios is that ideally they do not have exposure to the overall market risk as potential value increases (decreases) experienced by the companies in the long portfolio are automatically cancelled out by respective decreases (increases) in value in the short portfolio. Instead, long-short portfolios accentuate differences in performance that relate to the sorting criteria. Due to these special features, long-short portfolios have been frequently employed in the literature on exclusionary screening (e.g. Kempf and Osthoff 2007; Statman and Glushkov 2009; Hong and Kacperczyk 2009) and SRI more generally (e.g. Derwall et al. 2005). To illustrate the underlying logic of long-short portfolios, let us consider a portfolio that invests in the human rights exclusions and that is short in the labour rights exclusions. If it was financially harmful to exclude companies based on human rights issues relative to labour rights issues, we should find a positive abnormal return on this long-short portfolio. We construct long-short portfolios for all screen combinations within the norm-based screening category and the sector-based screening category in the same way. While the only pure sector-based screen in our sample is the tobacco screen, we also classify the controversial weapons screen as sector-based for the sake of this analysis. However, strictly speaking it should be considered a norm-based screen as funds do not systematically exclude the military and arms industry but only companies that are associated with the production and sale of weaponry that violates international conventions, such as cluster bombs and anti-personnel mines.

The results are presented in Table 5. Overall, we do not find a consistent differential performance effect within the two screening categories. All abnormal returns on specific long-short portfolios lose their statistical significance when changing the market model or the weighting scheme. This further supports our main finding that exclusionary screening does not significantly impact fund performance.

Industry-Specific Risk Factors

In our market models we employ risk factors that are constructed on a global economy level. Thus, we implicitly average the effects of these risk-factors over industries and regions. Adamsson and Hoepner (2015), however, show that risk characteristics, such as size, value and momentum, vary across sectors and that conditioning on industry-specific risk factors affects the performance implications of exclusionary screening (see also Li et al. 2006; Hanhardt and Ansotegui 2008). While this is unlikely to affect our findings regarding the norm-based screens—exclusions due to violations of norms are not industry-dependent—we cannot rule out that our sector-based screening results are driven by industry-specific risk factors. In fact, the tobacco analysis suggests that the way we control for the size of the companies affects our conclusion regarding the performance implications of this screen. To address this issue, we introduce industry-specific risk factors to the Four-Factor model and re-run the analysis for the controversial weapons and tobacco screens. For the industry market factor, we use the corresponding MSCI All Country World industry indices (i.e. aerospace & defence for controversial weapons and tobacco for the tobacco screen). To construct the industry-based size, value and momentum factors, we use the Style Research database and construct the factors in

Table 5 Long-short portfolios

	Panel A: CAPM model					Panel B: four-factor model				
	Equal weighted		Value weighted			Equal weighted		Value weighted		
	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.
	AP7									
Norm-based screens										
HR minus LR	**−0.00596*** (**−1.703**)	0.0454	−0.00185 (−0.625)	0.0477	166	**−0.00631*** (**−1.743**)	0.0926	−0.00375 (−1.214)	0.091	166
Env. minus HR	0.00284 (0.928)	0.00957	0.00266 (1.162)	0.0283	166	0.00224 (0.713)	0.0214	0.00235 (1.000)	0.0215	166
Env. minus LR	−0.00312 (−0.759)	0.00464	0.000811 (0.210)	0.0723	166	−0.00408 (−0.923)	0.00294	−0.00140 (−0.341)	0.0914	166
Sector-based screens										
Weap. minus Tob.	–	–	–	–	–	–	–	–	–	–
	AP1–4									
Norm-based screens										
HR minus LR	−0.00372 (−0.486)	0.00173	−0.00465 (−0.622)	0.0992	85	−0.00400 (−0.580)	0.0762	−0.00408 (−0.614)	0.200	85
Env. minus HR	**−0.0570**** (**−2.209**)	0.123	**−0.0498*** (**−1.915**)	0.127	25	−0.0192 (−1.021)	0.614	−0.0140 (−0.670)	0.549	25
Env. minus LR	**−0.0479*** (**−1.864**)	0.113	−0.0462 (−1.656)	0.115	25	−0.0136 (−0.657)	0.574	−0.00745 (−0.356)	0.643	25
Sector-based screens										
Weap. minus Tob.	–	–	–	–	–	–	–	–	–	–
	GPFG									
Norm-based screens										
HR minus LR	–	–	–	–	–	–	–	–	–	–
Env. minus HR	−0.0106 (−1.237)	0.159	−0.0127 (−1.386)	0.319	113	−0.0109 (−1.457)	0.304	−0.0119 (−1.448)	0.416	113
Env. minus LR	–	–	–	–	–	–	–	–	–	–
Sector-based screens										
Weap. minus Tob.	−0.00222 (−0.470)	0.0599	−0.000187 (−0.035)	0.0141	70	−0.00260 (−0.448)	0.0833	−0.000422 (−0.067)	0.0400	70

This table presents the results of the performance analysis where the dependent variable is the continuously compounded excess returns on a long-short portfolio that is long in one of the exclusion portfolios and short in another exclusion portfolio. For instance, the first row provides performance results for a portfolio that is long the companies excluded due to human rights issues (HR) and short companies excluded due to labour rights issues (LR). The returns of the long-short portfolios are calculated as the equal-weighted and value-weighted returns. HR stands for Human Rights, LR for Labour Rights, Env. for Environment, Weap. for Controversial Weapons, and Tob. for Tobacco, respectively. Panel A presents the estimates from a CAPM, while Panel B reports results from a global four-factor model. The table only reports the alpha estimates, the adjusted R^2 values and the number of observations per model. The coefficient estimates on the risk factors are omitted to preserve space, but are available upon request. Robust t ratios are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively. Alpha estimates that are significant at the 10, 5 or 1 % level are highlighted in bold

Table 6 Factor model with industry-based factor adjustment for sector-based exclusion portfolios

	Controversial weapons						Tobacco	
	AP7		AP1–4		GPFG		GPFG	
	Equal-weight (1)	Value-weight (2)	Equal-weight (3)	Value-weight (4)	Equal-weight (5)	Value-weight (6)	Equal-weight (7)	Value-weight (8)
Alpha	−0.000691 (−0.119)	−0.00321 (−0.547)	0.00276 (0.871)	−0.00163 (−0.639)	0.00248 (0.816)	0.00280 (0.989)	**0.00322*** (**1.677**)	0.000287 (0.403)
$MSCI_{all\ industries}$	1.619*** (6.528)	1.771*** (7.288)	1.366*** (15.246)	1.316*** (18.457)	1.108*** (15.389)	1.132*** (16.946)	1.616*** (15.843)	1.957*** (42.177)
$SMB_{all\ industries}$	−0.0428 (−0.058)	−0.178 (−0.239)	0.900* (1.761)	−0.205 (−0.598)	0.746* (1.789)	0.271 (0.749)	0.870*** (3.042)	0.365*** (2.786)
$HML_{all\ industries}$	0.359 (0.478)	0.137 (0.177)	−0.380 (−0.900)	−0.404 (−1.018)	−1.031** (−2.376)	−0.742* (−1.893)	−0.852** (−2.024)	−0.200 (−1.342)
$WML_{all\ industries}$	−0.605** (−2.327)	−0.444* (−1.682)	−0.183 (−1.158)	−0.0588 (−0.598)	0.00899 (0.082)	0.0863 (1.023)	−0.198 (−1.547)	−0.0281 (−0.600)
Ortho. $MSCI_{aerospace}$	−0.946** (−2.226)	−1.050** (−2.459)	−0.775*** (−3.894)	−1.316*** (−10.046)	−0.692*** (−4.644)	−0.971*** (−6.651)		
Ortho. $SMB_{industrials}$	0.579 (1.078)	0.688 (1.240)	−0.196 (−0.384)	0.420 (1.221)	−0.324 (−0.740)	−0.236 (−0.650)		
Ortho. $HML_{industrials}$	−1.445 (−1.557)	−1.298 (−1.363)	0.440 (0.849)	0.312 (0.778)	1.395*** (2.937)	0.955** (2.231)		
Ortho. $WML_{industrials}$	0.177 (0.588)	0.166 (0.550)	0.0512 (0.318)	0.139 (1.037)	−0.00696 (−0.048)	0.0263 (0.209)		
Ortho. $MSCI_{tobacco}$							−1.305*** (−8.106)	−1.902*** (−28.559)
Ortho. $SMB_{consumer\ staples}$							−0.433 (−1.392)	−0.189 (−1.512)
Ortho. $HML_{consumer\ staples}$							1.020** (2.277)	0.316 (1.557)
Ortho. $WML_{consumer\ staples}$							−0.0547 (−0.383)	−0.0656 (−0.948)
Observations	166	166	94	94	162	162	70	70
R^2	0.547	0.556	0.850	0.863	0.712	0.712	0.893	0.980
Adj. R^2	0.524	0.533	0.836	0.851	0.697	0.697	0.880	0.978

This table reports performance results for the exclusion portfolios of controversial weapons and tobacco, where the performance is measured according to an extended four-factor model that in addition to the global market, size, value and risk factors includes orthogonalised industry-style factors, i.e. factors reflecting the excess industry market return, the industry-specific size factor, the industry-specific value factor and the industry-specific momentum factor relating to the aerospace and defence / industrials industry and the tobacco / consumer staples industry. The results are presented by fund and by weighting scheme. Robust *t*-ratios are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively. Alpha estimates that are significant at the 10, 5 or 1 % level are highlighted in bold

accordance with the global size, value and momentum factors, described in "Methodology" section.[29] As the global risk factors are likely highly correlated with the industry-specific risk factors, we only add the differential industry effects of the risk factors to the model, using the orthogonalisation approach suggested by Elton et al. (1993) and applied in Adamsson and Hoepner (2015).[30] Table 6 presents the results when adding industry-style factors to the Four-Factor model.

We find that our main results do not significantly change although the *t*-statistics on the value-weighted portfolios shrink significantly and are now much closer to zero than to common significance levels. Still, only the equal-weighted portfolio of excluded tobacco companies generates a positive abnormal return of about 3.9 % per annum, confirming the patterns observed in the main analysis. Again, due to the low practical relevance of equal-weighted portfolios for the funds in our sample, we are cautious in drawing too strong performance implications based on this estimate. In contrast, these results are consistent with the finding by Adamsson and Hoepner (2015) that tobacco portfolios do not outperform in a real-world setting on a risk-and-factor-adjusted basis and hence their exclusion is not financially detrimental.

Sub-Sample Analysis

As another robustness test, we check whether our findings are the result of individual company effects due to the low number of excluded companies in the early part of the sample. To rule out this possibility, we restrict our sample to the years 2008–2015. From 2008 onwards, each fund excluded at least 13 companies, while most had a considerably larger number of exclusions (Table 1), assuring a reasonably diversified portfolio. The results of this sub-sample analysis are presented in Table 7.

The majority of the estimates remain qualitatively unchanged. Individual estimates become marginally significant or lose their significance over the sub-period. However, the cases of significant abnormal performance still tend to be fund-specific and/or depend on the weighting of returns. Thus, the sub-sample analysis indicates that our main results are unlikely to be driven by the dominance of single excluded companies in the early part of the sample.

Risk Comparison

While our main analysis focuses on the impact of exclusionary screening on funds' (risk-adjusted) returns, as a final robustness test, we address the question of whether the exclusion of unethical companies affects funds' risk characteristics. This analysis is partially motivated by the view that exclusionary screening is less a return enhancing but rather a risk-management tool. In line with this argument, Boutin-Dufresne and Savaria (2004) show that socially responsible portfolios have lower total risk as these portfolios are not exposed to the risks associated with companies' unethical business practices such as legal actions, strikes, boycotts and reputational damages, which the authors refer to as the unethical component of total risk. In addition, Lee et al. (2010) and Humphrey and Lee (2011) analyse the risk implications of exclusionary screening for samples of U.S. and Australian SRI mutual funds, respectively. However, the two studies arrive at different conclusions as to whether exclusionary screening increases or decreases portfolio risk, suggesting that the risk implications of exclusionary screening might depend on the way that exclusionary screens are applied in practice.

Inspired by Blake et al. (2013) and Hoepner et al. (2013), we test the risk implications of exclusionary screening by comparing the riskiness of the exclusion portfolios to that of the funds' benchmark index. Since the concept and definition of financial risk is not undisputed and many different risk measures have been suggested over the years, we employ a variety of risk measures that capture different aspects of financial risks. Firstly, following Lee et al. (2010) and Humphrey and Lee (2011), we examine the total risk of the portfolios as measured by the standard deviation of returns. The standard deviation of returns is a conventional risk measure in the finance literature to capture any deviations from an expected return, both negative and positive. We calculate the standard deviations of returns in the following way:

$$\mathrm{sd}_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(r_{xp,t} - \bar{r}_{xp}\right)^2}, \tag{5}$$

where sd_p is the standard deviation of daily excess returns of portfolio p over the recent month, $r_{xp,t}$ is the daily return in excess of the risk-free rate of portfolio p on day t, $\bar{r}_{\mathrm{xp}}$ is the average daily excess return of portfolio p over the

[29] Style Research only allows us to construct style factors at a broad sector level, i.e. for the tobacco industry we are only able to construct customised factors at the sector level of consumer staples and for the controversial weapons we are only able to generate the style factors for the industrial goods sector. In contrast, the MSCI industry indices are available for the specific industries, i.e. tobacco and aerospace and defence. To rule out that our results are affected by these different industry classifications, we replace the more specific MSCI industries with the MSCI sector returns matching the style sectors. The results remain qualitatively unchanged, except for AP1–AP4's equal-weighted portfolio of controversial weapons which now generates a significantly positive abnormal return.

[30] Each industry-level style factor is regressed on the corresponding economy-level factor. The clean, orthogonalised factor is represented by the residual plus the intercept of that regression.

Table 7 Subsample analysis—2008–2015

	Panel A: CAPM model					Panel B: four-factor model				
	Equal weighted		Value weighted			Equal weighted		Value weighted		
	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.	Alpha	Adj. R^2	Alpha	Adj. R^2	Obs.
	AP7									
All exclusions	0.00269 (1.319)	0.901	**0.00341*** (**1.874**)	0.849	93	0.00291 (1.500)	0.912	**0.00341*** (**1.927**)	0.867	93
Human rights issues	−0.00106 (−0.331)	0.805	−0.00229 (−0.700)	0.610	93	−0.000461 (−0.151)	0.825	−0.00211 (−0.664)	0.636	93
Labour rights issues	**0.00681**** (**2.115**)	0.695	0.00464 (1.497)	0.484	93	**0.00646*** (**1.964**)	0.699	**0.00467*** (**1.680**)	0.592	93
Environmental issues	−0.00364 (−0.847)	0.648	−0.00307 (−0.806)	0.587	93	−0.00375 (−0.866)	0.650	−0.00333 (−0.883)	0.590	93
Controversial weapons	0.00256 (0.901)	0.838	−0.000404 (−0.124)	0.818	93	0.00291 (1.051)	0.847	−0.000534 (−0.159)	0.813	93
Tobacco	–	–	–	–	–	–	–	–	–	–
	AP1–4									
All exclusions	**0.00588*** (**1.815**)	0.774	0.00468 (1.458)	0.448	93	**0.00658**** (**2.172**)	0.808	**0.00528*** (**1.790**)	0.557	93
Human rights issues	−0.00145 (−0.287)	0.487	−0.00297 (−0.572)	0.473	85	−0.00119 (−0.248)	0.498	−0.00236 (−0.477)	0.483	85
Labour rights issues	0.00490 (1.032)	0.357	0.00441 (1.005)	0.175	93	0.00567 (1.238)	0.403	0.00490 (1.260)	0.353	93
Environmental issues	**−0.0500*** (**−1.976**)	0.249	**−0.0506*** (**−1.992**)	0.250	25	−0.0135 (−0.794)	0.695	−0.0137 (−0.806)	0.697	25
Controversial weapons	0.00479 (1.305)	0.770	0.00283 (0.741)	0.645	93	0.00510 (1.491)	0.792	0.00314 (0.822)	0.636	93
Tobacco	–	–	–	–	–	–	–	–	–	–
	GPFG									
All exclusions	0.00326 (1.405)	0.881	0.00244 (1.341)	0.875	93	**0.00335*** (**1.701**)	0.907	0.00239 (1.406)	0.896	93
Human rights issues	0.00763 (1.490)	0.519	0.00367 (0.880)	0.212	93	**0.00875*** (**1.729**)	0.546	0.00420 (1.132)	0.374	93
Labour rights issues	–	–	–	–	–	–	–	–	–	–
Environmental issues	−0.00784 (−1.078)	0.618	**−0.0153*** (**−1.891**)	0.565	93	−0.00904 (−1.376)	0.682	**−0.0167**** (**−2.177**)	0.594	93
Controversial weapons	0.00430 (1.438)	0.793	**0.00544*** (**1.674**)	0.724	93	0.00473 (1.648)	0.806	**0.00588*** (**1.820**)	0.724	93
Tobacco	**0.00718**** (**2.309**)	0.591	**0.00833**** (**2.169**)	0.461	70	**0.00746*** (**1.948**)	0.608	0.00679 (1.598)	0.553	70

This table presents the results of the performance analysis for the sub-sample spanning January 2008–September 2015. Panel A presents the estimates from a CAPM, while Panel B reports results from a global four-factor model. The dependent variables are the continuously compounded excess returns on the funds' exclusion portfolios for all exclusions as well as for exclusions sorted by the reason for exclusion. The results are presented by fund and by weighting scheme. The table only reports the alpha estimates, the adjusted R^2 values and the number of observations per model. The coefficient estimates on the risk factors are omitted to preserve space, but are available upon request. Robust t ratios are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively. Alpha estimates that are significant at the 10, 5 or 1 % level are highlighted in bold

recent month, and T is equal to the number of trading days of the recent month.

Following Hoepner et al. (2013), we also employ several downside risk measures. These measures only account for the risk of negative deviations of returns from investors' expectation. In this sense, these measures better capture the risks associated with unethical business practices, such as unexpected and large negative shocks to returns, e.g. due to costs of lawsuits, strikes and boycotts. They also more strongly reflect investors' real attitudes towards risk as investors tend to fear losses but welcome larger than expected gains.

One measure that accounts for this asymmetry is the semi standard deviation, which can be regarded as a special case of the conventional standard deviation discussed above. The semi standard deviation only accounts for the negative deviations from expected returns and is computed as follows:

$$\text{ssd}_p = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\max\left[\left(\bar{r}_{xp} - r_{xp,t}\right), 0\right]^2}, \tag{6}$$

where ssd_p is the semi standard deviation of daily excess returns of portfolio p over the recent month. The maximum function assures that only returns below $\bar{r}_{xp}$ are considered.

In addition, we rely on several versions of the Lower Partial Moment (LPM^3) which is a commonly applied downside risk measure in more severe market conditions (Hoepner et al., 2013). The LPM^3 is calculated as:

$$\text{LPM}_p^3(\varphi) = \frac{1}{T-1}\sum_{t=1}^{T}\max\left[\left(\varphi - r_{xp,t}\right), 0\right]^3, \tag{7}$$

where LPM_p^3 is the lower partial moment of daily excess returns of portfolio p over the recent month and φ is the investor's minimally acceptable return.

The LPM^3 assumes highly risk-averse investors as it punishes large negative returns more strongly than small negative returns (i.e. it cubes instead of squares downside deviations). Lower Partial Moments are generally highly customisable and thus allow us to capture a variety of investor expectations and levels of risk aversion, whereby the magnitude of risk aversion increases with higher exponents (Eling and Schuhmacher 2007; Kaplan and Knowles 2004). Following Kaplan and Knowles (2004) and Hoepner et al. (2013), we choose an exponent of three (i.e. LPM^3), though our results are qualitatively unchanged when using a less conservative exponent of two instead. We use two alternatives for the minimally acceptable return φ to capture different investor expectations. Firstly, we employ the average monthly excess return of the portfolio p (i.e. $\varphi = \bar{r}_{xp}$). Secondly, we require returns to be non-negative (i.e. $\varphi = 0$). The latter case indirectly accounts for the possibility that the asset owners in our study might not be return maximising but invest against their share of a notional long-term liability. While we do not have access to the liability data of the AP-funds or the GPFG and hence cannot study this ambition in more detail, it seems reasonable to assume that asset owners investing against their share of notional long-term liabilities do not want to see the assets diminished in absolute terms.

Finally, we are interested in the highest possible loss that the portfolios might incur over a given investment period. This is captured by the minimum daily excess return of a portfolio over the recent month. This minimum return provides a good indication of whether excluding unethical companies protects the funds from incurring very large losses. The minimum return is calculated as:

$$\text{min.return}_{xp} = \text{min}_{xp,T}, \tag{8}$$

where $\text{min}_{xp,T}$ represents the minimum daily excess return on portfolio p over the recent month with T number of days.

Table 8 presents the estimates of the various risk measures for the MSCI index and the exclusion portfolios. We only report results on the value-weighted exclusion portfolios as they are more practically relevant for the funds' performance measurement and more suitable when compared to the (value-weighted) MSCI index. Panel A reports the monthly averages of the risk characteristics (together with their standard deviations in brackets) while Panel B shows the results of a paired t-test on the mean values for the MSCI index vis-à-vis the exclusion portfolios. The paired t-test is a standard statistical test that allows a comparison of mean values derived from different samples. It indicates whether the difference between the mean values is statistically significant or whether it could also simply be a result of large measurement error. Thus this test is particularly applicable in our case where the sample lengths of the exclusion portfolios differ across funds. As expected, the MSCI index exhibits the lowest risk based on all risk measures since it represents a more diversified portfolio compared to the exclusion portfolios, with a total of 2491 constituents as of the end of 2015. However, the daily returns of AP7's exclusion portfolio show comparable risk features, with only a slightly higher standard deviation and a slightly lower minimum return. In comparison, GPFG's exclusion portfolio appears the riskiest of all as it has the greatest average standard deviation, the lowest minimum daily returns, and features the highest values for the LPM^3 measures.

Next, we assess whether the riskiness of the funds' exclusion portfolios is statistically different from that of the MSCI benchmark index and hence, whether excluding these companies likely increases or decreases the funds'

Table 8 Risk measures for the MSCI index and the exclusion portfolios

	MSCI	GFPG	AP7	AP1–4
Panel A				
Standard deviation	0.0088 (0.0055)	0.0105 (0.0056)	0.0093 (0.0055)	0.0096 (0.0053)
Semi standard deviation	0.0062 (0.0039)	0.0073 (0.0042)	0.0065 (0.0040)	0.0067 (0.0039)
LPM^3 ($\bar{r}_p$ as min. acc. return)	0.0000 (0.0000)	0.0000 (0.0001)	0.0000 (0.0000)	0.0000 (0.0000)
LPM^3 (0 as min. acc. return)	0.0000 (0.0000)	0.0000 (0.0001)	0.0000 (0.0000)	0.0000 (0.0000)
Minimum return	−0.0175 (0.0124)	−0.0213 (0.0135)	−0.0183 (0.0108)	−0.0195 (0.0124)
		Difference between MSCI and…		
		GPFG	AP7	AP1–4
Panel B				
Standard deviation		−0.0017*** (−5.9022)	−0.0005*** (−3.7693)	−0.0002 (−0.6570)
Semi standard deviation		−0.0011*** (−4.4803)	−0.0002 (−1.5690)	0.00003 (0.1254)
LPM^3 ($\bar{r}_p$ as min. acc. return)		−0.000001* (−1.6752)	0.0000 (−0.6526)	0.0000 (0.0444)
LPM^3 (0 as min. acc. return)		0.000000 (−1.1054)	0.00000 (0.2371)	0.000001 (1.6091)
Minimum return		0.0038*** (3.3318)	0.0008 (1.0965)	−0.0001 (−0.1262)

Panel A reports mean values and standard deviations (in brackets) of the monthly risk measures for the MSCI index and the value-weighted exclusion portfolios of the three fund groups. Panel B reports mean differences in the monthly risk measures for the MSCI index and these exclusion portfolios. The numbers in brackets represent *t* values for a paired *t* test of the mean values of the MSCI vis-à-vis the funds' exclusion portfolios. The calculation of the risk measures is described in "Risk Comparison" section. *, **, *** indicate statistical significance at the 10, 5 and 1 % levels, respectively

risk. To do so, we turn to the results of the paired *t*-test on the means of the risk measures, presented in Panel B of Table 8. Only GPFG's exclusion portfolio seems to systematically differ from the MSCI. In particular, as indicated by the majority of risk measures, the GPFG's exclusion portfolio tends to be riskier than the MSCI index. This implies that excluding these companies might protect GPFG from incurring losses. In comparison, the exclusions of the AP-funds are unlikely to result in significant risk implications for their overall portfolios.

Conclusion

Divesting from companies that are associated with unethical business practices, such as the violation of human and labour rights or environmental pollution, represents one way to protect investors against complicity in these activities. In particular, two of the world's largest public asset owners, Norway's GPFG and Sweden's AP-funds, have adopted such exclusionary screening to ensure that their investments live up to the ethical standards expected from them by the general public. However, the funds also need to meet the financial objectives set out by the national legislation which requires them to maximise financial returns. As previous research suggests that exclusionary screening harms financial performance, the conflicting expectations of meeting ethical standards while maximising financial wealth present the funds with a dilemma: Does the exclusion of unethical companies inevitably mean sacrificing financial returns or can investors achieve both, their financial and ethical objectives? This is the question that we address in this study. In particular, we empirically analyse the performance effect of excluding companies from the investment universe of the GPFG and the AP-funds. We find that these exclusions neither financially harm the funds nor do they increase fund performance. This finding holds, both across funds and across different screening types. The only exception is the equal-weighted exclusionary screen of tobacco, which tends to outperform the fund's benchmark. While this finding provides initial evidence that the performance effect differs between norm-based and sector-based exclusionary screens we are very cautious when interpreting this finding, since the respective value-weighted portfolio does not outperform and hence this finding is more likely to result from small stocks effects than any tobacco characteristics (see also Adamsson and Hoepner 2015). Overall, we conclude that the exclusionary screening practiced by the GPFG and the AP-funds enables the funds to incorporate their beneficiaries' interest without compromising returns and might provide a promising route for other (non-SRI) investors to avoid criticism regarding their legitimacy and social usefulness that has emerged after the financial crisis.

However, our findings are subject to several limitations. Firstly, we are cautious in extending our findings of an insignificant performance effect of exclusionary screening on any form of exclusionary screens adopted by investors. Instead, we acknowledge that the relation between performance and exclusionary screens depends on the type and extent of the screens (Barnett and Salomon 2006; Renneboog et al. 2008b; Capelle-Blancard and Monjon 2014; Trinks and Scholtens 2015). For instance, investors from other societal backgrounds might be bound by different ethical obligations, whose impact on performance has not been analysed in this study (e.g. Salaber 2013; Fauver and McDonald 2014; Liu et al. 2014; Adamsson and Hoepner 2015). In addition, in unreported results we find great differences across the exclusion lists of comparable investors even in the Scandinavian SRI market which is known for its uniform approach towards exclusionary screening and a relatively homogeneous set of ethical standards (Bengtsson 2008a; Jensen 2016a).[31] Thus, while exclusionary screening offers a promising way to align ethical and financial objectives, the performance implications might strongly depend on the fund's particular screening approach as well as the ethical norms it represents.

Secondly, our study, in line with the majority of the academic literature, has only evaluated the financial implications of exclusionary screening, hence implicitly assuming that the applied screens satisfy the ethical demands of investors. However, given that these funds represent the interests of the entire population, including future generations, this assumption cannot be easily satisfied. To overcome this problem, the funds base their ethical standards on a set of minimally agreed principles, which are defined by the national laws as well as the states' commitments to international conventions. However, since the funds only react in hindsight (and often with a significant time lag) to accusations of breaches of these standards it would be an interesting route for future research to investigate whether exclusionary screening actually reduces funds' exposure to unethical business practices and thus achieves the objective of avoiding complicity in severe violations of ethical standards.

Thirdly, our findings do not provide any normative guidance as to what objectives should be given priority to, the ethical objectives or the financial objectives. This question is particularly relevant for the funds in our sample and distinguishes our study from the numerous studies on SRI mutual funds, as contrary to mutual fund investors, the beneficiaries of the GPFG and the AP-funds cannot exit the funds if they disagree with the funds' investment approach. While the legal guidelines of the Swedish AP-funds can be understood as prioritising financial objectives over ethical ones (e.g. Du Rietz 2016), Sandberg et al. (2014) criticise these regulations as too abstract and vague. In comparison, the guidelines given to the GPFG do not provide any instructions on how to resolve conflicts between ethical and financial objectives (Richardson 2011). Thus, a clarification of the funds' objectives and a clear prioritisation regarding ethical and financial demands by the legislator would not only relieve the funds from this conflict. This clarification might have the additional benefit of improving fund governance by reducing the scope to which other interests, especially political interests, might influence the funds' exclusion decisions. The latter has been a constant point of criticism that these funds have to face and that undermines their legitimacy with the general public (e.g. Clark and Monk 2010; Richardson 2011).[32]

Moreover, while exclusionary screening can represent a powerful tool for legislators and policy makers to safeguard themselves against accusations of complicity in unethical behaviour, exclusionary screening, by itself, does not represent an appropriate tool for addressing societal and social change. For instance, considering the issue of climate change which both, the GPFG and the AP-funds, acknowledge as one of their major challenges in the future, Richardson (2011) points out that climate change is caused by the aggregate of small-scale environmental damages while exclusions only target "severe environmental damage". In other words, the threshold that leads to action is too high to meaningfully tackle climate change. Thus, in order to target social challenges such as climate change, diversity and equality, exclusionary screening has to be combined with other approaches such as engagement and dialogue that encourage companies to change their business practices.

Finally, an interesting question, though not the focus of our study, are the implications of the exclusionary

[31] In unreported results, we compare the most recent exclusion lists of 12 Scandinavian asset owners and asset managers to the exclusion lists of the GPFG and the AP-funds and find great heterogeneity across the lists. For instance, of the 191 different companies that were excluded, only 1.6 % were excluded by all funds and 36 % only appeared on the list of a single fund. The exclusion list of GPFG covers 34 % of all excluded companies, while AP7 covers 24 % and AP1–4 only 11 %, respectively. These findings are in line with those of Blanc and Cozic (2012) who compare the exclusion lists of several European investors. Limits in data availability and quality do not allow us to perform a more formal analysis of the performance implications of the exclusions adopted by the Scandinavian investors. Results of this preliminary comparison are available from the authors upon request.

[32] A striking illustration of political interests impacting exclusion decisions is the AP-funds' different treatment of SAS, Scandinavian Airlines System, in which the Swedish state is a large shareholder. While AP7 excluded the company as early as 2004 after SAS has been accused of breaching international competition law, SAS is still part of the other AP-funds' portfolios (Bengtsson 2008a).

screening for the excluded companies. Proponents of the exclusionary screening approach often claim that coordinated exclusions by investors might depress the stock price of the company and put pressure on the company to change its business practices. However, prior studies that analyse such coordinated divestments of large investor groups, e.g. the divestment of U.S. public asset owners from companies in South Africa during the Apartheid regime (Teoh et al. 1999; Grossman and Sharpe 1986; Ennis and Parkhill 1986; Wagner et al. 1984) or the Sudan Divestment Act in 2007 (GAO 2010), found little impact of these actions on the divested companies. This is in line with theoretical findings by Heinkel et al. (2001) who conclude that divestments only have the potential to change corporate behaviour when they are adopted by a critical number of investors representing a significant share of a company's shareholdings. Nevertheless, there is some anecdotal evidence that exclusions can occasionally initiate the desired change. For instance, after GPFG had excluded Rio Tinto, the company sought re-inclusion and GPFG entered a dialogue with Rio Tinto about how it could redeem itself (Richardson 2011). A more uniform approach of norm-based divestments among global asset owners might increase their influence on corporations and the reputational costs to the shunned company, and lead to more companies like Rio Tinto entering into a dialogue with the asset owners. First attempts of creating a universal list of "unethical" companies to guide exclusion decisions have been discussed by Belgian policy makers and provide a promising route for future regulations (Blanc and Cozic 2012).

Thus, to conclude our findings have important implications for the funds in our sample and especially their fund governance, for legislator and policy-makers, for other global investors, as well as for the excluded companies. Having focused on the exclusionary screens of two large public asset owners, we have extended the literature on exclusionary screening by studying a widely underrepresented investor group and underexplored type of screen. However, as the above discussion highlights our findings pose additional questions that represent interesting opportunities for future research.

Acknowledgments We would like to thank Hampus Adamsson, Marcus Nilsson, Simone Varotto, Pei-Shan Yu, Yeqin Zeng and two anonymous referees for helpful comments as well as Peter Hopkins and Robert Schwob for providing us access to the Style Research database. We are responsible for all remaining errors and omissions. The views expressed in this paper are not necessarily shared by the Principles for Responsible Investment. Authors are listed in alphabetical order.

References

Adamsson, H., & Hoepner, A. (2015). The 'price of sin' aversion: Ivory tower illusion or real investable alpha? *Working paper*.

Adler, T., & Kritzman, M. (2008). The cost of socially responsible investing. *Journal of Portfolio Management, 35*(1), 52–56.

Amer, E. (2015). The penalization of non-communicating UN Global Compact's companies by investors and its implications for this initiative's effectiveness. *Business & Society, 58*, 1–37.

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

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

Bengtsson, E. (2008a). Socially responsible investing in Scandinavia—A comparative analysis. *Sustainable Development, 16*, 155–168.

Bengtsson, E. (2008b). A history of Scandinavian socially responsible investing. *Journal of Business Ethics, 82*, 969–983.

Blake, D., Rossi, A. G., Timmermann, A., Tonks, I., & Wermers, R. (2013). Decentralized investment management: evidence from the pension fund industry. *Journal of Finance, 68*(3), 1133–1178.

Blanc, D., & Cozic, A. (2012). Norm-based exclusions: How responsible investors handle controversial companies. *Novethic Reseach Report*. Retrieved November 6, 2015, from http://www.google.co.uk/url? http://www.novethic.fr.

Boutin-Dufresne, F., & Savaria, P. (2004). Corporate social responsibility and financial risk. *Journal of Investing, 13*(1), 57–66.

Capelle-Blancard, G., & Monjon, S. (2014). The performance of socially responsible funds: Does the screening process matter? *European Financial Management, 20*(3), 494–520.

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

Chambers, D., Dimson, E., & Ilmanen, A. (2012). The norway model. *Journal of Portfolio Management, 38*(2), 67–81.

Clark, G. L. (2004). Pension fund governance: Expertise and organizational form. *Journal of Pension Economics and Finance, 3*(2), 233–253.

Clark, G. L., & Monk, A. (2010). Norway's government pension fund global: Ethics over efficiency. *Rotman International Journal of Pension Management, 3*(1), 14–19.

Dasgupta, S., Laplante, B., & Mamingi, N. (2001). Pollution and capital markets in developing countries. *Journal of Environmental Economics and Management, 42*(2), 310–335.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analyst Journal, 61*(2), 51–63.

Derwall, J., Koedijk, K., & Ter Horst, J. (2011). A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance, 35*(8), 2137–3147.

Dimson, E., Kreutzer, I., Lake, R., Sjo, H., & Starks, L. (2013). Responsible investment and the Norwegian government pension fund global. Main Report and Executive Summary, Strategy Council of the Norwegian Government Pension Fund.

Du Rietz, S. (2016). Responsible investment in Sweden. In T. Hebb, et al. (Eds.), *The Routledge handbook of responsible investment* (1st ed., pp. 336–344). New York: Routledge.

Durand, R. B., Koh, S., & Limkriangkrai, M. (2013a). Saints versus sinners. Does morality matter? *Journal of International Financial Markets, Institutions and Money, 24*(4), 166–183.

Durand, R. B., Koh, S., & Tan, P. L. (2013b). The price of sin in the Pacific-Basin. *Pacific-Basin Finance Journal, 21*(1), 899–913.

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

Elton, E. J., Gruber, M. J., Das, S., & Hlavka, M. (1993). Efficiency with costly information: A reinterpretation of evidence from managed portfolios. *Review of Financial Studies, 6*(1), 1–22.

Ennis, R. M., & Parkhill, R. L. (1986). South African divestment: Social responsibility or fiduciary folly? *Financial Analyst Journal, 42*(4), 30–38.

Fabozzi, F. J., Ma, K. C., & Oliphant, B. J. (2008). Sin stock returns. *Journal of Portfolio Management, Fall, 35*, 82–94.

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

Fauver, L., & McDonald, M. B. (2014). International variation in sin stocks and its effects on equity valuation. *Journal of Corporate Finance, 25*, 173–187.

Flammer, C. (2013). Corporate social responsibility and shareholder reaction: The environmental awareness of investors. *Academy of Management Journal, 56*(3), 758–781.

Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Returns and risk in socially responsible investment. *Journal of Banking & Finance, 32*(12), 2646–2654.

GAO. (2010). United States government accountability office: Testimony before the subcommittee on international monetary policy and trade on sudan divestment. GAO-11-245T. Retrieved November 6, 2015, from www.gao.gov/new.items/d11245t.pdf.

Grossman, B. R., & Sharpe, W. F. (1986). Financial implications of South African Divestment. *Financial Analyst Journal, 42*(4), 15–29.

Gupta, S., & Goldar, B. (2005). Do stock markets penalize environment-unfriendly behaviour? Evidence from India. *Ecological Economics, 52*(1), 81–95.

Hanhardt, A., & Ansotegui, C. (2008). Do the Fama and French factors proxy for state variables that predict macroeconomic growth in the Eurozone? *Working paper*.

Hawley, J. P. (2016). Setting the scene: The basics and basis of responsible investment. In T. Hebb, et al. (Eds.), *The Routledge handbook of responsible investment* (1st ed., pp. 16–33). New York: Routledge.

Heinkel, R., Kraus, A., & Zechner, J. (2001). The effect of green investment corporate behaviour. *Journal of Financial and Quantitative Analysis, 36*(4), 431–449.

Hirsh, C. E., & Cha, Y. (2015). Employment discrimination lawsuits and corporate stock prices. *Social Currents, 2*(1), 40–57.

Hoepner, A., Rezec, M., & Siegl, K. S. (2013). Does pension funds' fiduciary duty prohibit the integration of environmental responsibility criteria in investment processes? A realistic prudent investment test. *Working paper*.

Hoepner, A., & Zeume, S. (2014). Fiduciary duty and 'sin stocks': Is vice really nice? In J. P. Hawley, et al. (Eds.), *Cambridge handbook of institutional investment and fiduciary duty* (1st ed., pp. 181–206). Cambridge: Cambridge University Press.

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

Humphrey, J. E., & Lee, D. D. (2011). Australian socially responsible funds: performance, risk and screening intensity. *Journal of Business Ethics, 102*(4), 519–535.

Humphrey, J. E., & Tan, D. T. (2014). Does it really hurt to be responsible? *Journal of Business Ethics, 122*(3), 375–386.

Ince, O. S., & Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research, 29*(4), 463–479.

Jensen, A. (2016a). Introduction to Scandinavia. In T. Hebb, et al. (Eds.), *The Routledge handbook of responsible investment* (1st ed., pp. 313–316). New York: Routledge.

Jensen, A. (2016b). Responsible investment in Norway. In T. Hebb, et al. (Eds.), *The Routledge handbook of responsible investment* (1st ed., pp. 345–354). New York: Routledge.

Kaplan, P. D., & Knowles, J. A. (2004). Kappa: A generalized downside risk-adjusted performance measure. *Journal of Performance Measurement, 8*, 42–54.

Kappel, V., Schmidt, P., & Ziegler, A. (2009). Human rights abuse and corporate stock performance—An event study analysis. *Working paper*.

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management, 13*(5), 908–922.

Klassen, R. D., & McLaughlin, C. P. (1996). The impact of environmental management on firm performance. *Management Science, 42*(8), 1199–1214.

Konar, S., & Cohen, M. (1997). Information as regulation: The effect of community right to know laws on toxic emissions. *Journal of Environmental Economics and Management, 32*(1), 109–124.

Kurtz, L. (2005). Answers to four questions. *Journal of Investing, 14*(3), 125–139.

Lee, D. D., Humphrey, J. E., Benson, K. L., & Ahn, J. Y. (2010). Socially responsible investment fund performance: The impact of screening intensity. *Accounting & Finance, 10*(2), 351–370.

Li, Q., Vassalou, M., & Xing, Y. (2006). Sector investment growth rates and the cross section of equity returns. *Journal of Business, 79*(3), 1637–1665.

Liu, Y., Lu, H., & Veenstra, K. (2014). Is sin always a sin? The interaction effect of social norms and financial incentives on market participants' behaviour. *Accounting, Organizations and Society, 39*, 289–307.

Lobe, S., & Walkshäusl, C. (2011). Vice versus virtue investing around the world. *Working paper*.

Meller, S., & Husson-Traore, A.-C. (2013). Controversial companies: Do investor blacklists make a difference. *Novethic Reseach Report*. Retrieved November 6, 2015, from http://www.google.co.uk/url? http://www.novethic.fr.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance, 42*, 483–510.

Myklebust, T. (2010). The Norwegian government pension fund: Moving forward on responsible investing and governance. *Rotman International Journal of Pension Management, 3*(1), 20–23.

Norwegian Ministry of Finance. (2015). The Management of the Government Pension Fund in 2014. Report no. 21 (2014–2015), Report to the Storting (white paper).

Norwegian Ministry of Finance. (2016). The Management of the Government Pension Fund in 2015. Report no. 23 (2015–2016), Report to the Storting (white paper).

Petersen, H. L., & Vredenburg, H. (2009). Morals or economics? Institutional investor preferences for corporate social responsibility. *Journal of Business Ethics, 90*, 1–14.

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

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

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

Richardson, B. J. (2011). Sovereign wealth funds and the quest for sustainability: Insights from Norway and New Zealand. *Nordic Journal of Commercial Law, 2*, 1–27.

Salaber, J. (2013). Religion and returns in europe. *European Journal of Political Economy, 32*, 149–160.

Sandberg, J., Siegl, S., & Hamilton, I. (2014). The regulation of institutional investment in Sweden: A role model for the promotion of responsible investment? In J. P. Hawley, et al. (Eds.), *Cambridge handbook of institutional investment and fiduciary duty* (1st ed., pp. 59–71). Cambridge: Cambridge University Press.

Scholtens, B., & Sievänen, R. (2013). Drivers of socially responsible investing: A case study of four Nordic countries. *Journal of Business Ethics, 115*, 605–616.

Severinson, C., & Stewart, F. (2012). Review of the Swedish National Pension Funds. *OECD working papers on finance, insurance and national pension funds*, No. 17, OECD Publishing.

Sparkes, R., & Cowton, C. J. (2004). The maturing of socially responsible investment: A review of the developing link with corporate social responsibility. *Journal of Business Ethics, 52*, 45–57.

Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analyst Journal, 65*(4), 33–46.

Teoh, S. H., Welch, I., & Wazzan, C. P. (1999). The effect of socially activist investment policies on the financial markets: Evidence from the south african boycott. *Journal of Business, 72*(1), 35–89.

Trinks, P. J., & Scholtens, B. (2015). The opportunity cost of negative screening in socially responsible investing. *Journal of Business Ethics, 30*, 1–16.

Visaltanachoti, N., Zou, L., & Zheng, Q. (2009). The performances of sin stocks in China. *Working paper*.

Wagner, W. H., Emking, A., & Dixon, R. L. (1984). South African divestment: the investment issues. *Financial Analyst Journal, 40*(6), 14–22.