

Ethical Exclusions and Realized Excess Returns: Evidence from the Norwegian Government Pension Fund Global

Adrian Ruud Wiker

Dissertation written under the supervision of Professor Eva Schliephake

Dissertation submitted in partial fulfilment of requirements for the MSc in International Finance, at Universidade Católica Portuguesa and for the MSc in Business, major in finance at BI Norwegian Business School, May 2026.

Ethical Exclusions and Realized Excess Returns: Evidence from the Norwegian Government Pension Fund Global

by Adrian Ruud Wiker

May 2026

Abstract

This thesis examines whether companies excluded from the Norwegian Government Pension Fund Global (GPFG) earned measurable realized excess returns after exclusion. I construct a monthly dataset of 205 excluded firms from January 2005 to February 2026 by combining exclusion data from Norges Bank Investment Management (NBIM) and the Council on Ethics with firm-level return and market value data from Refinitiv Datastream. Each firm is tracked only during its active exclusion period, and monthly exclusion portfolios are formed by exclusion criterion.

The main analysis estimates Fama-French five-factor alphas and applies mean-variance spanning tests using MSCI ACWI as a broad global equity benchmark proxy. Active returns relative to this benchmark are also reported as a supplementary comparison. The aggregate value-weighted exclusion portfolio earns a positive and statistically significant annualized abnormal return of 6.82%. The strongest evidence is concentrated in Coal and Weapons. The Coal result remains positive after controlling for a coal-industry portfolio, while the Weapons result is more sensitive to defense-sector controls. Several excluded portfolios also show return and risk characteristics not fully captured by MSCI ACWI.

Overall, the findings suggest that GPFG-excluded firms earned measurable realized excess returns, but the evidence is not uniform across exclusion criteria. The thesis therefore provides a criterion-specific estimate of one financial dimension of ethical exclusions, rather than an exact estimate of the GPFG's fund-level opportunity cost.

Keywords: Ethical exclusions, GPFG, sovereign wealth funds, responsible investment, abnormal returns, Fama-French five-factor model, mean-variance spanning

J.E.L. Classification: G11, G12, G23, Q56

Ethical Exclusions and Realized Excess Returns: Evidence from the Norwegian Government Pension Fund Global

por Adrian Ruud Wiker

Maio de 2026

Resumo

Esta dissertação examina se as empresas excluídas do Norwegian Government Pension Fund Global (GPFG) obtiveram retornos excedentes realizados mensuráveis após a exclusão. Construo uma base de dados mensal com 205 empresas excluídas entre janeiro de 2005 e fevereiro de 2026, combinando dados de exclusão da Norges Bank Investment Management (NBIM) e do Conselho de Ética com dados de retornos e valor de mercado provenientes da Refinitiv Datastream. Cada empresa é acompanhada apenas durante o seu período ativo de exclusão, e as carteiras mensais são formadas por critério de exclusão.

A análise principal estima alfas de Fama-French de cinco fatores e aplica testes de abrangência média-variância, utilizando o MSCI ACWI como proxy de um índice global de ações. Os retornos ativos face a este índice são reportados como comparação suplementar. A carteira agregada ponderada pelo valor de mercado apresenta um retorno anormal anualizado positivo e estatisticamente significativo de 6,82%. A evidência mais forte concentra-se em Carvão e Armas. O resultado para Carvão permanece positivo após controlar por uma carteira da indústria do carvão, enquanto o de Armas é mais sensível a controlos do setor da defesa. Várias carteiras excluídas apresentam características de retorno e risco não totalmente capturadas pelo MSCI ACWI.

Os resultados sugerem que as empresas excluídas pelo GPFG obtiveram retornos excedentes mensuráveis, embora a evidência não seja uniforme entre critérios. Assim, a dissertação fornece uma estimativa por critério de uma dimensão financeira das exclusões éticas, não uma estimativa exata do custo de oportunidade do GPFG.

Palavras-chave: Exclusões éticas, GPFG, fundos soberanos, investimento responsável, retornos anormais, modelo de cinco fatores de Fama-French, mean-variance spanning

Classificação J.E.L.: G11, G12, G23, Q56

Acknowledgements

I would first like to thank my supervisor, Professor Eva Schliephake, for her guidance, feedback, and support throughout this thesis. Our meetings, email exchanges, and discussions have been highly valuable in shaping the direction and quality of this work.

I am also grateful to the professors at BI and Católica for a challenging and rewarding learning journey. Their courses have given me a strong academic foundation and helped deepen my interest in finance.

I would also like to thank my colleagues at Sumo Capital in Lisbon for their flexibility, support, and encouragement, which made it possible to work full-time while completing this thesis.

Finally, I am deeply thankful to my family and friends for their constant support, encouragement, and patience. Their belief in me has meant more than they know.

Declaration of AI Use

AI tools that have been used in the work on assignment/exam:

Claude Code Opus 4.6, ChatGPT 5.5, Grammarly

Purpose of AI use:

“I have used AI while undertaking my assignment in the following ways (circle YES or NO):

- To develop research questions on the topic – NO
- To generate ideas – NO
- To create an outline of the topic – YES
- To explain concepts – YES
- To support my use of language – YES
- To organize data – YES
- To analyse data – YES
- To visualize data- YES
- In other ways, as described below:
 - Review and discuss with me to improve my text
 - Help create and improve Python code
 - To translate abstract to Portuguese
 - Help with LaTeX code

Table of Contents

1	Introduction	9
2	Literature Review	11
2.1	Negative screening, sin stocks, and expected returns	11
2.2	GPFG exclusions and previous empirical evidence	11
2.3	Mean-variance spanning and portfolio opportunity sets	12
3	Data	13
3.1	Exclusion list and sample definition	13
3.2	Firm-level returns and market values	14
3.3	Exclusion windows and special cases	15
3.4	Portfolio classification	16
3.5	Risk factors, benchmark data, and industry controls	17
3.6	Descriptive statistics	17
3.7	Data limitations	18
4	Methodology	19
4.1	Calendar-time portfolio construction	19
4.2	Hypotheses	20
4.3	Abnormal-return tests	20
4.4	Sector-control regressions	21
4.5	Supporting benchmark-proxy active returns	21
4.6	Benchmark-proxy spanning tests	21
4.7	Interpretation of the empirical design	22
5	Empirical Results	23
5.1	Fama-French five-factor alphas	23
5.2	Sub-period evidence	24
5.3	Sector-control results for Coal and Weapons	25
5.4	Supporting benchmark-relative active returns	26
5.5	Mean-variance spanning results	26
5.6	Robustness checks	28
5.6.1	Equal-weighted results	28
5.6.2	Concentration and top-name exclusion	29
5.6.3	Newey-West lag test	31
5.7	Summary of empirical findings	32
6	Discussion	33

6.1	Interpreting the main finding	33
6.2	Mechanisms	33
6.3	Criterion-level interpretation	34
6.4	Return effects versus benchmark spanning effects	36
6.5	Relation to previous literature	36
6.6	Policy relevance for the ethical-framework review	37
6.7	Limitations and alternative explanations	38
7	Conclusion	40
	Appendix	41

Table of Tables

3.1	Sample-selection waterfall	14
3.2	Exclusion portfolios and sample coverage	16
3.3	Descriptive return statistics	17
5.1	Full-sample Fama-French five-factor alphas, value-weighted	23
5.2	Fama-French alphas across sample windows	24
5.3	Sector-control regressions for Coal and Weapons	25
5.4	Active returns relative to MSCI ACWI	26
5.5	Mean-variance spanning tests, market-value-weighted specification	27
5.6	Equal-weighted Fama-French five-factor alphas	29
5.7	Portfolio concentration measures	29
5.8	Top-name exclusion robustness, value-weighted specification	30
5.9	Fama-French five-factor alphas: Newey-West lag robustness	31
5.10	Mean-variance spanning tests: Newey-West lag robustness	31
A1	Removed firms by reason	41
A2	Portfolio classification mapping	42
A3	Special data cases	43
A4	Back-of-the-envelope calculation	44
A5	Full Fama–French five-factor regression estimates, value-weighted portfolios .	45
A6	Full Fama–French five-factor regression estimates, equal-weighted portfolios .	46
A7	Extended sector-control regressions for Coal and Weapons	47

Introduction

The Norwegian Government Pension Fund Global (GPFG), known as the Oil Fund, invests Norway's petroleum revenues in global financial markets. Its purpose is to transform temporary petroleum income into lasting financial wealth for current and future generations. The fund's financial objective is to achieve the highest possible return after costs at an acceptable level of risk, with responsible management as part of its mandate (Ministry of Finance, 2026).

Ethical exclusions create a financial dilemma. The GPFG is a large global investor managed closely to a benchmark but cannot invest in companies that meet certain exclusion criteria. While these exclusions may be ethically justified, they reduce the fund's investable universe. The central financial question is whether excluded firms earned higher returns after exclusion.

The institutional setup is divided between three main actors. The Ministry of Finance sets the investment mandate and the ethical guidelines. NBIM manages the fund within this framework, while the Council on Ethics assesses companies for observation or exclusion. Once a company is excluded, the fund is no longer allowed to invest in it (Ministry of Finance, 2026).

Norway has excluded companies from the GPFG since the early 2000s. The framework includes product-based criteria, such as tobacco, cannabis, certain weapons, and coal, and conduct-based criteria, such as human rights violations, severe environmental damage, gross corruption, and serious breaches in situations of war or conflict (Ministry of Finance, 2022). This distinction matters because the financial effects of exclusion may differ across criteria.

The question is especially relevant because the ethical framework is now under formal review. On 4 November 2025, the Storting asked the government to review the framework and pause new observation and exclusion decisions until a new framework is decided (Ministry of Finance, 2025b). On 7 November 2025, the government appointed a committee chaired by Svein Gjedrem to assess the framework and propose changes by 15 October 2026 (Ministry of Finance, 2025a). The committee's mandate is forward-looking, asking how exclusions may affect the fund's risk, return, and diversification. This is why historical evidence on the financial consequences of past exclusions is relevant to the current review.

Recent events show why the review has become urgent. In August 2025, NBIM announced the exclusion of Caterpillar under the war-and-conflict criterion, based on serious violations of individuals' rights in situations of war or conflict (NBIM, 2025; Council on Ethics, 2025). The case illustrates a broader tension in the current framework: exclusions may increasingly involve large global companies whose ordinary products or services are connected to complex conflict situations. While Caterpillar only adds a few months of data to this thesis, and War and Conflict is not where the strongest return evidence is found, the case illustrates the motivation for this research.

This policy context leads to the thesis’s main empirical question: if exclusions reduce the GPFG’s investable universe, how did the excluded companies perform afterwards? More specifically, the thesis asks:

Have firms excluded from the GPFG earned measurable realized excess returns, either through abnormal returns relative to standard risk factors or through risk and return not already captured by a broad global equity benchmark proxy?

This question has two parts. First, did excluded firms earn abnormal returns after controlling for standard risk factors? Second, did excluded portfolios differ from the benchmark in risk and return? Together, these tests assess the performance of firms outside the GPFG’s investable universe.

Previous studies reach different conclusions. Hoepner and Schopohl (2018) study exclusions by the GPFG and Swedish AP funds and find limited evidence that exclusionary screening harmed performance. Berle et al. (2025) study a more recent GPFG sample and report a positive annualized alpha of around 5% for an “unethical” portfolio. This leaves three questions open. Does the positive return evidence remain when the sample is extended to the current review period? Does the evidence differ across exclusion criteria? And did excluded firms have risk and return not already captured by broad global equity exposure?

To answer these questions, this thesis extends previous work by updating the sample until February 2026, tracking firms only during active exclusion, reporting results by exclusion reason, and using both Fama-French abnormal return tests and mean-variance spanning tests. Active returns relative to the benchmark are also reported as a supplementary comparison.

Empirically, excluded firms are matched to return and market-value data. Monthly portfolios are constructed for firms under active exclusion. The analysis examines both the aggregate portfolio and results by exclusion criterion.

The main finding is a positive and statistically significant annualized abnormal return of 6.82% for the aggregate exclusion portfolio. The strongest evidence is in Coal and Weapons, though the Weapons result is sensitive to defense-sector controls. Several portfolios are also not redundant relative to the benchmark. These findings show that excluded firms earned higher returns, but the contribution is not simply that excluded firms outperformed. Rather, the realized financial effects of ethical exclusions are criterion-specific and benchmark-dependent.

The remainder of the thesis is organized as follows. Chapter 2 reviews related literature. Chapter 3 describes the data and portfolio construction. Chapter 4 presents the methodology. Chapter 5 reports empirical results. Chapter 6 discusses interpretation, policy relevance, and limitations. Chapter 7 concludes.

Literature Review

This thesis builds on three strands of literature. The first covers negative screening, sin stocks, and investor constraints. The second examines ethical exclusions in the GPFG context. The third develops mean-variance spanning tests, which are used to assess whether adding an asset changes an investor's opportunity set.

2.1 Negative screening, sin stocks, and expected returns

A central question for responsible investors is whether excluding firms for ethical or sustainability reasons lowers returns. Hong and Kacperczyk (2009) argue that social norms can affect asset prices because some investors avoid “sin” stocks. When fewer investors are willing to hold these firms, they may trade at lower valuations and offer higher expected returns. This premium is not caused by a single fund's decision, but reflects broader investor constraints and social norms.

The type of exclusion also matters. Trinks and Scholtens (2017) show that the financial effect of negative screening depends on which firms are excluded. For the GPFG, exclusions cover a wide range of companies. Removing coal firms, weapons producers, tobacco companies, or firms linked to human-rights violations may each yield different return effects. This thesis therefore examines exclusion criteria separately.

Pástor et al. (2021) reach a similar conclusion. Their model of sustainable investing distinguishes between “green” firms, which investors may prefer, and “brown” firms, which are more likely to be avoided. If many investors avoid controversial firms, these companies may offer higher expected returns to attract capital. This framework supports the idea that excluded firms can earn higher returns not only because of one fund's decision, but because many investors are reluctant to hold them.

However, exclusion performance depends not only on the excluded firms, but also on the alternatives. Blitz and Swinkels (2023) note that a loss only occurs if the excluded firms outperform the replacement investments. This distinction is central here because the thesis examines the returns of excluded firms, not the overall effect on the GPFG portfolio.

2.2 GPFG exclusions and previous empirical evidence

Empirical studies of GPFG exclusions reach different conclusions. Hoepner and Schopohl (2018) find limited evidence that screening harmed performance. In their sample, excluded firms did not clearly outperform the relevant benchmarks, suggesting that exclusions did not create a large financial cost.

Berle et al. (2025) provide the closest reference for this thesis. They study a portfolio of GPFG-excluded, or “unethical,” stocks and report a positive annualized alpha of about 5%. This means excluded firms earned more than standard risk factors predict. They also caution against

interpreting high post-exclusion returns as caused by GPFG actions. Their analysis finds no strong evidence of a lasting price effect around exclusion events. In this thesis, GPFG exclusions are used to identify firms outside the fund's universe, not to attribute causality for their returns.

This thesis follows Berle et al. (2025) in constructing calendar-time portfolios of excluded firms and estimating abnormal returns. Here, the sample is extended to February 2026, beyond their ending in 2021. Exclusion criteria are analyzed separately, and benchmark-based tests are included. The extended sample allows for assessment of whether the main result persists, where the strongest evidence lies, and whether excluded portfolios differ in risk and return from broad global equity.

2.3 Mean-variance spanning and portfolio opportunity sets

Mean-variance spanning tests assess whether adding an asset changes the investor's risk-return opportunity set. The test asks if an investor holding the benchmark could form a meaningfully different portfolio by adding the excluded portfolio. Huberman and Kandel (1987) formalize this with the minimum-variance frontier. If the benchmark already spans the excluded portfolio, adding it does not change the opportunity set. In this thesis, the test helps determine if excluded firms offer risk and return not already captured by broad global equity.

Kan and Zhou (2012) provide methods to test whether a portfolio is already covered by a benchmark. For this thesis, the question is whether excluded firms add risk or return not provided by broad global equities. If so, they may matter for diversification. However, different does not always mean better. A spanning rejection means the excluded portfolio differs from the benchmark in return, risk, or both.

Together, these strands of literature show that the thesis is not only about whether excluded firms performed well. The negative-screening literature suggests that controversial firms may earn higher returns when many investors avoid them. GPFG-specific studies show mixed evidence, with some finding positive abnormal returns. The spanning literature adds a portfolio perspective: even if excluded firms earn high returns, it matters whether they offer risk and return not already captured by broad global equity. This motivates the focus on both abnormal returns and benchmark-based portfolio effects.

Data

The empirical analysis relies on three core elements: which firms were excluded, the timing of their active exclusion, and their listed equity performance during those periods. Official exclusion data are combined with firm-level market data, benchmark returns, risk factors, and industry-portfolio data to construct the return series for empirical testing.

3.1 Exclusion list and sample definition

The analysis starts with the historical exclusion list of NBIM, which identifies excluded companies, exclusion reasons, decision types, and publication dates. This list is cross-checked with the recommendations of the Council on Ethics to verify the reason for each exclusion and, when relevant, the reason for any revocation (Council on Ethics, nd; NBIM, nd).

The sample runs from January 2005 to February 2026. The start date follows the establishment of the Council on Ethics in late 2004 and the operational start of the ethical framework in 2005 (Ministry of Finance, 2022). February 2026 is the final month included in the data collection and matching process.

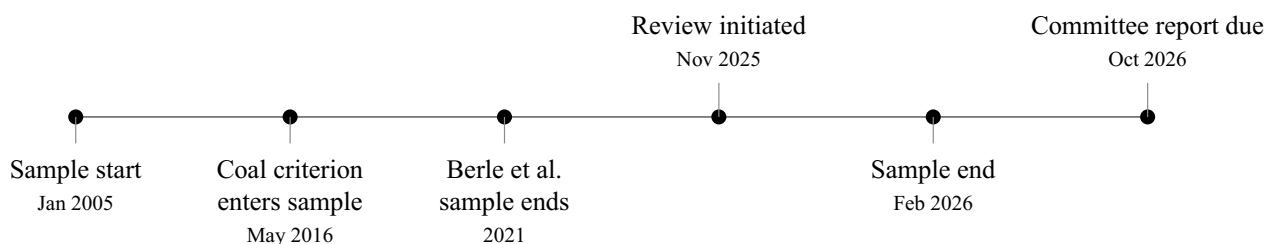


Figure 3.1: Timeline of the empirical sample and the GPFG ethical-framework review. The spacing is illustrative and not to scale.

The ethical framework distinguishes between observation and exclusion. Companies under observation are monitored but remain in the investable universe. Only formally excluded firms are included in the sample. If a company is first observed and later excluded, it enters the sample from the exclusion date. If an exclusion is revoked, the company is included only for the months when exclusion was active.

Table 3.1 summarizes the sample selection. The raw NBIM file contains 239 exclusion-related rows, corresponding to 236 unique firms because three firms are excluded more than once. Observation-only firms, exclusions announced after the sample period, firms without listed-equity data, and firms with unusable return series are removed. The final sample contains 205 unique firms and 20,171 firm-month observations.

Table 3.1: Sample-selection waterfall

Step	Selection criterion	Firms removed	Firms remaining
1	Raw NBIM master file	–	236
2	Observation-only firms	-19	217
3	Post-window exclusions	-5	212
4	No listed-equity data	-4	208
5	Unusable return series	-3	205
Final	Final empirical panel	-31	205

Note: Counts refer to unique firms. See appendix A1 for explanation.

3.2 Firm-level returns and market values

For each excluded company, the GPFG exclusion list is manually matched to Refinitiv Datastream. This is necessary because the exclusion list identifies companies, while Datastream provides data at the security level. Each firm is matched to a Total Return Index (RI) series and a market value (MV) series, both measured in U.S. dollars.

The RI series is used to calculate simple monthly total returns, including dividends. For each firm, the monthly return is the percentage change in RI from the last available trading day of the previous month to the last available trading day of the current month:

$$r_{i,t} = \frac{RI_{i,t}}{RI_{i,t-1}} - 1 \quad (3.1)$$

Simple returns are used because portfolio returns are calculated as weighted averages of firm returns. Market value is measured at month-end to determine portfolio weights. Returns are measured in USD, consistent with the Fama-French factors and Berle et al. (2025). Results are reported as USD listed-equity returns. Since the GPFG reports in its own currency basket and is often discussed in NOK, these estimates do not represent NOK accounting returns for the fund.

3.3 Exclusion windows and special cases

A firm enters the portfolio in the month after the public exclusion announcement, following the GPFG-specific approach in Berle et al. (2025). NBIM publishes exclusion decisions after the fund has sold its holdings, so the first full month after announcement provides a clean starting point for post-exclusion returns. A firm exits when the exclusion is revoked, when usable listed-equity data is no longer available, or at the end of the sample period. If a firm is excluded, reinstated, and later excluded again, each exclusion spell is treated as a separate active window.



Figure 3.2: Illustration of exclusion-window construction.

A few firms require manual adjustments due to listing changes, unusable return series, or corporate events. These cases are handled according to the active-window rules and documented in the appendix.

3.4 Portfolio classification

Firm-level observations are grouped into monthly exclusion portfolios. The broadest portfolio, ALL, includes all firms under active exclusion each month and represents the listed excluded firms in the dataset.

The ALL portfolio is split into Product-based and Conduct-based exclusions, following the main ethical distinction. Product-based exclusions are linked to what firms produce, while Conduct-based exclusions relate to firm behavior.

Six sub-portfolios are then created to test whether results differ by exclusion reason: Coal, Tobacco and Cannabis, Weapons, Environmental, War and Conflict, and Human Rights. These groups are based on the stated exclusion reasons in NBIM and Council on Ethics documents. Each firm belongs to only one sub-portfolio at a time. Tobacco and Cannabis are combined because both are product-based exclusions and the cannabis group is too small to analyze separately. The mapping from official reasons to empirical categories is provided in the appendix.

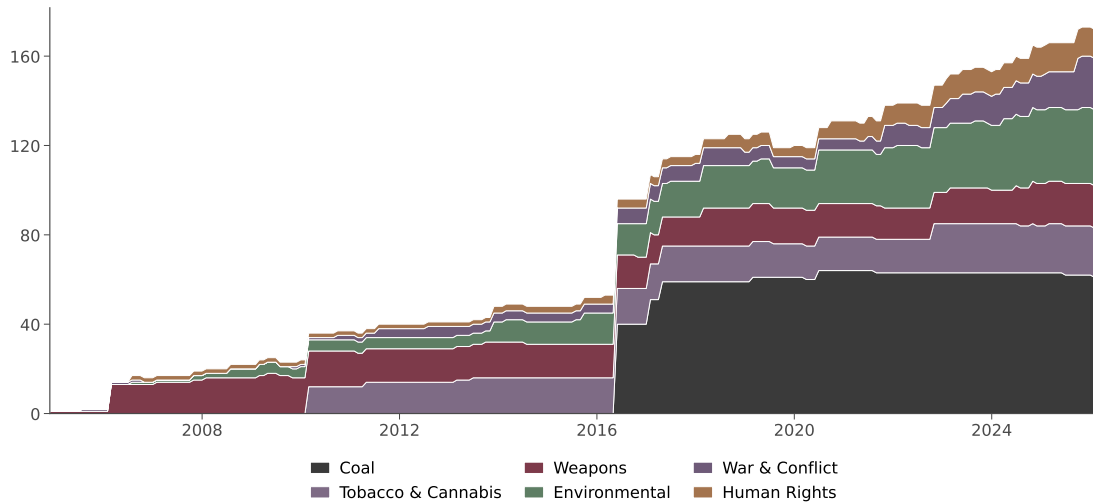


Figure 3.3: Active excluded firms over time by exclusion criterion.

Table 3.2: Exclusion portfolios and sample coverage

Portfolio	Firms	Firm-months	Months	First active month	Last active month
ALL	205	20,171	254	2005-01	2026-02
Product-based	115	13,863	254	2005-01	2026-02
Conduct-based	90	6,308	248	2005-07	2026-02
Coal	65	7,094	118	2016-05	2026-02
Tobacco and Cannabis	25	3,185	193	2010-02	2026-02
Weapons	29	3,742	254	2005-01	2026-02
Environmental	40	3,512	236	2006-07	2026-02
War and Conflict	30	1,426	212	2005-07	2026-02
Human Rights	16	1,212	236	2006-07	2026-02

Note: Firm counts refer to unique firms assigned to each portfolio.

3.5 Risk factors, benchmark data, and industry controls

To estimate abnormal returns, the Developed 5 Factors dataset from the Kenneth R. French Data Library is used, following Berle et al. (2025) and using the data provided by French (2026). The factors cover developed markets in North America, Europe, Japan, and Asia Pacific. Because some excluded firms are listed outside developed markets or have concentrated commodity exposure, the reported alphas should be interpreted as abnormal returns relative to this developed-market factor model, not relative to every possible global risk factor.

The benchmark proxy is MSCI ACWI. The GPFG's official equity benchmark is based on the FTSE Global All Cap Index, but a historical version suitable for this analysis was not accessible. MSCI ACWI is therefore used as a transparent public proxy for broad global listed equity exposure. This proxy is used for the active-return comparison and the mean-variance spanning tests.

Sector-control tests use Kenneth French industry portfolios. Coal is matched to the 30-industry Coal-mining portfolio. Weapons is matched to Aerospace and Guns from the 49-industry portfolios. These are U.S. sector proxies and do not perfectly control for global firms.

3.6 Descriptive statistics

Table 3.3 summarizes the raw return properties of the exclusion portfolios and MSCI ACWI for the full sample. For each portfolio, it reports the number of monthly observations, annualized average return, annualized volatility, and unconditional Sharpe ratio. The exclusion portfolios use value-weighted returns, consistent with the main analysis.

Table 3.3: Descriptive return statistics

Portfolio	N	Mean (%)	Vol. (%)	Sharpe
ALL	254	17.21	15.20	1.02
Product-based	254	16.31	15.91	0.92
Conduct-based	248	20.66	19.78	0.96
Coal	118	17.35	12.88	1.17
Tobacco and Cannabis	193	14.78	16.02	0.84
Weapons	254	17.66	19.68	0.81
Environmental	236	19.14	30.15	0.58
War and Conflict	212	10.30	23.53	0.37
Human Rights	236	12.10	17.85	0.59
MSCI ACWI	254	9.86	15.38	0.53
Risk-free rate	254	1.68	—	—

Note: Means and volatilities are annualized from monthly returns. Sharpe ratios use the Kenneth French one-month risk-free rate. MSCI ACWI is the benchmark.

Descriptive statistics show that exclusion portfolios had higher raw returns than MSCI ACWI before risk adjustment. The ALL portfolio's annualized average return is 17.21% versus 9.86% for MSCI ACWI, with a higher unconditional Sharpe ratio. These figures do not control for systematic risk or factor exposures, but they motivate the formal tests in Section 5.

3.7 Data limitations

The data and benchmark construction have four main limitations. First, the analysis relies on Refinitiv Datastream coverage, so firms without usable RI and MV data are excluded. Second, some category portfolios are small, especially Human Rights and War and Conflict. Third, the active-return and spanning tests use MSCI ACWI rather than the official GPFG equity benchmark. Fourth, the sector-control tests use U.S. industry portfolios because comparable global industry portfolios are not available at the same level of detail.

These limitations affect interpretation. The evidence is strongest for the ALL portfolio, which has the largest sample and longest time series. Category-level results matter for identifying where return evidence is concentrated, but are more sensitive to small samples, timing, sector exposure, and portfolio composition.

Methodology

The empirical design uses two main tests and one supporting benchmark comparison. Fama-French five-factor regressions test whether excluded portfolios earned abnormal returns after controlling for standard equity risk factors. Mean-variance spanning tests examine whether excluded portfolios are redundant relative to the benchmark proxy. Active returns relative to MSCI ACWI are also reported as a benchmark-relative comparison.

4.1 Calendar-time portfolio construction

Using the portfolios defined in Chapter 3, the analysis is carried out at the portfolio-month level. Portfolio returns are calculated as weighted averages of the simple monthly total returns of firms under active exclusion:

$$r_{p,t} = \sum_{i=1}^{N_{p,t}} w_{i,t} r_{i,t} \quad (4.1)$$

Here, $r_{i,t}$ is the monthly total return of firm i , and $w_{i,t}$ is the firm's weight in portfolio p in month t . The weights sum to one within each portfolio-month.

The main specification uses market-value weights. Each firm is weighted by its market value within the relevant exclusion portfolio:

$$w_{i,t} = \frac{MV_{i,t}}{\sum_{j=1}^{N_{p,t}} MV_{j,t}} \quad (4.2)$$

Value-weighting is appropriate for the GPFG context because larger firms are more relevant for a benchmark-oriented investor. This approach follows Berle et al. (2025), who report value-weighted returns for GPFG-excluded firms. Equal-weighted portfolios are used as a robustness check because they describe the average excluded firm rather than the size-weighted exclusion portfolio.

4.2 Hypotheses

The hypotheses translate the research question into two testable components. First, if excluded firms did not earn returns beyond what standard risk factors predict, the Fama-French five-factor alpha should be zero:

$$H_0 : \alpha_p = 0 \quad (4.3)$$

Rejecting this null, with a positive alpha, means the excluded portfolio earned abnormal returns relative to the five-factor model.

The second hypothesis concerns benchmark redundancy. If broad global equity exposure already captures the excluded portfolio, the spanning restrictions should hold:

$$H_0 : \alpha_{S,p} = 0 \quad \text{and} \quad \beta_{S,p} = 1 \quad (4.4)$$

Rejecting this joint null means the excluded portfolio is not fully captured by the benchmark proxy. In other words, it contains risk, return, or both that are not fully replicated by broad global equity exposure.

4.3 Abnormal-return tests

The regression follows the five-factor asset pricing model of Fama and French (2015). To test the first H_0 , Fama-French five-factor regressions are estimated for each exclusion portfolio. The aim is to assess whether the portfolio earned returns beyond what standard equity risk factors would predict:

$$r_p - r_f = \alpha + \beta MKT + \beta SMB + \beta HML + \beta RMW + \beta CMA + \varepsilon \quad (4.5)$$

The dependent variable is the monthly excess return on portfolio p . The explanatory variables are the developed-market Fama-French five factors: market excess return, size, value, profitability, and investment. The main coefficient of interest is α_p , which captures the portion of portfolio return not explained by these factors. A positive and statistically significant Fama-French five-factor alpha suggests that the excluded portfolio earned abnormal returns relative to the factor model.

Alphas are reported in annualized terms. Monthly alphas are annualized by compounding:

$$\alpha_{\text{annual}} = (1 + \alpha_{\text{monthly}})^{12} - 1 \quad (4.6)$$

Following Berle et al. (2025), Newey-West HAC standard errors are used for the main tables (Newey and West, 1987).

The main regressions cover January 2005 to February 2026. Sub-period results are also reported. The post-2021 window begins in January 2022 because it captures the extension beyond earlier GPFG-specific evidence and the period most relevant to the current policy debate.

4.4 Sector-control regressions

Coal and Weapons require additional sector controls because they are closely linked to specific industries. For Coal, returns may partly reflect energy and commodity-market dynamics. For Weapons, returns may reflect broader aerospace and defense-sector trends. The Fama-French five-factor model controls for common equity risk factors, but it may not fully capture these industry-specific effects.

Additional regressions are therefore estimated for Coal and Weapons:

$$r_p - r_f = \alpha + \beta FF5 + \gamma IndustryExcess + \varepsilon \quad (4.7)$$

The industry controls are the Kenneth French portfolios described in Section 3.5. These regressions test whether the Coal and Weapons alphas persist after accounting for sector returns. Results are interpreted in Chapter 5 and discussed in Chapter 6.

4.5 Supporting benchmark-proxy active returns

As a supporting benchmark comparison, active returns relative to MSCI ACWI are calculated:

$$ActiveReturn_p = r_p - r_{ACWI} \quad (4.8)$$

This comparison shows how each excluded portfolio performed relative to broad global equities. Although not the primary risk-adjusted assessment, it serves as a straightforward benchmark-relative evaluation. This approach aligns with GPFG performance evaluation, in which active return is measured relative to the benchmark established by the Ministry of Finance (Døskeland and Sjuve, 2025).

4.6 Benchmark-proxy spanning tests

The mean-variance spanning test asks whether MSCI ACWI already captures the excluded portfolio's risk-return profile. If the joint null is rejected, the excluded portfolio contains return behaviour not fully reproduced by the benchmark proxy.

I use the framework developed by Huberman and Kandel (1987) and extended by Kan and Zhou (2012). The test is estimated without a risk-free asset, using raw monthly returns. In this setting, spanning requires both $\alpha_{S,p} = 0$ and $\beta_{S,p} = 1$.

I use this version because the GPFG is not an unconstrained investor choosing freely between equities and cash. It is a large benchmark-oriented equity investor. The relevant question is therefore whether excluded firms had risk and return not already captured by broad global equity exposure.

For each excluded portfolio, the following regression is estimated using raw monthly returns:

$$r_p = \alpha + \beta r_{MSCI} + u_p \quad (4.9)$$

Here, r_p is the raw monthly return on excluded portfolio p , and r_b is the raw monthly return on the benchmark proxy.

The source of rejection matters. An alpha-driven rejection indicates a return component not captured by MSCI ACWI, while a beta-driven rejection indicates a different benchmark exposure. Because each excluded portfolio is compared to a single benchmark, the test can reject even when the difference is mainly beta-related. For this reason, alpha-driven and beta-driven rejections are separated in the results. A beta-driven rejection alone does not imply improved performance, but it may still matter for portfolio construction and diversification analysis.

4.7 Interpretation of the empirical design

The tests provide complementary evidence on the research question. Fama-French five-factor regressions are the main test of abnormal returns. Mean-variance spanning tests add a portfolio perspective by examining whether excluded portfolios are redundant relative to the benchmark proxy. Active returns provide a simple benchmark-relative comparison. This structure separates factor-adjusted performance, benchmark non-redundancy, and broad-equity comparison while keeping the main focus on realized excess returns.

Empirical Results

The empirical results are presented in five steps. The chapter first reports the full-sample Fama-French alphas, followed by sub-period evidence and sector-control tests for Coal and Weapons. Next, it presents the supporting active-return comparison and the mean-variance spanning results. The chapter ends with robustness checks on weighting and concentration.

5.1 Fama-French five-factor alphas

Table 5.1 reports the full-sample Fama-French five-factor alphas. The main result is clear. The aggregate ALL portfolio earns a positive and statistically significant alpha in the market-value-weighted specification.

Table 5.1: Full-sample Fama-French five-factor alphas, value-weighted

Portfolio	N	Alpha (%)	t-stat	Adj. R^2
ALL	254	6.82***	3.80	0.727
Product-based	254	5.73***	3.15	0.692
Conduct-based	248	10.75***	2.99	0.430
Coal	118	8.51***	3.19	0.482
Tobacco and Cannabis	193	2.66	0.82	0.472
Weapons	254	6.51***	2.80	0.670
Environmental	236	9.48*	1.68	0.484
War and Conflict	212	1.10	0.24	0.280
Human Rights	236	4.26	1.31	0.185

Note: The table reports annualized alphas from Fama-French five-factor regressions using market-value-weighted exclusion portfolios. Three, two, and one asterisks denote significance at the 1%, 5%, and 10% levels, respectively. Standard errors are Newey-West HAC adjusted with four monthly lags. Equal-weighted results are reported in Section 5.6.1.

The ALL portfolio earns an annualized alpha of 6.82%, with a t-statistic of 3.80. This is the main aggregate finding: excluded firms earned positive abnormal returns after controlling for standard equity risk factors. Both Product-based and Conduct-based exclusions contribute to this result. Product-based exclusions earn an alpha of 5.73%, while Conduct-based exclusions earn a larger alpha of 10.75%. At the criterion level, the aggregate alpha is concentrated in a few categories. Coal and Weapons provide the clearest evidence of positive abnormal returns, with alphas of 8.51% and 6.51%, respectively. Environmental is also positive, but only significant at the 10% level. Tobacco and Cannabis, War and Conflict, and Human Rights do not have statistically significant full-sample alphas.

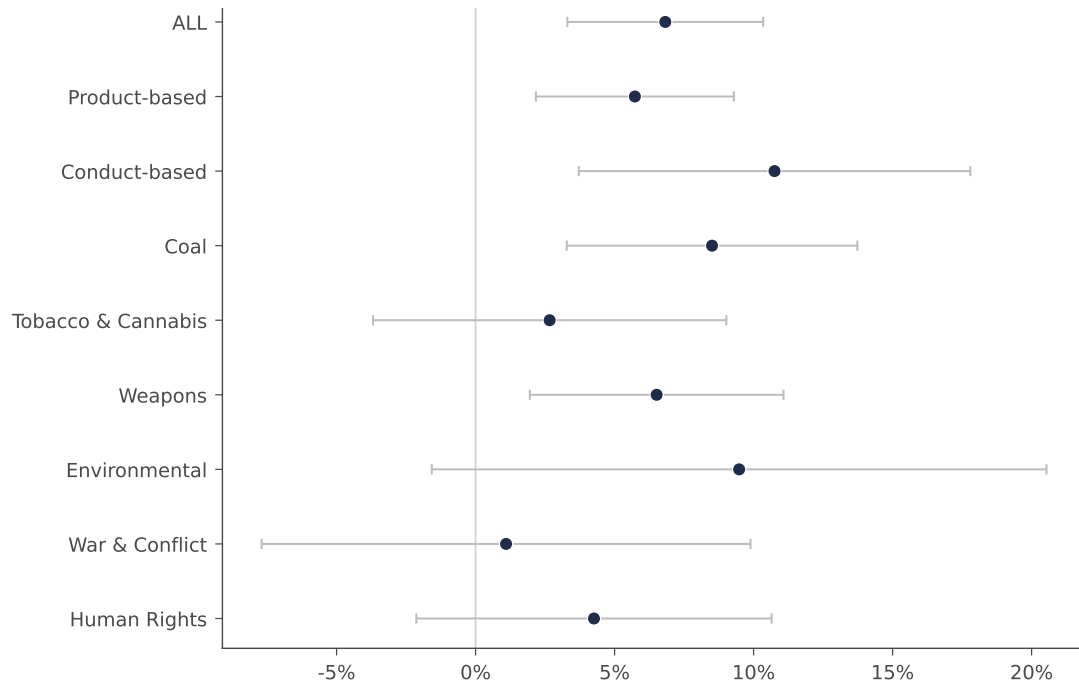


Figure 5.1: Annualized Fama-French five-factor alphas by exclusion criterion. Dots show alpha estimates and horizontal bars show 95% Newey-West HAC confidence intervals.

5.2 Sub-period evidence

Table 5.2 examines whether the abnormal-return evidence is present across sample periods and whether it changes after the end of Berle et al.'s sample period. This matters because the post-2021 period fills the main data gap relative to earlier GPFG-specific evidence and is also relevant to the current policy debate.

Table 5.2: Fama-French alphas across sample windows

Portfolio	Full sample alpha (%)	Pre-2022 alpha (%)	Post-2021 alpha (%)
ALL	6.82*** (3.80)	5.86*** (2.97)	9.60*** (2.85)
Product-based	5.73*** (3.15)	5.03** (2.48)	8.78*** (2.81)
Conduct-based	10.75*** (2.99)	10.02** (2.45)	10.03 (1.56)
Coal	8.51*** (3.19)	9.01** (2.48)	9.85** (2.07)
Weapons	6.51*** (2.80)	7.08** (2.21)	10.37** (2.42)
Tobacco and Cannabis	2.66 (0.82)	2.57 (0.72)	4.58 (0.77)
Environmental	9.48* (1.68)	9.52 (1.46)	9.66 (1.18)
War and Conflict	1.10 (0.24)	3.42 (0.68)	-3.70 (-0.44)
Human Rights	4.26 (1.31)	5.17 (1.36)	3.14 (0.36)

Note: The table reports annualized Fama-French five-factor alphas. T-statistics are reported in parentheses. Three, two, and one asterisks denote significance at the 1%, 5%, and 10% levels. Standard errors are Newey-West HAC adjusted.

The aggregate result appears in both sub-periods. The ALL portfolio earns a positive and statistically significant alpha before 2022, and a larger alpha after 2021. This suggests the main result

is not only driven by the earlier part of the sample. At the criterion level, Coal and Weapons remain the clearest return categories. War and Conflict remains weak and is negative after 2021.

5.3 Sector-control results for Coal and Weapons

Coal and Weapons are the strongest criterion-level results, but they are also the categories where sector exposure is most likely to matter. The sector-control regressions therefore test whether the alphas remain after accounting for coal-industry and defense-related returns. Table 5.3 reports the results.

Table 5.3: Sector-control regressions for Coal and Weapons

Panel A: Coal

Model	Alpha (%)	t-stat	Industry beta	Adj. R^2
Baseline FF5	8.51***	3.19	—	0.482
FF5 + Coal industry	7.78***	3.42	0.057**	0.505

Panel B: Weapons

Model	Alpha (%)	t-stat	Industry beta	Adj. R^2
Baseline FF5	6.51***	2.80	—	0.670
FF5 + Aerospace	4.48***	3.40	0.659***	0.860
FF5 + Guns	2.75	1.41	0.382***	0.772

Note: The table reports annualized alphas from Fama–French five-factor regressions with additional Fama–French industry controls. Standard errors are Newey–West HAC adjusted with four monthly lags. Three, two, and one asterisks denote significance at the 1%, 5%, and 10% levels.

The Coal alpha remains positive and statistically significant after adding the coal-industry proxy. The full-sample alpha falls only modestly, from 8.51% to 7.78%. This suggests that the Coal result is not primarily absorbed by the sector-control specification, although sector exposure remains relevant for interpretation.

The Weapons result is more sensitive. The alpha remains positive and statistically significant when Aerospace is used as the control, but becomes smaller and statistically insignificant when Guns is used. This suggests that part of the Weapons result reflects defense-sector exposure. The evidence supports a positive return result for Weapons, but not a weapons-specific exclusion premium independent of sector exposure.

5.4 Supporting benchmark-relative active returns

Table 5.4 reports active returns relative to MSCI ACWI. This gives a simple benchmark-relative comparison between the excluded portfolios and broad global equities.

Table 5.4: Active returns relative to MSCI ACWI

Portfolio	N	Mean active return (%)	t-stat
ALL	254	7.61	3.61
Product-based	254	6.65	3.10
Conduct-based	248	11.10	3.03
Coal	118	4.34	1.14
Tobacco and Cannabis	193	3.25	0.87
Weapons	254	8.09	3.15
Environmental	236	9.85	1.59
War and Conflict	212	-1.52	-0.34
Human Rights	236	2.42	0.60

Note: Active return is defined as the exclusion-portfolio return minus MSCI ACWI. Mean active return is the average active return over the sample, annualized. The t-statistic tests whether this average active return is different from zero and is Newey-West HAC adjusted.

The benchmark-relative results are positive for the aggregate portfolio and for both broad exclusion groups. The ALL portfolio has an annualized mean active return of 7.61%, meaning that it outperformed MSCI ACWI by 7.61 percentage points per year on average with a statistically significant t-stat. The broad split shows the same pattern: Product-based exclusions have a positive active return of 6.65%, while Conduct-based exclusions have a larger active return of 11.10%. Both are statistically significant. At the criterion level, Weapons is the clearest positive result, with a significant active return of 8.09%. Coal, Tobacco and Cannabis, Environmental, and Human Rights also have positive active returns, but their t-stats are lower, so the evidence is weaker. War and Conflict is the only category with a negative active return.

5.5 Mean-variance spanning results

Table 5.5 reports the mean-variance spanning results. The joint p-value tests whether each excluded portfolio is redundant relative to MSCI ACWI. The channel column shows whether the rejection is driven by the return component, benchmark exposure, or both.

Table 5.5: Mean-variance spanning tests, market-value-weighted specification

Portfolio	N	Alpha (%)	Beta	Joint p-value	Channel
ALL	254	9.56***	0.817	< 0.001***	Both
Product-based	254	8.44***	0.830	< 0.001***	Both
Conduct-based	248	13.09***	0.823	< 0.001***	Both
Coal	118	10.22***	0.579	< 0.001***	Both
Tobacco and Cannabis	193	7.46**	0.653	< 0.001***	Both
Weapons	254	8.06***	1.003	0.007***	Alpha-only
Environmental	236	6.65	1.307	< 0.001***	Beta-only, high beta
War and Conflict	212	-0.28	0.894	0.473	No rejection
Human Rights	236	8.03**	0.449	< 0.001***	Both

Note: The table reports spanning regressions of each exclusion portfolio's raw monthly return on the raw monthly return of MSCI ACWI. Alpha is annualized. The joint p-value tests $\alpha = 0$ and $\beta = 1$.

The spanning null is rejected for the ALL portfolio. This is driven by both a positive return component and beta below one, meaning the aggregate exclusion portfolio is not redundant relative to MSCI ACWI.

Product-based and Conduct-based exclusions show the same pattern. Both groups differ from MSCI ACWI in return component and benchmark exposure.

At the criterion level, the results are more mixed. Coal, Tobacco and Cannabis, and Human Rights also reject through both channels, so they are not fully captured by MSCI ACWI. Weapons is mainly a return channel result: its beta is close to one, while its alpha is positive as well as statistically significant. Environmental is different: its rejection is driven by beta above one, not by significant alpha. In simple terms, the portfolio behaved like a more market sensitive version of our benchmark. This means that the spanning rejection reflects higher exposure to global equity movements. War and Conflict is the only portfolio for which the joint null is not rejected, meaning that MSCI ACWI captures its risk-return profile in this test.

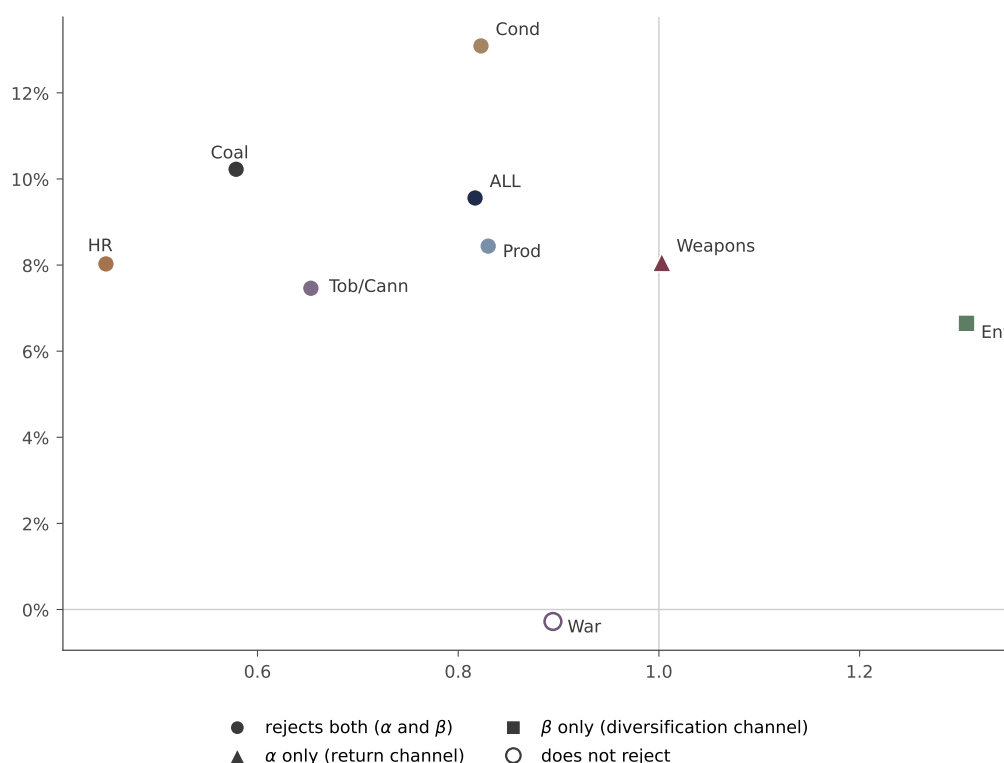


Figure 5.2: Mean-variance spanning interpretation. The figure plots annualized spanning alphas against beta on MSCI ACWI. The vertical line marks $\beta = 1$, and the horizontal line marks $\alpha = 0$. Marker types indicate whether the joint spanning test rejects and through which channel.

5.6 Robustness checks

The robustness checks examine whether the main results depend on two modelling choices: value weighting and portfolio concentration. These checks are especially important because the main specification gives more weight to larger excluded firms. Value weighting is relevant for the GPFG setting, but it also means that a few large companies can have substantial influence on the results.

5.6.1 Equal-weighted results

The first robustness check gives each excluded firm the same weight. This shows whether the results are also present for the average excluded firm, rather than mainly reflecting the largest firms in the value-weighted portfolios.

The equal-weighted results are weaker than the value-weighted results. Only ALL, Product-based, and Weapons remain statistically significant, while Conduct-based, Coal, Tobacco and Cannabis, Environmental, War and Conflict, and Human Rights are not significant. This shows that the main result is stronger when larger excluded firms receive more weight. For the GPFG, larger companies are more relevant for benchmark exposure and for the scale of an investment restriction. The distinction also shapes the interpretation: the evidence is not a uniform premium across the average excluded firm, but is stronger among the larger firms in the excluded universe.

Table 5.6: Equal-weighted Fama-French five-factor alphas

Portfolio	N	Alpha (%)	t-stat	Adj. R^2
ALL	254	4.54**	2.22	0.750
Product-based	254	4.13**	2.09	0.684
Conduct-based	248	6.64	1.63	0.533
Coal	118	4.95	1.48	0.489
Tobacco and Cannabis	193	0.73	0.26	0.492
Weapons	254	5.07**	2.35	0.727
Environmental	236	5.27	0.91	0.478
War and Conflict	212	4.34	1.12	0.410
Human Rights	236	4.66	1.28	0.285

Note: The table reports annualized alphas from Fama-French five-factor regressions using equal-weighted exclusion portfolios. Standard errors are Newey-West HAC adjusted.

5.6.2 Concentration and top-name exclusion

The second robustness check examines whether the value-weighted results are driven by a few large firms, since value weighting gives more influence to larger companies. Table 5.7 reports concentration measures for the main contributors to the empirical results: the aggregate portfolio, the product/conduct split, and the two strongest criterion-level categories.

Table 5.7: Portfolio concentration measures

Portfolio	Avg. active firms	Top-1 weight	Top-5 weight	Top-10 weight
ALL	79.3	19.0%	44.4%	62.3%
Product-based	54.5	18.1%	49.7%	71.6%
Conduct-based	25.4	41.2%	73.1%	88.2%
Coal	60.0	10.2%	35.7%	55.3%
Weapons	14.7	25.0%	72.6%	94.8%

Note: Weights are within-portfolio value shares by month, averaged across sample months.

The concentration measures show that the ALL portfolio is relatively broad, while Conduct-based and Weapons are more concentrated. Conduct-based has the highest top-name exposure, with the largest firm accounting for 41.2% of the portfolio on average and the top five firms accounting for 73.1%. Weapons is also concentrated, with the top five firms accounting for 72.6%. Because both portfolios have strong value-weighted results, the robustness question is whether the alphas depend too much on the largest names.

Table 5.8 reports the corresponding top-name exclusion tests. Conduct-based is included because it is the most concentrated broad portfolio and has a large value-weighted alpha. It is therefore useful to test whether the result is driven by one or a few large firms.

Table 5.8 shows that the main value-weighted results are not driven by a single large firm. The ALL and Product-based alphas remain positive and statistically significant after removing the

Table 5.8: Top-name exclusion robustness, value-weighted specification

Portfolio	Baseline alpha (%)	Excl. top 1 alpha (%)	Excl. top 5 alpha (%)	Excl. top 10 alpha (%)
ALL	6.82*** (3.80)	6.49*** (3.63)	6.45*** (3.05)	7.30*** (3.64)
Product-based	5.73*** (3.15)	5.67*** (3.01)	6.15*** (3.57)	6.89*** (3.68)
Conduct-based	10.75*** (2.99)	8.81*** (2.60)	10.44** (2.37)	8.59 (1.27)
Coal	8.51*** (3.19)	7.87*** (3.10)	6.70** (2.42)	7.44*** (2.60)
Weapons	6.51*** (2.80)	6.94*** (3.31)	7.36*** (2.92)	5.73** (2.02)

Note: The table reports annualized FF5 alphas. T-statistics are in parentheses. Standard errors use Newey-West HAC. Top firms are ranked by average value weight.

largest firm, the top five firms, and the top ten firms. Conduct-based also remains significant after removing the largest firm and the top five firms, but not after the top ten. This suggests the Conduct-based result is not driven by a single company, but does depend partly on the largest names. Coal remains statistically significant across all top-name exclusions, and Weapons is not weakened by removing its largest firms.

5.6.3 Newey-West lag test

The main regressions use Newey-West standard errors with four lags, which is common for monthly return data and helps account for patterns over time. To see if this choice matters, I also run the same tests with one lag. This change does not affect the main estimates, but it does alter the standard errors, t-statistics, and p-values. As shown in Tables 5.9 and 5.10, the key results are unchanged: the exclusion portfolio stays positive and significant, Coal and Weapons are still among the strongest categories, and the spanning tests give the same overall interpretation. The results therefore do not depend on using four lags.

Table 5.9: Fama-French five-factor alphas: Newey-West lag robustness

Portfolio	N	Alpha ann. (%)	<i>t</i> NW(4)	<i>t</i> NW(1)	Adj. R^2
ALL	254	6.82	3.80	3.44	0.727
Product-based	254	5.73	3.15	2.80	0.692
Conduct-based	248	10.75	2.99	2.88	0.430
Coal	118	8.51	3.19	2.79	0.482
Tobacco and Cannabis	193	2.66	0.82	0.84	0.472
Weapons	254	6.51	2.80	2.55	0.670
Environmental	236	9.48	1.68	1.73	0.484
War and Conflict	212	1.10	0.24	0.22	0.280
Human Rights	236	4.26	1.31	1.14	0.185

Notes: The table reports annualized value-weighted Fama-French five-factor alphas. The coefficient estimates are unchanged across the two specifications; only the Newey-West lag length changes.

Table 5.10: Mean-variance spanning tests: Newey-West lag robustness

Portfolio	N	Alpha ann. (%)	Beta	<i>p</i> NW(4)	<i>p</i> NW(1)
ALL	254	9.56	0.817	0.000	0.000
Product-based	254	8.44	0.830	0.000	0.000
Conduct-based	248	13.09	0.823	0.000	0.000
Coal	118	10.22	0.579	0.000	0.000
Tobacco and Cannabis	193	7.46	0.653	0.000	0.000
Weapons	254	8.06	1.003	0.007	0.010
Environmental	236	6.65	1.307	0.000	0.000
War and Conflict	212	-0.28	0.894	0.473	0.478
Human Rights	236	8.03	0.449	0.000	0.000

Notes: The table reports value-weighted spanning tests relative to MSCI ACWI. The p-values refer to the joint spanning test. The coefficient estimates are unchanged across the two specifications; only the Newey-West lag length changes.

5.7 Summary of empirical findings

The empirical results show a clear but uneven pattern. The aggregate exclusion portfolio earns positive realized excess returns, and the result remains strong after controlling for standard equity risk factors. The evidence is not evenly distributed across criteria. Coal and Weapons provide the clearest return evidence, though the Weapons result is more sensitive to the defense-sector control. The active-return comparison also shows positive benchmark-relative performance for the aggregate portfolio, while the spanning tests show that several excluded portfolios are not redundant relative to MSCI ACWI. The robustness checks show that the results are weaker under equal weighting, but the value-weighted aggregate result is not driven by a single large firm. The main inference is not materially affected by the Newey–West HAC lag choice. Overall, excluded firms earned measurable realized excess returns, but interpretation must be criterion-specific.

Discussion

The results point to a clear but uneven financial pattern. Excluded firms earned measurable realized excess returns in the aggregate, but the evidence is concentrated in specific criteria, especially Coal and Weapons. This chapter interprets the financial meaning of the findings, discusses possible mechanisms, and explains their relevance for the GPFG ethical-framework review.

6.1 Interpreting the main finding

The aggregate result shows that excluded firms earned measurable realized excess returns while they were outside the GPFG's investment universe. For an investor constrained from holding these firms, this can be interpreted as a negative ethical premium: firms avoided by some investors may need to offer higher returns to attract capital. The result is strongest in the value-weighted specification, which is the most relevant for a large benchmark-oriented fund.

The distinction between value-weighted and equal-weighted returns is important. Equal-weighted estimates describe the average excluded firm, while value-weighted estimates give more influence to larger excluded firms. Since the results are stronger in the value-weighted specification, the evidence is not a uniform premium across all excluded firms. It is stronger among larger firms in the excluded universe. This does not weaken the relevance for the GPFG, since large companies matter more for benchmark-relative exposure and the scale of an investment restriction, but it does shape the interpretation of the mechanism.

The result is not an exact estimate of the GPFG's fund-level opportunity cost. The true effect depends on what the fund holds instead, how exclusions are reflected in the benchmark, actual ownership weights, implementation timing, and currency reporting. The estimates in this thesis should therefore be interpreted as evidence on excluded-firm returns and benchmark-proxy return differences.

6.2 Mechanisms

The positive post-exclusion returns can be explained through three possible mechanisms.

The first mechanism is a persistent exclusion or sin-stock premium. If norm-constrained investors avoid certain firms, these firms may trade at lower valuations and earn higher expected returns. This is the mechanism described by Hong and Kacperczyk (2009), and it fits the broad logic of this thesis: firms outside the investable universe of responsible investors may earn a return premium.

The second mechanism is recovery from controversy or exclusion-related overreaction. Council on Ethics recommendations often follow public controversies, meaning that negative information may already be reflected in prices by the time a firm enters the exclusion portfolio. If markets initially react strongly to the controversy, some of the later return may reflect

recovery rather than a permanent premium.

The third mechanism is sector tailwinds unrelated to exclusion. This is especially relevant for Coal and Weapons. Coal firms are exposed to energy and commodity markets, while weapons-related firms are exposed to aerospace and defense demand. Strong returns in these sectors can raise excluded-firm returns even if the driver is sector performance rather than exclusion status itself.

These mechanisms are best understood as complementary explanations. They can operate at the same time, and their importance can differ across exclusion criteria. The empirical design cannot fully separate them: the estimated alphas show how excluded firms performed after exclusion, but they do not identify whether the returns were caused by investor avoidance, recovery from earlier controversy, or sector-specific shocks. Coal and Weapons are more exposed to sector-specific explanations, while the broader aggregate result is also consistent with the exclusion-premium logic. This is why the next section interprets the results criterion by criterion.

6.3 Criterion-level interpretation

The aggregate result masks important differences across the exclusion criteria. The same return estimate does not have the same economic meaning across all types of exclusions. A coal result, a weapons result, and a human-rights result may all enter the aggregate portfolio, but they represent very different firms, risks, sectors, and moral concerns. The criterion-level split is therefore key to understanding what drives the aggregate finding.

Coal has the strongest criterion-level return evidence. The portfolio earns a positive and statistically significant factor alpha, and the result remains positive after adding the coal-industry control. Economically, this means that the Coal exclusions removed firms that performed strongly in listed equity markets after exclusion. The result is not fully explained by the Kenneth French coal-mining industry proxy. At the same time, Coal is closely linked to energy and commodity markets, and the post-2021 period includes large energy-market movements. The Coal result should therefore be read as a strong listed-equity return result with an important sector component, not as a pure exclusion effect.

Coal also has a special interpretation for the GPFG. The exclusion removed firms that performed strongly after exclusion, but coal is not just another sector for a petroleum-funded sovereign wealth fund. Norway's broader public wealth is already exposed to petroleum revenues and energy-transition risk. From a listed-equity perspective, the Coal exclusion was associated with foregone exposure to high realized returns. From a total-wealth perspective, however, holding more coal-related equity exposure may not have reduced Norway's overall energy risk. The Coal alpha therefore measures the listed-equity return earned by excluded coal firms in the sample, not the full total-wealth effect for the Norwegian state.

Weapons also show positive abnormal returns, though the interpretation is different. The alpha continues to be significant with the Aerospace control, but is largely absorbed by the Guns

proxy. This suggests that part of the Weapons return result is tied to defense sector performance. The spanning result points in the same direction: Weapons is not redundant relative to MSCI ACWI, mainly because of its return component, while its beta is close to one. Economically, Weapons appear to be a category where financial exclusion mattered, but the return evidence is difficult to separate from broader defense-sector conditions.

Tobacco and Cannabis are more mixed. The category does not have a statistically significant Fama-French alpha, but it is not redundant relative to MSCI ACWI in the value-weighted spanning test. This means that the category does not provide strong evidence of factor-adjusted abnormal returns, but it still differs from the broad equity benchmark in risk and return. The grouping also matters. Tobacco fits the classic sin-stock literature more directly, while cannabis is a newer and more speculative industry. Combining them is necessary for sample-size reasons, but it may weaken a cleaner tobacco-specific signal.

Environmental points in another direction. The factor alpha is positive, although only weakly significant, while the spanning rejection is driven mainly by beta above one. Economically, this means that Environmental differs from MSCI ACWI mostly through a higher benchmark exposure, not through a clear return component. Its result should therefore not be interpreted as a low-risk diversification benefit. It is better understood as a category with a higher systematic exposure in the spanning regression.

Conduct-based exclusions are harder to summarize because they are not tied to one stable sector. This portfolio has a high aggregate alpha and is also not redundant relative to MSCI ACWI. However, the concentration tests show that large names matter. The result remains significant after dropping the top one and top five firms, but becomes less precise after dropping the ten largest firms. The implication is that Conduct-based exclusions are financially important in the value-weighted results, but the result is not a broad premium shared evenly across all conduct-related exclusions.

War and Conflict is the weakest category in the quantitative results. It does not earn a significant full-sample factor alpha and does not reject the spanning null. In this test, MSCI ACWI captures its risk-return profile. This does not imply the category is ethically less important, only that, historically, War and Conflict has not been the main source of measurable listed-equity excess returns in this sample.

Human Rights show a significant spanning result, but the category is small and the factor alpha is not statistically significant. This means that the portfolio has risk and return not entirely captured by the benchmark proxy, but the evidence is not strong enough to support a clear factor-adjusted return premium. The small number of firms makes precise criterion interpretation difficult.

Taken together, the criterion-level results show that exclusion effects are not uniform. Coal and Weapons yield the clearest return evidence, but they also require sector-specific interpretation. Tobacco and Cannabis, Environmental, and Human Rights matter more through the spanning results than through clear factor-adjusted alphas. War and Conflict does not contribute

materially to the measured excess-return results. The financial consequences of the ethical framework should therefore be evaluated criterion by criterion, not only through the aggregate exclusion portfolio.

6.4 Return effects versus benchmark spanning effects

The Fama-French regressions and the spanning tests address related but distinct aspects of the research question. The Fama-French regressions examine whether excluded portfolios earned abnormal returns after controlling for standard equity risk factors. The spanning tests assess whether these return patterns were already captured by broad global equity exposure. An excluded portfolio can earn abnormal returns in the factor model, but still be less relevant for a benchmark-oriented investor if its risk-return profile closely matches MSCI ACWI.

The two tests are therefore interpreted together. A positive Fama-French alpha indicates abnormal performance relative to standard risk factors. A spanning rejection shows that the excluded portfolio is not redundant relative to the benchmark proxy. When both are present, the financial evidence is stronger: the portfolio earned abnormal returns and was not fully captured by broad global equities. When only one is present, interpretation is more cautious.

This distinction is important at the criterion level. Weapons has both a positive Fama-French alpha and a spanning rejection driven mainly by returns, which strengthens the financial interpretation: excluded Weapons firms not only performed well after exclusion, but also offered return behaviour not simply replicated by MSCI ACWI. Environmental does not show a clear abnormal-return result, but mainly moves more strongly with MSCI ACWI. War and Conflict shows neither a significant Fama-French alpha nor a spanning rejection. Relative to the research question, this means War and Conflict firms do not provide evidence of measurable realized excess returns after exclusion, either through abnormal returns or through risk and return not already captured by the benchmark proxy. For the aggregate ALL portfolio, both tests point in the same direction: excluded firms earned positive abnormal returns and were not fully captured by the benchmark proxy.

6.5 Relation to previous literature

The results are closest to Berle et al. (2025), who estimate a positive return premium for firms excluded by the GPFG. This thesis extends their approach by updating the sample to February 2026, decomposing the exclusion universe by criterion, adding benchmark-proxy spanning tests, and including supporting active-return comparisons.

The results differ from Hoepner and Schopohl (2018), who find limited evidence that exclusions harmed performance. This difference should not be read as a direct contradiction. The studies differ in sample period, exclusion universe, benchmark choices, and market environment. The later sample includes the coal criterion and the post-2021 period, when energy and defense-related firms performed strongly.

The broader literature on negative screening also supports a heterogeneous interpretation.

Hong and Kacperczyk (2009) show that social norms can affect expected returns. Trinks and Scholtens (2017) show that negative-screening costs differ across controversial issues. Blitz and Swinkels (2023) emphasize that exclusion costs depend on what replaces the excluded firms. The results here are consistent with that literature, but do not indicate a universal exclusion premium.

6.6 Policy relevance for the ethical-framework review

The current review of the ethical framework is broader than the financial analysis presented here. The Gjedrem committee must consider ethical obligations, legal constraints, ownership engagement, the role of the Council on Ethics, legitimacy, foreign-policy concerns, and operational feasibility. This thesis does not address those questions. Its contribution is narrower: it estimates one financial dimension of the trade-off.

That contribution remains useful because the return-risk profiles differ across criteria. The GPFG is a broad, global, benchmark-oriented fund. If exclusions remove firms from the benchmark-relevant opportunity set, their financial effects should be measured and made transparent.

A simple back-of-the-envelope calculation illustrates the scale of the result. This is not a separate empirical test, but a way to translate the estimated alpha into an approximate dollar value. Appendix A4 shows the calculation using the active excluded market value and an illustrative GPFG ownership assumption of 1.5%. At the end of 2025, the active excluded universe had a listed market value of about USD 4.2 trillion. Under the 1.5% ownership assumption, this corresponds to about USD 62 billion in equity exposure. Applying the aggregate annualized alpha of 6.82% gives an illustrative return difference of about USD 4.3 billion per year. This should not be read as the exact opportunity cost to the GPFG, since the true effect depends on benchmark weights, ownership shares, implementation, currency, and the assets held instead. The purpose is only to show that the estimated return differences are economically meaningful for a fund of the GPFG's size.

The policy implication is not that exclusions should be avoided. A criterion can be ethically justified even if excluded firms later earn high returns. A criterion can also have little measured financial cost without being ethically more important. A financial cost in one criterion does not have the same policy meaning as the same cost in another criterion. The ethical issue involved in holding a tobacco or coal company is different from the ethical issue involved in holding a company linked to serious human-rights violations or war and conflict. The point is that financial consequences should be explicit, criterion-specific, and interpreted alongside ethical, legal, and operational considerations.

Exclusion is not the only tool available to a responsible investor. Observation, ownership dialogue, voting, risk-based divestment, and formal exclusion serve different purposes. This thesis does not test whether ownership engagement could achieve similar ethical objectives at lower financial cost. It only shows that the financial consequences of exclusions can be measured and that they differ across criteria.

6.7 Limitations and alternative explanations

The analysis uses realized return data, showing how excluded firms performed after exclusion, but not fully explaining why they performed that way. Positive realized excess returns may reflect investor neglect, compensation for unmodelled risks, recovery from controversy, sector timing, geopolitical shocks, litigation risk, or transition risk.

Benchmark choice is another limitation. MSCI ACWI serves as a public proxy for broad global equity exposure and is used in related GPFG-specific work, but it is not the official GPFG benchmark. The official equity benchmark is FTSE-based, and a historical version was not accessible for this thesis. The spanning results should therefore be interpreted as benchmark-proxy evidence, not as exact estimates against the official GPFG equity benchmark.

The exclusion-window rule is another limitation. Firms are included from the month after the public exclusion announcement to give a clean post-exclusion return window. However, the Council on Ethics notes that Norges Bank needs time to assess recommendations and divest the GPFG's shareholdings before decisions are published (Council on Ethics, 2024). Any price effects during this divestment period, or during the controversy period before exclusion, are therefore outside the return window. The analysis may not capture the full return effect around the exclusion process, but focuses on how excluded firms performed after they were formally outside the GPFG's investment universe.

The Fama-French factors are developed-market factors. Some excluded firms are listed outside developed markets, and some categories have sector-specific exposures that may not be fully captured. The sector-control tests help address this for Coal and Weapons, but do not resolve all benchmark and factor-matching issues. Some excluded firms, especially emerging-market firms and commodity-exposed coal companies, may have risk exposures not fully captured by the factor model.

Some category portfolios are small or have short histories. Coal starts in 2016, the post-2021 period contains only 50 monthly observations, and Human Rights and War and Conflict contain relatively few firms. Category-level estimates are therefore less precise than the aggregate ALL result.

There is also a possible selection issue in the active-window construction. Firms leave the return sample when they no longer have usable listed-equity data, for example after delisting, bankruptcy, or corporate restructuring. If firms with poor subsequent performance are more likely to disappear from the listed-equity data, the estimated returns may be biased upward. This does not overturn the main evidence, but means that the results should be interpreted as returns on the listed excluded firms that remain observable in the data.

Finally, the analysis focuses only on listed-equity performance. It does not measure reputational benefits, reduced legal risk, political legitimacy, ownership effectiveness, ethical value, or wider social consequences. These dimensions are central to the policy debate, but they are outside the empirical scope of this thesis.

Overall, ethical exclusions have been associated with measurable realized excess returns

in some criteria. The financial side of the trade-off should be explicit, criterion-specific, and interpreted with appropriate caution.

Conclusion

This thesis examines whether firms excluded from the GPFG have earned measurable realized excess returns in listed equity markets. The analysis employs Fama-French five-factor regressions and mean-variance spanning tests against MSCI ACWI. Benchmark-proxy active returns relative to MSCI ACWI are also reported for comparison.

The results show that excluded firms earned positive and statistically significant realized excess returns in aggregate, with an annualized abnormal return of 6.82%. The clearest criterion evidence is found in Coal and Weapons. Coal continues to be positive and statistically significant after controlling for the Kenneth French Coal-mining industry portfolio, while the Weapons result is more sensitive to sector definition: the alpha survives an Aerospace control, but is largely absorbed by the Guns proxy.

Several excluded portfolios are also not redundant relative to MSCI ACWI. These spanning results are important, but they should not be read as proof that holding the excluded firms would have automatically improved the GPFG portfolio. In several cases, the rejection reflects both return differences and different benchmark exposure.

The value-weighted results are especially relevant for the GPFG because larger companies matter more for benchmark-relative exposure. Equal-weighted results are generally weaker, showing that larger excluded firms contribute materially to the headline result. The estimates are not the exact fund-level effect. That effect depends on benchmark treatment, ownership weights, implementation, and the assets held instead.

The policy implication is not that ethical exclusions should be avoided. A criterion can be ethically justified even if excluded firms later earn high returns. A criterion can also have little measured return effect without being ethically more important. The contribution is therefore more limited but more precise: the thesis does not simply show that excluded firms outperformed, but that the realized financial effects of ethical exclusions are criterion-specific and benchmark-dependent.

For the current review of the ethical framework, the main implication is transparency. If exclusions restrict the GPFG's opportunity set, their financial effects should be monitored and reported alongside the ethical, legal, and operational reasons for exclusion. This would make the trade-off clearer without reducing the ethical framework to a return-maximization exercise.

Overall, the thesis shows that excluded firms have earned measurable realized excess returns over the sample period. This does not determine where the ethical boundary should be drawn, but it makes one financial part of the trade-off visible.

Appendix

A1 Removed firms

Table A1 lists the 31 firms removed before the final 205-firm sample is formed.

Table A1: Removed firms by reason

Firm	Step	Reason
Berkshire Hathaway Energy Co	Observation	Observation only
Bombardier Inc	Observation	Observation only
CMS Energy Corp	Observation	Observation only
Fincantieri SpA	Observation	Observation only
KDDI Corp	Observation	Observation only
Kyushu Electric Power Co Inc	Observation	Observation only
Marfrig Global Foods SA	Observation	Observation only
MidAmerican Energy Co	Observation	Observation only
NorthWestern Corp	Observation	Observation only
OGE Energy Corp	Observation	Observation only
Pan Ocean Co Ltd	Observation	Observation only
Petrofac Ltd	Observation	Observation only
Pinnacle West Capital Corp	Observation	Observation only
Semen Indonesia Persero Tbk PT	Observation	Observation only
Sumitomo Corp	Observation	Observation only
Tohoku Electric Power Co Inc	Observation	Observation only
Toronto-Dominion Bank	Observation	Observation only
Uniper SE	Observation	Observation only
Vistra Corp	Observation	Observation only
Adani Green Energy	Post-window	Entry after sample end
Bolloré SE	Post-window	Entry after sample end
Compagnie de l'Odé SE	Post-window	Entry after sample end
Dian Swastatika Sentosa Tbk PT	Post-window	Entry after sample end
Ecopetrol SA	Post-window	Entry after sample end
Great River Energy	No listed equity	No exchange-listed equity
PacifiCorp	No listed equity	Not separately listed
Petróleos Mexicanos	No listed equity	State-owned issuer
Tri-State Generation & Transmission Assoc.	No listed equity	No exchange-listed equity
Empire District Electric Co	Unusable RI	Unusable RI series
Evraz PLC	Unusable RI	Suspended / non-tradeable RI
Korea Line Corp	Unusable RI	Unusable RI series

A2 Portfolio classification

Table A2 summarizes how official exclusion reasons are grouped into the empirical criterion portfolios.

Table A2: Portfolio classification mapping

Portfolio	Main official reasons included
Coal	Coal and coal-based energy.
Tobacco/Cannabis	Tobacco and cannabis.
Weapons	Weapons production, including nuclear weapons, cluster munitions, and antipersonnel landmines. Also includes selected conduct-based weapons-sales criteria.
Environmental	Severe environmental damage, unacceptable greenhouse gas emissions, and related environmental criteria.
War / Conflict	Serious violations in situations of war or conflict, including selected cases related to Western Sahara.
Human Rights	Human-rights violations, serious or systematic human-rights violations, and gross corruption.

A3 Special data cases

Table A3 summarizes the main firm-level data cases requiring manual treatment.

Table A3: Special data cases

Firm	Treatment
Peabody Energy Corp	Chapter 11 months with missing RI/MV are left missing.
Poongsan Corp	Early Refinitiv RI gaps are left missing.
Potash Corp / Nutrien Ltd	The two Datastream series are spliced; the transition month is dropped.
Vector Group Ltd	The active window is capped at the last clean RI month.
ALLETE Inc	The active window is capped at the last clean RI month.
RWE AG	Stale RI segments are removed by splitting/capping the active window.
Korea Line Corp	Dropped because the RI series is unusable.
Empire District Electric Co	Dropped because the RI series is unusable.
Evrz PLC	Dropped because the post-suspension RI is not tradeable.
Danya Cebus Ltd	Old and new listing series are stitched into one firm history.

Notes: Missing returns are not imputed. Portfolios are rebuilt each month using firms with valid returns and active exclusion status.

A4 Back-of-the-envelope calculation

Table A4 shows the illustrative scale calculation used in Section 6.6.

Table A4: Back-of-the-envelope calculation

Input	Value	Calculation
Active excluded listed market value	≈ USD 4.2 trillion	December 2025 cross-section
Illustrative GPFG ownership share	1.5%	Approximate ownership assumption
Implied excluded equity exposure	≈ USD 62 billion	$\text{USD 4.2 trillion} \times 1.5\%$
Aggregate FF5 alpha, ALL VW	6.82%	Headline alpha estimate
Illustrative annual return difference	≈ USD 4.3 billion	$\text{USD 62 billion} \times 6.82\%$

Notes: This is only an illustrative scale calculation. It is not an exact estimate of the GPFG's opportunity cost.

A5 Full Fama–French five-factor estimates, value-weighted portfolios

Table A5: Full Fama–French five-factor regression estimates, value-weighted portfolios

Portfolio	Alpha	MKT	SMB	HML	RMW	CMA	Adj. R2	N
ALL	6.82 (3.80)	0.864 (22.26)	0.017 (0.14)	0.334 (3.45)	0.416 (3.00)	0.280 (1.76)	0.727	254
Product-based	5.73 (3.15)	0.896 (23.45)	0.032 (0.27)	0.263 (2.66)	0.370 (2.46)	0.375 (2.54)	0.692	254
Conduct-based	10.75 (2.99)	0.841 (11.94)	0.004 (0.02)	0.449 (2.73)	0.412 (1.60)	0.178 (0.73)	0.430	248
Coal	8.51 (3.19)	0.618 (10.49)	0.196 (1.14)	0.089 (0.52)	0.298 (1.19)	0.548 (1.98)	0.482	118
Tobacco & Cannabis	2.66 (0.82)	0.765 (15.84)	-0.069 (-0.40)	0.084 (0.50)	0.686 (2.89)	1.035 (4.62)	0.472	193
Weapons	6.51 (2.80)	1.068 (18.30)	0.066 (0.46)	0.292 (2.00)	0.133 (0.77)	0.354 (2.18)	0.670	254
Environmental	9.48 (1.68)	1.077 (9.09)	0.443 (1.85)	1.158 (3.09)	0.494 (1.33)	-1.326 (-1.81)	0.484	236
War & Conflict	1.10 (0.24)	0.811 (8.51)	0.060 (0.19)	0.546 (2.10)	-0.240 (-0.65)	-0.281 (-0.76)	0.280	212
Human Rights	4.26 (1.31)	0.559 (7.02)	-0.287 (-1.23)	-0.046 (-0.22)	0.186 (0.70)	0.747 (2.85)	0.185	236

Notes: Alpha is annualized and reported in percent. The remaining coefficients are monthly factor loadings. Newey–West HAC t-statistics with four monthly lags are reported in parentheses.

A6 Full Fama–French five-factor estimates, equal-weighted portfolios

Table A6: Full Fama–French five-factor regression estimates, equal-weighted portfolios

Portfolio	Alpha	MKT	SMB	HML	RMW	CMA	Adj. R2	N
ALL	4.54 (2.22)	0.897 (18.73)	0.374 (3.13)	0.419 (4.09)	0.251 (1.72)	-0.140 (-0.69)	0.750	254
Product-based	4.13 (2.09)	0.866 (19.90)	0.353 (2.82)	0.268 (2.30)	0.212 (1.33)	0.114 (0.63)	0.684	254
Conduct-based	6.64 (1.63)	0.958 (9.55)	0.367 (1.65)	0.805 (4.44)	0.401 (1.48)	-0.763 (-2.18)	0.533	248
Coal	4.95 (1.49)	0.650 (10.05)	0.512 (2.89)	0.025 (0.19)	0.324 (1.26)	0.585 (2.54)	0.489	118
Tobacco & Cannabis	0.73 (0.26)	0.766 (14.91)	0.339 (1.92)	0.123 (0.89)	0.384 (2.11)	0.527 (2.46)	0.492	193
Weapons	5.07 (2.35)	1.073 (19.79)	0.258 (1.77)	0.394 (2.84)	0.114 (0.62)	0.003 (0.02)	0.727	254
Environmental	5.27 (0.91)	1.052 (7.48)	0.746 (2.98)	0.956 (3.67)	0.484 (1.35)	-1.032 (-1.89)	0.478	236
War & Conflict	4.34 (1.12)	0.977 (12.10)	0.191 (0.73)	0.281 (1.11)	-0.372 (-1.17)	-0.408 (-1.17)	0.410	212
Human Rights	4.66 (1.28)	0.678 (11.31)	-0.201 (-0.99)	0.170 (0.92)	0.001 (0.01)	0.298 (1.40)	0.285	236

Notes: Alpha is annualized and reported in percent. The remaining coefficients are monthly factor loadings. Newey–West HAC t-statistics with four monthly lags are reported in parentheses.

A7 Extended sector-control coefficients

Table A7: Extended sector-control regressions for Coal and Weapons

Specification	Alpha	Industry beta	MKT	SMB	HML	RMW	CMA	Adj. R^2	N
<i>Panel A: Coal</i>									
Baseline FF5	8.51 (3.19)	– –	0.618 (10.49)	0.196 (1.14)	0.089 (0.52)	0.298 (1.19)	0.548 (1.98)	0.482	118
FF5 + Coal industry	7.78 (3.42)	0.057 (2.13)	0.556 (8.96)	0.117 (0.67)	0.067 (0.37)	0.338 (1.37)	0.523 (1.94)	0.505	118
<i>Panel B: Weapons</i>									
Baseline FF5	6.51 (2.80)	– –	1.068 (18.30)	0.066 (0.46)	0.292 (2.00)	0.133 (0.77)	0.354 (2.18)	0.670	254
FF5 + Aerospace	4.48 (3.40)	0.659 (16.53)	0.319 (5.71)	0.037 (0.33)	-0.056 (-0.70)	0.043 (0.40)	0.293 (2.72)	0.860	254
FF5 + Guns	2.75 (1.41)	0.382 (10.51)	0.764 (11.46)	0.172 (1.58)	0.367 (2.61)	0.307 (2.08)	0.087 (0.59)	0.772	254

Note: Alpha is annualized and reported in percent. The remaining coefficients are monthly factor or industry loadings. Newey–West HAC t-statistics with four monthly lags are reported in parentheses. The specifications match Table 5.3.

Bibliography

- Berle, E., He, W., and Ødegaard, B. A. (2025). The stock market and corporate consequences of ethical exclusions by the world's largest fund. *Journal of International Financial Markets, Institutions & Money*, 102:102174.
- Blitz, D. and Swinkels, L. (2023). Does excluding sin stocks cost performance? *Journal of Sustainable Finance & Investment*, 13(4):1693–1710.
- Council on Ethics (2024). Annual report 2024. Council on Ethics for the Norwegian Government Pension Fund Global. Available at: <https://www.regjeringen.no/en/documents/annual-report-2024/id3090916/>. Retrieved April 10, 2026.
- Council on Ethics (2025). Recommendation to exclude caterpillar inc. from the government pension fund global. Available at: <https://files.nettsteder.regjeringen.no/wpuploads01/sites/275/2025/08/Caterpillar-ENG.pdf>. Retrieved March 18, 2026.
- Council on Ethics (n.d.). Recommendations. Available at: <https://etikkradet.no/recommendations/>. Retrieved March 18, 2026.
- Døskeland, T. and Sjuve, A. W. (2025). Evaluating the performance of active management in norway's sovereign wealth fund (gpfg). Technical report, NHH Norwegian School of Economics. Available at: https://www.regjeringen.no/contentassets/f353169233704a55b3af6b0b36fb3129/evaluation_report_gpfg_2025.pdf. Retrieved May 3, 2026.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- French, K. R. (2026). Data library: Developed 5 factors, 30 industry portfolios, and 49 industry portfolios. Tuck School of Business, Dartmouth College. Available at: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Retrieved April 10, 2026.
- Hoepner, A. G. F. and Schopohl, L. (2018). On the price of morals in markets: An empirical study of the swedish ap-funds and the norwegian government pension fund. *Journal of Business Ethics*, 151:665–692.
- Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1):15–36.
- Huberman, G. and Kandel, S. (1987). Mean-variance spanning. *The Journal of Finance*, 42(4):873–888.
- Kan, R. and Zhou, G. (2012). Tests of mean-variance spanning. *Annals of Economics and Finance*, 13(1):139–187.

- Ministry of Finance (2022). Guidelines for observation and exclusion of companies from the government pension fund global. Government of Norway. Available at: https://www.regjeringen.no/contentassets/9d68c55c272c41e99f0bf45d24397d8c/2022.09.05_gpfg_guidelines_observation_exclusion.pdf. Retrieved April 20, 2026.
- Ministry of Finance (2025a). The government follows up on the parliament's decision regarding the ethical framework for the pension fund. Government of Norway. Available at: <https://www.regjeringen.no/en/whats-new/the-government-follows-up-on-the-parliaments-decision-regarding-the-ethical-framework-for-the-pension-fund/id3136331/>. Retrieved April 10, 2026.
- Ministry of Finance (2025b). Mandate for the committee appointed to assess the ethical framework for the government pension fund global. Government of Norway. Available at: <https://www.regjeringen.no/no/dep/fin/org/styrer-rad-og-utvalg-oppevnt-av-finansdepartementet2/utvalg-skal-vurdere-det-etiske-rammeverket-for-statens-pensjonsfond-utland/id3138380/>. Retrieved April 10, 2026.
- Ministry of Finance (2026). Meld. st. 7 (2025–2026): Statens pensjonsfond 2026. Government of Norway. Available at: <https://www.regjeringen.no/contentassets/f143d0a97dae41e38658800102dcd6d8/en-gb/pdfs/meld-st-7-20252026-executive-summary.pdf>. Retrieved April 10, 2026.
- NBIM (2025). Decisions on exclusion. Norges Bank Investment Management. Available at: <https://www.nbim.no/en/news-and-insights/the-press/press-releases/2025/decisions-on-exclusion/>. Retrieved April 10, 2026.
- NBIM (n.d.). Exclusion of companies. Norges Bank Investment Management. Available at: <https://www.nbim.no/en/responsible-investment/exclusion-of-companies/>. Retrieved April 10, 2026.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.
- Pástor, L., Stambaugh, R. F., and Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2):550–571.
- Trinks, P. J. and Scholtens, B. (2017). The opportunity cost of negative screening in socially responsible investing. *Journal of Business Ethics*, 140:193–208.