The Cost of ESG Investing*

Laura Lindsey
Arizona State University
laura.lindsey@asu.edu

Seth Pruitt
Arizona State University
seth.pruitt@asu.edu

Christoph Schiller
Arizona State University
christoph.schiller@asu.edu

August 20, 2022

Abstract

Socially-responsible investment mandates can cost nothing. Optimal systematic portfolios, using many approaches and rich information in asset characteristics, can be tilted to achieve ESG investing goals with negligible effects on performance. Nonetheless, strategies based on ESG-based mispricing can be profitable if we pool information across ESG scores or use specific environmental criteria. Our evidence is inconsistent with ESG measures conveying novel information about systematic risk; but it is consistent with investors placing significant weight on certain ESG subcomponent information, and with ESG-driven mispricing having occurred.

Keywords: ESG, IPCA, tangency portfolio, portfolio tilt, responsible investing, sustainable investing

*We thank Florian Berg (discussant), Sam Hartzmark (discussant), Philipp Illeditsch, Wenxi (Griffin) Jiang (discussant), Andrew Karolyi, Pedro Matos (discussant), Markus Pelger, Roberto Rigobon, Laura Starks, Luke Taylor, Jeffrey Wurgler (discussant), and Feng Zhang (discussant) for many valuable suggestions. We also thank seminar participants at Georgetown University and CUHK, and participants at the 2022 AFA Meeting, 2022 Kentucky Finance Conference, 2022 Oklahoma Energy and Climate Finance Research Conference, 2022 ABFER Meeting, 2022 Mid-Atlantic Research Conference, 2022 GSU-CEAR Asset Pricing Conference, 2022 Young Scholars Finance Consortium, 2022 Rotman Sustainability Roundtable, and 2022 Pacific Center for Asset Management (PCAM) research conference for many helpful comments. All remaining errors are our own. We are grateful for financial support from PCAM and the International Centre for Pension Management (ICPM), and data provision from MSCI ESG Research LLC.

1 Introduction

Over the past two decades, the amount of investment linked to Environmental, Social, and Governance (ESG) goals has seen tremendous growth (see Bialkowski and Starks, 2016). According to the 2020 Global Sustainable Investment Review, sustainable-investing assets reached $35.3 trillion globally at the start of 2020, a 15% increase over 2018 to represent almost 36% of total assets under management. Similarly, the number of signatories of the United Nations 'Principles for Responsible Investment' (PRI), institutional investors committed to ESG-oriented investment decisions, has increased from 734 to over 3000 between 2010 and 2020.

With rapidly growing demand from clients, fund managers are increasingly looking for ways to integrate ESG goals into their investment strategies. However, the implications for portfolio efficiency and performance from such actions are unclear. Economic theory generally argues that, all else equal, high-ESG firms should have lower expected returns since socially-oriented investors require less compensation for holding high-ESG firms (e.g. Fama and French, 2007; Pedersen et al., 2020; Pastor et al., 2021). While Hong and Kacperczyk (2009), Luo and Balvers (2017), Bolton and Kacperczyk (2020), and Pastor et al. (2022) empirically document higher risk-adjusted returns for "sin" stocks and high-carbon-emissions firms, Edmans (2011) and Glossner (2021), among others, find that high-ESG firms outperform. Perhaps unsurprisingly given this mixed evidence, many fund managers who publicly commit to responsible-investment goals do little to improve the ESG performance of their portfolios (Kim and Yoon, 2020; Brandon et al., 2021). Without a clear picture on the costs of pursuing ESG strategies, fiduciaries without a specific ESG mandate are naturally reluctant to cater to changing investor demands. Hence, there may be a growing tension between green-washing managers and clients who "want to own ethical companies in a saintly effort to promote good corporate behavior, while hoping to do so in a guiltless way that does not sacrifice returns" (Pedersen et al., 2020).

Our main result is that optimal stock portfolios *can be adjusted* to achieve responsible-investment goals *without sacrificing returns*—that is, ESG investing can have no cost. In contrast to previous research, which has primarily relied on Fama and French (1993) factors with static regression-based betas, we consider ESG ratings within the context of a conditional asset-pricing model, using the instrumented principal components analysis (IPCA) approach of Kelly et al. (2019, 2020). Our empirical methodology allows us to bring rich conditioning information into estimates of firms' risk exposures, which in turn drive systematic-portfolio weights. By including (what must be) just some of asset-managers' large information set, we find optimal portfolios with performance robust to a range of ESG-investing mandates.

We follow the taxonomy of ESG mandates as surveyed in Dimson et al. (2020). They report that *negative-screening* accounts for nearly one-half of global assets tied to ESG-investing: we find that these mandates have small effects on the portfolio's average return, broadly across ESG data providers.[1] *Positive-screening* also exhibits modest effects on performance. Going further, we consider the *responsible-investing models* of Pedersen et al. (2020) and Pastor et al. (2021), which derive optimal portfolios incorporating ESG preferences, and implement them using our empirical model estimates. Similar to before, we find small performance effects for a range of reasonable ESG-preference settings. Taken altogether, the evidence indicates that many systematic responsible investment strategies are possible with little sacrifice to returns.

Next, we broaden our investigation to evaluate the information content of ESG scores with respect to either systematic or non-systematic portfolios—allowing us to consider *ESG-integration* mandates which account for another one-third of global assets according to Dimson et al. (2020). We find that ESG measures do not tell us anything about systematic risk that is new to asset-managers' existing and rich information set. However, we show that particular ESG scores or combinations of ESG scores *can* define non-systematic portfolios

[1] We group together negative- and norms-based screening discussed by Dimson et al. (2020).

with economically- and statistically significant average returns.[2] This implies that *some* ESG attributes lead to mispricing and can generate abnormal profits *if information is carefully chosen or elicited*—which echoes results in Pastor et al. (2022) and Berg et al. (2021), as we describe further below.

How might investors care about the ESG performance of firms, and yet prices fail to adjust such that ESG measures predict returns? To explain this observation, we consider the equilibrium model of Pastor et al. (2021) and propose a simple extension: disagreement across ESG criteria. As our robust empirical findings across different ESG data providers and ESG mandates indicate, there are many ways to "do ESG". If investors do not agree on the definition of ESG criteria, equilibrium pricing need not reflect their ESG concerns, as our systematic-portfolio results suggest (and essentially consistent with Berg et al., 2022; Avramov et al., 2021; Christensen et al., 2021; Gibson et al., 2021). Yet, if there is information upon which more investors agree, there could be exploitable mispricing—in line with our non-systematic portfolio results. These contrasting ideas might be rationalized by an idea as in Pastor et al. (2022), who measure significant ESG alpha over the past decade but argue it results from unforecastable shocks that cannot be expected to persist.

Overviewing our results in more detail, the core of our analysis is estimating a conditional asset-pricing model via IPCA, following Kelly et al. (2019) who show the model works well for stock returns as we consider here. Our benchmark analysis estimates both factors and betas (loadings) on those factors using returns and lagged characteristic data.[3] As we detail below, the estimated betas and factor means and covariances define a tangency portfolio on the mean-variance efficient frontier. It is the weights of this tangency portfolio that we negatively- or positively screen to achieve an ESG-investing mandate. Going further, adding in the estimated idiosyncratic variances delivers stocks' conditional mean vector and

[2]These portfolios are non-systematic because they are orthogonal to the estimated risk factors, as explained further below.

[3]The main results are robust to instead taking the factors as given from Fama and French (2015) and Carhart (1997).

covariance matrix. These two objects deliver Markowitz portfolio weights that the Pedersen et al. (2020) and Pastor et al. (2021) models tilt using ESG information, thereby delivering responsible-investing portfolios. Common across these approaches is that they start with optimal systematic portfolios that can be derived in the absence of ESG information, and tilt those portfolios in response to an ESG-investing mandate or preference. As mentioned above, we find a range of mandates/preferences that deliver small effects on portfolio performance.

Alternatively, we allow ESG information to directly determine portfolio weights from the beginning—what is called ESG-integration. Our first way of accomplishing this is systematic, by including ESG measures in our beta and factor estimation via IPCA. Should our estimated tangency portfolio improve markedly, then the ESG measures are providing incrementally relevant information on systematic risk exposures. Our second way of integrating ESG is non-systematic, by including ESG measures in our estimation of alpha. That is, we find an ESG-based portfolio that is orthogonal to factor risk, and see if this portfolio delivers significant profits. In order to find this portfolio, which we label *beta-neutral*, we develop an extension of IPCA that separates the characteristics entering alpha and beta, as in concurrent work by Chini and Rubin (2022).

An important aspect of our work is the large extent of ESG information we consider. Our ESG-tilted (screened- and responsible-investing) portfolio results and systematic ESG integration results are robust to: using data from seven different ESG providers; using different ESG index subcomponents; using industry-adjusted measures or not; and, various choices of missing-value imputation. Our beta-neutral portfolio results are also extensive, but there we find important distinctions: certain providers' individual E subcomponent indices deliver positive alpha, whereas the topline or S and G subcomponent indices do not; and, combining topline, E, or S subcomponents across data providers also show evidence of mispricing. A few of these last observations cleanly connect to Pastor et al. (2022) and Berg et al. (2021), whose conclusions are supported and enhanced by our results.

Our paper fits in a vibrant literature investigating the growing importance of responsible investing and ESG information. Despite extensive research, there is widespread disagreement in the literature on the return predictability of ESG characteristics. The lack of return predictability in our systematic portfolios echoes Hartzmark and Sussman (2019) who find no evidence that sustainable funds outperform non-sustainable funds, Pedersen et al. (2020) who find that the KLD ESG scores do not significantly predict returns and carbon emissions do not yield value-weighted alphas, and Gorgen et al. (2020) who find insignificant differences in average returns for high- and low-carbon-emissions firms. In contrast, a long literature on so-called "sin" stocks has found a premium for firms in industries like alcohol or tobacco.[4] In a similar vein, Zerbib (2020) uses the holdings of "green" institutional investors to show that excluded firms have a significantly higher average returns. Glossner (2021) documents a negative Carhart (1997) alpha of -3.5% for firms with high reputation risk using RepRisk ratings, and Baker et al. (2018) and Zerbib (2019) find a positive "greenium" for green bonds over similar non-green bonds. Giglio et al. (forthcoming) provide a recent survey of the large and growing literature on the effects of ESG-investing across many different asset classes.

We contribute to this rapidly growing literature along several important dimensions. First, we use IPCA to extract aggregate risks that better-capture the mean-variance-efficient frontier, as has been argued in Kelly et al. (2019, forthcoming). It is crucial to have the best-possible depiction of systematic risks when we evaluate how firms' ESG scores lead to differences in average returns, so as to appropriately understand ESG's effects if they are risk-based, rather than inappropriately attribute them to an alpha because one's factor model is poor. Second, we take into account a large set of other firm characteristics, controlling for a substantial amount of the conditioning information investors have at their disposal *already* in addition to ESG scores. Third, we use data from seven major ESG providers (and evaluate both aggregate and subcomponent performance) in our empirical analysis, making

[4] Among others, Fabozzi et al. (2008), Luo and Balvers (2017), and Pedersen et al. (2020) find that non-sin stocks earn negative CAPM and Fama and French (1993) alphas.

our conclusions broad. Fourth—and this pertains even when we use the same factor models as other papers—we explicitly allow for ESG measures and other firm characteristics to drive cross-sectional and time-series variation in alphas, betas, or both. This way we can comprehensively evaluate ESG's role in pricing assets, and distinguish whether a *conditional* risk-based or mispricing-based explanation best fits ESG's impact on returns.

Our paper also contributes to the literature on the costs of implementing ESG investment mandates. Kim and Yoon (2020) and Brandon et al. (2021) document that signatories of the UN Principles of Responsible Investment in the U.S. experience a significant increase in fund inflows, but do not significantly increase fund-level ESG performance in their portfolios after committing to ESG-investment goals, while also experiencing a decrease in returns (Kim and Yoon, 2020). Ceccarelli et al. (2021) show that funds that received a 'low-carbon' label by Morningstar in 2018 experienced significant fund inflows. While these funds outperformed conventional funds in months with high salience of climate change risk, they offered significantly lower diversification benefits throughout the sample. Similarly, Aragon et al. (2020) find that university endowments receive higher donations following the adoption of socially responsible investment (SRI) policies but exhibit greater management costs and portfolio return volatility. Our results demonstrate how fund managers can implement a wide range of ESG mandates without substantially compromising Sharpe ratios relative to the tangency portfolio.

The paper proceeds as follows. Section 2 discusses the data and availability of ESG measures. Section 3 discusses the empirical factor model, responsible-investing models, the formation of systematic and non-systematic portfolios, and how ESG screens can be implemented. Section 4 presents the empirical results and extensive robustness analysis, with further pointers to the online appendix. Section 5 discusses the effect of responsible-investing mandates on portfolios' ESG performance, and explores how disagreement can drive some of the results we find. Section 6 concludes.

2 Data

2.1 Returns and firm characteristics

Our data for returns and firm characteristics are obtained from CRSP and Compustat via the codes provided by Jensen et al. (forthcoming). We select fifty characteristics, based on those that provide the greatest firm-month coverage, which we refer to by their names in Jensen et al. (forthcoming). They are: `market_equity` and `assets`; cash-flow variables `net_income`, `sales`; pay-out ratios `eqnpo_1m`, `eqnpo_3m`, `eqnpo_6m`, `eqnpo_12m`, `ni_at`; change in shares `chcsho_1m`, `chcsho_3m`, `chcsho_6m`, `chcsho_12m`; valuation ratios `div3m_me`, `div6m_me`, `div12m_me`, `at_me`, `ni_me`, `nix_me`, `sale_me`, `xido_at`; leverage ratios `debt_me`, `netdebt_me`, `debt_at`; turnover, trading, and volume variables `tvol`, `zero_trades_21d`, `zero_trades_126d`, `dolvol_126d`, `turnover_126d`, `dolvol_var_126d`, `turnover_var_126d`, `zero_trades_252d`, `bidaskhl_21d`, `rvolhl_21d`; past return variables `ret_1_0`, `ret_2_0`, `ret_3_0`, `ret_3_1`, `ret_6_0`, `ret_6_1`, `ret_9_0`, `ret_9_1`, `ret_12_0`, `ret_12_1`, `ret_12_7`; quality-minus-junk `qmj_safety`, `qmj_prof`; and, other variables `seas_1_1an`, `age`, `mispricing_perf`. In robustness checks, we restrict our attention to a subset of "slow" characteristics, defined as having a low time-series volatility—this excludes all of the past-return variables, some trading variables, and most valuation ratios. These slow characteristics are: `market_equity`, `div3m_me`, `div6m_me`, `div12m_me`, `qmj_safety`, `tvol`, `dolvol_126d`, `zero_trades_252d`, `age`, `assets`, `net_income`, `qmj_prof`, `ni_at`, `debt_me`, `netdebt_me`, `sales`, and `sale_me`.

In order to estimate IPCA we require a firm-month observation to have all lagged characteristics and the month's return to be nonmissing. Figure 1 reports the time-series of the number of all firms' observations as the solid black line. As will shortly become evident, it is useful to also restrict attention to a sample of large firms. To do so, we obtain NYSE breakpoints from Ken French's data library and define the large-firm cut-off as the median. The number of large firm observations is plotted in Figure 1 as the solid red line.

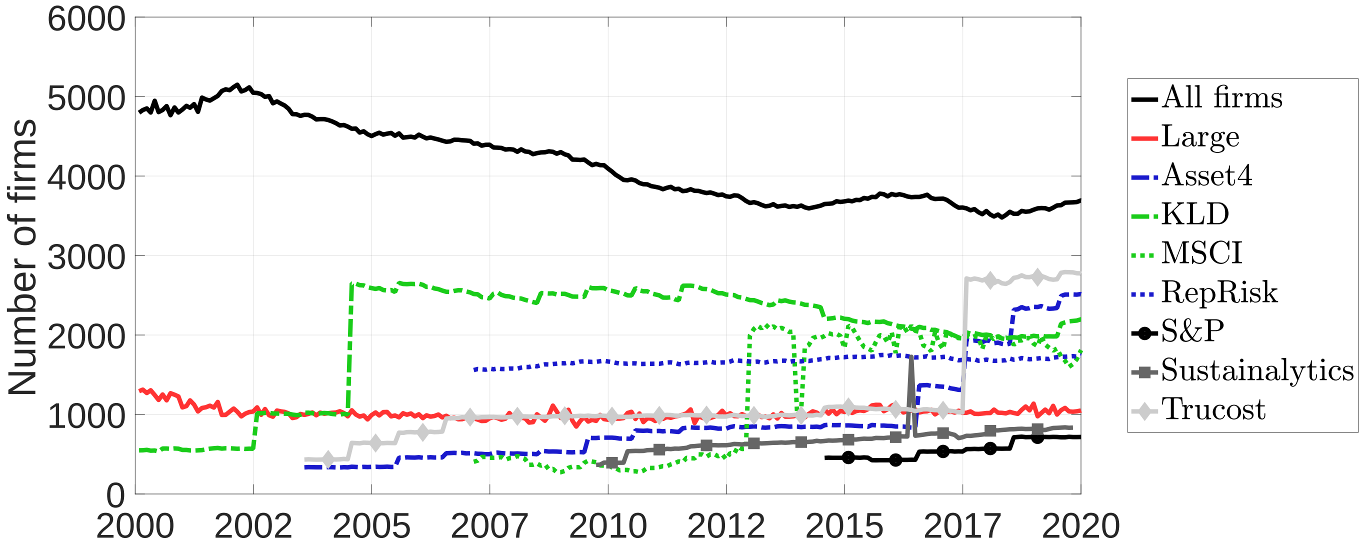

Figure 1: ESG availability and firm size

Notes – The total number of firm observations from the Jensen et al. (forthcoming) data (subject to our characteristic choice) is represented by the black solid line. The red solid line is the number of firms in the data above the NYSE median size break point. The remaining lines (of varying type, marker, and color) show how many firms have ESG data available from each provider. MSCI data are reproduced by permission of MSCI Research LLC ©2022 MSCI Research LLC All rights reserved.

2.2 ESG characteristics

We obtain data on firm-level ESG scores from seven major data providers commonly used by investors and in the academic literature (see e.g. Berg et al., 2022; Huang et al., 2021). Our first data source is MSCI ESG KLD STATS (KLD), which is available from 1992 to 2018. KLD was the first provider of socially responsible investing information in North America and continues to be widely used in academic settings given its length of coverage. For each covered firm, KLD uses a binary system to evaluate "strengths" and "concerns" across a wide range of firm attributes within the following six dimensions of ESG: environmental impact (E), community relations, product characteristics, employee relations, diversity, and governance (G). Following the literature (e.g. Di Giuli and Kostovetsky, 2014), we construct the scores for each of the six dimensions as well as the overall ESG score as the sum of strengths minus the sum of concerns. We summarize community relations, product characteristics, employee relations, and diversity as the "social" (S) category, as is standard in the literature.[5]

Second, we obtain ESG scores from MSCI, which is widely used by both investors and academics (e.g. Berg et al., 2022). MSCI generates ESG scores by assessing thousands of data points across environmental (climate change, natural capital, pollution & waste, environmental opportunities), social (human capital, product liability, stakeholder opposition, social opportunities), and governance (corporate governance, corporate behavior) issues, using data from government sources, NGOs, corporate disclosure, models, and media sources. Following the literature (e.g. Berg et al., 2021), we primarily focus on MSCI's topline ESG score, which is materiality-weighted and industry-adjusted relative to peers' ESG performance.[6] While ESG scores are available at the monthly frequency starting in 2007, coverage increased substantially in 2013, with a temporary dip in late 2013, as shown in Figure 1.

Third, we construct ESG scores using data from Thomson Reuters Asset4 (now published under the name 'Refinitiv ESG').[7] Asset4 coverage starts in 2003 and includes ESG information based on over 450 individual data points across three "pillars": 'E' (emissions, resource use, product innovation), 'S' (workforce, human rights, community, product responsibility) and 'G' (management, shareholders, CSR strategy).[8] In a survey of investors by SustainAbility, this data source was noted for its granular quantitative data.[9] In our main analysis we focus on the topline ESG score and the E, S, and G "pillar" scores from Asset4, which are adjusted for materiality and ESG performance relative to peers at the industry- and country (for governance) level for each data item.[10]

[5]In our robustness tests we also consider alternative KLD scores that are adjusted for changes in KLD's methodology and underlying data items.

[6]In additional tests we also consider MSCI ESG scores without industry adjustment as well as individual category scores for E, S, and G.

[7]We continue to refer to this dataset as 'Asset4' in our analysis for consistency with prior literature.

[8]Asset4 further overlays their ESG score with indicators for ESG-related controversies.

[9]"Rate the Raters 2020" available at https://www.sustainability.com/thinking/rate-the-raters-2020/.

[10]Following Dyck et al. (2019), we also construct unadjusted ESG scores using an equal weighting scheme of the underlying ESG data items in additional robustness tests. This also helps us address concerns about changes in the Asset4 data aggregation methodology throughout our sample period as highlighted by Berg et al. (2020).

Fourth, we obtain ESG scores from Sustainalytics (now owned by Morningstar). Sustainalytics constructs ESG scores based on hundreds of individual data items according to a proprietary weighting scheme. We obtain the E, S, and G category scores as well as the aggregated topline ESG score for the sample period from 2009 to 2019. In the SustainAbility survey, Sustainalytics was mentioned as one of the most high-quality and useful providers by both investors and industry experts.

Fifth, we obtain ESG data from RepRisk, which is available to us for the sample period from 2007 to 2020. RepRisk uses both algorithms and analysts to monitor company-specific news events related to 28 ESG issues (e.g. air pollution, product controversies, discrimination, and labor practices) using over 80,000 public sources in 20 languages such as print and social media, regulators, think tanks, and newsletters. The company advertises its transparency of methods and the external nature of its sources, which provide a counterpoint to providers relying primarily on company-provided data. Based on the occurrence of ESG-related controversies, RepRisk provides a Reputation Risk Rating (RRR) using a letter rating (AAA to D). We translate this letter scale to a numerical scale (ranging from 1 to 10 in one-unit increments) such that a higher number indicates a better rating.

Sixth, we add ESG scores from S&P Global. S&P Global uses an industry-specific weighting scheme across 130 question-level items per firm to construct scores across the E, S, and G dimensions, considering a firm's ESG data availability, quality, relevance, and performance. We obtain the topline ESG score as well as the E,S, and G dimension scores. Compared to the other ESG data sources in this paper, S&P Global ESG coverage starts late (coverage begins in 2013) and is very sparse even among large firms, as shown in Figure 1. Hence, results based on S&P should be interpreted with caution. We primarily include this dataset for completeness.

Seventh, and last, we include data on greenhouse gas (GHG) emissions at the firm-year level from S&P Trucost. Compared to the other six ESG data sources, Trucost GHG data does *not* aggregate ESG-related data items across a wide range of topics, but rather focuses on a single, quantifiable issue, i.e. greenhouse gas emissions. Given its nature, 'Trucost GHG' is therefore primarily a measure of environmental (E) performance. S&P Trucost uses a variety of data sources, including the Carbon Disclosure Project (CDP), EPA filings, corporate social responsibility and sustainability reports, and financial disclosure documents to collect GHG emissions information. The data is available starting in 2004. Our main measure of firms' GHG emissions is the sum of direct and tier-1 supplier emissions, scaled by revenue. This includes both the emissions resulting directly from the firms' own operations (i.e. 'scope 1') and from the firm's first-tier upstream supply chain—their direct suppliers. We choose this measure to capture all GHG emissions over which the firm has some control, and multiply it by -1 so that higher emissions mean a lower score.[11]

The ESG measures are reported at different frequencies. In particular, KLD, Asset4, and S&P ESG scores and Trucost GHG emissions are reported annually, while MSCI, Sustainalytics, and RepRisk scores are available at the monthly frequency. For the latter, timing them is quite simple: they are in the investor's information set at the end of that month in which they are reported. But the former are tougher to time definitively. In fact, this issue is quite similar to the well-known issue of timing firms' accounting variables for the purpose of portfolio sorting, for instance as done by the seminal Fama and French (1993)—therefore we adopt their well-known convention. If a score that is available at the annual frequency (KLD, Asset4, S&P, Trucost) is given for year y, we assume that it is observed by the investor starting in June of year $y + 1$ and remains constant for the subsequent twelve months.

2.3 ESG coverage and summary statistics

The availability of our ESG measures varies tremendously over the sample period, as Figure 1 makes plain. The KLD measures (green dashed-dot line) are available starting in 1992 for a

[11] In additional tests we also use scope 2 emissions, i.e. Greenhouse gas (GHG) emissions from consumption of purchased electricity, heat or steam, as an alternative measure of GHG emissions.

small number of firms, with noticeable increases in coverage in 1996, 2002, and particularly 2004. Asset4 measures (blue dashed-dot line) start in 2004, again for a small number of firms, and with a noticeable increase in coverage in 2016. Trucost (light gray marked line) also begins in 2004, with a gradual increase in coverage until a large expansion in 2017. MSCI (green dotted line) starts in 2007 with modest coverage, and expands in 2013 to reach coverage levels similar to KLD. RepRisk (blue dotted line) starts in 2007 with relatively large coverage that slightly declines over time. Sustainalytics (dark gray marked line) starts in 2009 with a small number of firms, bumps up in 2010, and remains steady. S&P (black dot marked line) begins much later and has lower coverage than other providers. This issue of ESG score availability is one we take seriously by a variety of means.

Figure 2 illustrates that ESG coverage is related to firm size, reported on a log scale. In each panel, percentiles of the distribution of all firms with available ESG coverage is reported by gray lines: the minimum and maximum (p_0 and p_{100}, respectively) as dotted lines at the top and bottom, the p_{10} and p_{90} as dashed lines closer to the middle, and the median p_{50} as a solid line. In addition, the NYSE-median large-firm cut-off is plotted as the red dashed line. The gray and red lines are identical in every panel. What changes between panels are the blue lines, which plot the percentiles of the size distribution of firms for which that provider's ESG score is nonmissing. As with the gray lines, we use dotted lines for p_0 and p_{100}, dashed-dot lines for p_{10} and p_{90}, and a solid line for p_{50}.

Figure 2 broadly says that ESG coverage is skewed towards large firms. We see this plainly for the KLD data panel. For the first several years, the median firm with a KLD rating is as big as the 90^{th} percentile of all firms, judging by the relationship of the solid blue line to the top gray dashed-dot line. A bit less than 90% of the KLD firms are above the NYSE median, judging from how the bottom blue dashed-dot line hovers below the red dashed line. The large expansion in 2004 noticeably drops the 10-50-90 percentiles, implying increased coverage of small firms. Nonetheless, throughout its history the KLD percentiles lie above

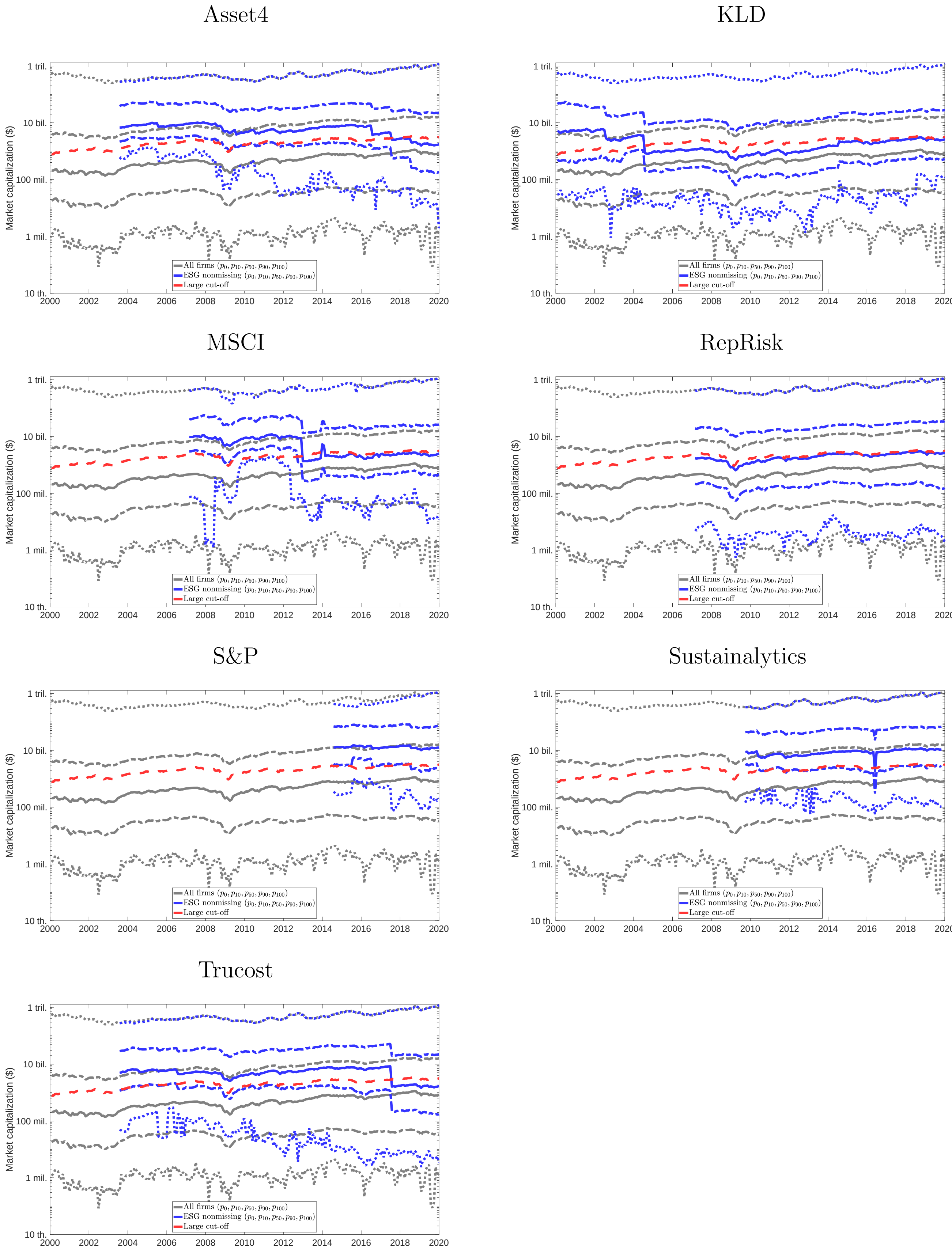

Figure 2: Firm size and ESG availability

Notes – Cross-sectional size distribution over time, for each ESG data provider. The dotted lines show the minimum p_0 and maximum p_{100}, the dashed lines show the 10th percentile p_{10} and 90th percentile p_{90}, and the solid line the median p_{50}. In all panels: the gray lines show the distribution of all firm data using our characteristics using Jensen et al. (forthcoming), and the red dashed line shows the NYSE median breakpoint. For each panel, the blue lines show the distribution of firms for that ESG provider. MSCI data are reproduced by permission of MSCI Research LLC ©2022 MSCI Research LLC All rights reserved.

the percentiles of all firms, showing us that the coverage is better for larger firms.

The remaining panels show that this feature is similar for other ESG providers. The largest 90% of Asset4 firms are larger than the median of the firm size distribution for almost its entire history, with this p_{10} line lying close to the NYSE median for the most part. The same is true for Trucost and S&P, at least until Trucost's expansion in coverage at that the end of the sample. The median of Sustainalytics firms lies just below the p_{90} of all firms, and the 10^{th} percentile of Sustainalytics firms is just about at the NYSE median. Similarly, for MSCI and RepRisk, half of firms are above the NYSE median and distributions are consistently skewed towards large firms.

The broad takeaway here is that firms that receive ESG coverage tend to be bigger. Hence, we will have a paucity of ESG information in the sample of all firms. For this reason, our main tests restrict attention to the sample of large firms, defined as those larger than the NYSE median. This reduces the impact of imputing ESG scores when we do so. Furthermore, Kelly et al. (2019) show that systematic-investment performance is lower in large firms, which we also observe in our data. Therefore large firms provide a more-stringent test of systematic strategies' profitability and the impact of ESG scores thereupon.

3 Model and Portfolio Construction

In this section we describe IPCA, responsible-investing models, and the systematic and non-systematic portfolios based thereupon.

3.1 Basic IPCA model

Our basic framework is Kelly et al. (2019)'s restricted IPCA model:

$$r_{n,t+1} = \beta'_{n,t} f_{t+1} + \varepsilon_{n,t+1}, \quad \text{where } \beta_{n,t} = \Gamma'_\beta z_{n,t}, \tag{1}$$

for the $K \times 1$ exposure $\beta_{n,t}$ to the $K \times 1$ factors f_{t+1}, and the $L \times 1$ firm characteristics $z_{n,t}$. The timing says that $\beta_{n,t}$ is known before f_{t+1}, which follows arbitrage-pricing theory. By estimating Γ_β we allow firm characteristics to give information on how a stock's exposure to aggregate factors varies both cross-sectionally and over time. The factors f could be jointly estimated along with Γ_β, or instead the factors could be exogenously specified as portfolio returns representing systematic risk. In either case, Γ_β is estimated by a large panel regression of stock returns on the interaction of factor realizations and lagged firm characteristics. Following Kelly et al. (2019), we stack $r_{n,t+1}$ into the vector r_{t+1} and $z'_{n,t}$ into the $N_t \times L$ matrix Z_t,[12] so we can concisely state the first-order conditions upon which one iterates until convergence to the least-squares estimates:

$$f_{t+1} = \left(\Gamma'_\beta Z'_t Z_t \Gamma_\beta\right)^{-1} \Gamma'_\beta Z'_t \left(r_{t+1} - Z_t \Gamma_\alpha\right) \tag{2}$$

$$\text{vec}(\Gamma'_\beta) = \left(\sum_{t=1}^{T-1} Z'_t Z_t \otimes f_{t+1} f'_{t+1}\right)^{-1} \left(\sum_{t=1}^{T-1} \left[Z_t \otimes f'_{t+1}\right]' r_{t+1}\right). \tag{3}$$

Kelly et al. (2019) and Kelly et al. (forthcoming) use this model to describe stock and bond returns, respectively, and find unprecedented success by a variety of measures.[13] Furthermore, they find that estimating f leads to significant gains, relative to instead exogenously taking the factors as well-known portfolios (such as Fama and French, 2015; Hou et al., 2015, amongst others). Kelly et al. (2021) emphasizes a key point: an element of Γ_β is nonzero only to the extent that the corresponding characteristic meaningfully drives differences in

[12] Empirically the number of stocks varies, hence the notation N_t.

[13] Kelly et al. (2020) provide asymptotic analysis of the estimator.

a return's covariance with aggregate risk. Hence, we construct a *systematic portfolio* when basing the portfolio weights on $\beta_{n,t}$, even when they are instrumented by characteristics.[14]

3.2 Systematic investment strategies

Systematic investment strategies are based only on stocks' aggregate risk exposures $\beta_{n,t}$ estimated in the restricted IPCA model. Theoretically, the mean-variance-efficient frontier is provided by the tangency portfolio constructed from systematic-risk factors. Indeed, the stock evidence in Kelly et al. (2019) and Kelly et al. (2021), and bond evidence in Kelly et al. (forthcoming), suggest that IPCA-based tangency portfolios are very profitable.

Suppose that the factors have excess return mean m and covariance S, which we take as static for simplicity. Then the $K \times 1$ *factor*-tangency portfolio weights are

$$w_{factan} = \frac{1}{\iota_K' S^{-1} m} S^{-1} m \tag{4}$$

for a $K \times 1$ ones vector ι_K. Meanwhile, the IPCA-model-implied factor weights are the projection onto betas. That is, stack $\beta_{n,t}'$ into the $N_t \times K$ matrix β_t, and the $K \times N_t$ factor weights are

$$W_{f,t} = \left(\beta_t' \beta_t\right)^{-1} \beta_t' \equiv \begin{bmatrix} w_{f,1,t} & \cdots & w_{f,K,t} \end{bmatrix} \tag{5}$$

where $w_{f,k,t}$ is the portfolio weight for the k^{th} factor. Therefore, the $1 \times N_t$ tangency portfolio weights combine (4) and (5):

$$w_{tan,t}' = w_{factan}' W_{f,t} \tag{6}$$

[14]Kelly et al. (2021) provide evidence that characteristics predict future covariance with the market portfolio, as well as a host of aggregate risk factors, supporting this view.

3.3 Negative and positive screening

The main idea of ESG screening approaches is that ESG measures are not used to estimate the model, but instead to achieve an ESG mandate. Therefore, we take the $w_{tan,t}$ vectors as given by the model estimation, and then adjust certain elements $w_{i,tan,t}$ according to the chosen screening specification. Screening occurs by choosing a desired ESG threshold and which weights one will allow to be screened. From the cross-sectional distribution of the ESG measures, call it the vector ζ_t, we choose the threshold as a certain percentile, called p_{screen}, on (0,100).

We consider two negative screens and one positive screen. The first negative screen zeros out a weight $w_{i,tan,t}$ if the ESG score of firm i in t is below the threshold, i.e. $\zeta_{i,t} < p_{screen}$. This means that both long ($w_{i,tan,t} > 0$) and short ($w_{i,tan,t} < 0$) positions are screened: Pedersen et al. (2020) note the importance of this for achieving ESG mandates because bad-ESG stocks (those with negative scores) can be shorted, thereby improving the portfolio's aggregate ESG performance.[15] The second negative screen zeros out $w_{i,tan,t}$ if $\zeta_{i,t} < p_{screen}$ and $w_{i,tan,t} > 0$. This means that only long positions are screened. This type of screen allows bad-ESG stocks to be shorted, but not to be held. Our positive screen is simple: we zero out $w_{i,tan,t}$ unless $\zeta_{i,t} > p_{screen}$.[16]

It is worth noting how these screens interact with the missing values in ESG measures. Portfolios implementing the two negative screens could include firms with missing ESG information depending on what one assumes about their values: if we don't *see* that it is a bad-ESG stock, we don't screen it. The positive screen, instead, excludes firms with missing ESG information if we impute it below the threshold: if we don't *see* that it is a good-ESG stock, we don't include it. Hence, negatively-screened portfolios will include more stocks than positively-screen portfolios in our setting.

[15]We look for evidence of this in our specific empirical application below.

[16]Note, we do this regardless of whether it is a long or short position.

3.4 Responsible-investing models

In addition to screened portfolios, we construct optimal portfolios from the responsible-investing models of Pedersen et al. (2020) and Pastor et al. (2021) (for which we use the shorthand PFP and PST, respectively). These take into account firms' expected returns, covariances, and ESG information in a way that is reminiscent of the screened portfolios above. That is, the expected returns and covariances are independent of ESG information and would naturally define an optimal Markowitz portfolio in the absence of ESG concerns. The two frameworks then include ESG information into investor preferences, and thereby tilt the Markowitz portfolio, similar to how the screens tilted tangency weights. To explain the responsible-investing portfolios, we simply repeat those papers' expressions, adjusting the notation for our explicitly conditional context.

The model of Pedersen et al. (2020) assumes that investors pursue the highest possible Sharpe ratio, subject to a target average ESG score. Using w_t as the $N_t \times 1$ portfolio weight vector, s_t as the $N_t \times 1$ vector of ESG scores, define the average ESG score $\bar{s} = \frac{w_t' s_t}{s_t' \iota_{N_t}}$. Their Proposition 3 expresses the optimal weights as

$$w_{PFP,t} = \Sigma_t^{-1} \left(\mu_t + \pi_t (s_t - \iota_{N_t} \bar{s}) \right) \tag{7}$$

for the scalar π_t defined in their paper, returns' covariance matrix Σ_t, and returns' mean μ_t.[17] This expression requires that the portfolio is net long, that is $w_t' \iota_{N_t} \geq 0$, so that the average ESG score $\bar{s}$ has a natural interpretation. Conveniently, we find that our model-implied Markowitz portfolio is net long over our entire sample. Reminiscently, Pastor et al. (2021) derive optimal weights as

$$w_{PST,t} = \Sigma_t^{-1} \left(\mu_t + d s_t \right) \tag{8}$$

[17] We set the investor's relative risk-aversion parameter to equal 1.

where the scalar $d \geq 0$ is the investor's "ESG taste."[18] Assume for the moment that "bad" ESG is denoted by negative values in s: the PST weight then reduces the *effective* expected return for bad-ESG stocks. While (7) and (8) are clearly conceptually similar, some differences emerge in Section 4.2 below.

Since we use the restricted model estimates for the screened tangency portfolios, we use the same estimates to give us the mean and covariance that (7) and (8) require. Therefore $\mu_t = \beta_t \lambda$ where λ is the $K \times 1$ price of factor risk, which we estimate as the factor mean since they are tradable. For simplicity, we use a strict factor decomposition of stocks' covariance matrix and assume the idiosyncratic return covariance matrix Σ_ε is diagonal: hence, $\Sigma_t = \beta_t S \beta_t' + \Sigma_\varepsilon$.[19]

When we use w_{PFP} and w_{PST}, we should note the units of the ESG score s. As noted above in Section 2, we follow previous studies in normalizing the other firm characteristics to be ranks translated to the $[-0.5, 0.5]$ interval. So when using ESG measures in the model, we do the same: at each time the median ESG score is 0, the minimum is -0.5, and the maximum is 0.5. Therefore, for the Pedersen et al. (2020) w_{PFP}, we will want to consider $\bar{s}$ in this interval. For the Pastor et al. (2021) w_{PST}, this scaling also seems appropriate: if a firm has median ESG, then the value of $s_{nt} = 0$ in (8) implies no tilt from the Markowitz weight. In this case, we will want to set the taste parameter d so that different values of s_{nt} do not lead to unreasonable tilts away from the Markowitz weights. For instance, with $d = 0.01$ the expected return is effectively shifted by $\pm 0.5\%$ *monthly* as the ESG moves from the median to an extreme, which could be considered a large ESG adjustment.

[18] Again, set relative risk-aversion to equal 1.

[19] For any firm in our data with fewer than ten monthly observations, we set its idiosyncratic variance equal to the average idiosyncratic variance of all firms—this only affects a few hundred stock-month observations.

3.5 ESG integration and a modified IPCA model

In addition to various ESG screens and tilts, an important category of ESG strategies is focused on 'ESG integration' (see Dimson et al., 2020). To explore how ESG integration may affect portfolio performance, we next examine if ESG measures provide useful information for optimal portfolio weights. There are two main routes we take to including ESG measures into the IPCA model.

The first route includes the ESG measures in firm characteristics $z_{n,t}$ and estimates Γ_β. In this way, ESG is treated like any other firm characteristic, and the model is just the basic restricted one presented in Section 3.1. For this ESG-integrated model, we report the performance of the tangency portfolio and compare to the performance when ESG measures were not included.

The second route uses both ESG and firm characteristics in a modified IPCA model as follows. This modified model is a special case of Kelly et al. (2019)'s unrestricted model that allows for characteristic-driven mispricing $\alpha_{n,t}$ (as well as $\beta_{n,t}$.) We impose that ESG information drives alpha while non-ESG characteristics drive beta. Concurrent work in Chini and Rubin (2022) pursues an extension like this as well.

Denote the ESG measures as $L_\zeta \times 1$ vector $\zeta_{n,t}$ and other characteristics as $z_{n,t}$. In this case, we impose the that $\alpha_{n,t} = \Gamma'_\alpha \zeta_{n,t}$ and $\beta_{n,t} = \Gamma'_\beta z_{n,t}$, and the modified model is, thus,

$$r_{n,t+1} = \Gamma'_\alpha \zeta_{n,t} + z'_{n,t} \Gamma_\beta f_{t+1} + \varepsilon_{n,t+1}. \tag{9}$$

This model is not exactly Kelly et al. (2019)'s unrestricted model because $z_{n,t}$ and $\zeta_{n,t}$ are different from each other. Hence we obtain modified versions of (2) and (3) along with a new first-order condition for Γ_α; let ζ_t be the $N_t \times L_\zeta$ matrix stacking the $\zeta'_{n,t}$:

$$f_{t+1} = \left(\Gamma'_\beta Z'_t Z_t \Gamma_\beta\right)^{-1} \Gamma'_\beta Z'_t \left(r_{t+1} - \zeta_t \Gamma_\alpha\right) \tag{2.1}$$

$$\text{vec}(\Gamma_\beta) = \left(\sum_{t=1}^{T-1} f_{t+1} f'_{t+1} \otimes Z'_t Z_t \right)^{-1} \left(\sum_{t=1}^{T-1} f_t \otimes \left[Z'_t r_{t+1} - Z'_t \zeta_t \Gamma_\alpha \right] \right) \tag{3.1}$$

$$\Gamma_\alpha = \left(\sum_{t=1}^{T-1} \zeta'_t \zeta_t \right)^{-1} \left(\sum_{t=1}^{T-1} \zeta'_t \left[r_{t+1} - Z_t \Gamma_\beta f_{t+1} \right] \right). \tag{10}$$

At this point, it may be clear to the reader that ESG-integraton as we define it does not necessarily improve a portfolio's ESG performance. ESG information matters in Γ_β to the extent that it describes aggregate return covariance (Kelly et al., 2021), and matters in Γ_α to the extent it predicts returns. We take this approach to represent, perhaps, a more opportunistic usage of ESG information than screened-portfolios or responsible-investing models envision—which makes our results on ESG integration also of interest to profit-maximizing investors with no ESG concerns themselves.

3.6 Non-systematic ESG-based portfolios

In the unrestricted model of Kelly et al. (2019), which we do not use in this paper, one uses all characteristics in the beta and alpha. With that unrestricted specification $\alpha_{n,t} = \Gamma'_\alpha z_{n,t}$, one must impose an assumption to separately identify Γ_β and Γ_α: Kelly et al. (2019) impose $\Gamma'_\alpha \Gamma_\beta = 0$, meaning that risk loadings "explain as much of the asset's mean returns as possible". Based on this, Kelly et al. (2019) define a non-systematic investment strategy called a "pure-alpha portfolio"

$$w_{\alpha,t} = Z_t \left(Z'_t Z_t \right)^{-1} \Gamma_\alpha, \tag{11}$$

because $\Gamma'_\alpha \Gamma_\beta = 0$, Γ_α is a combination of characteristics that is orthogonal to every risk exposure's combination of characteristics. Moreover, this condition ensures that $w_{\alpha,t}$ is cross-sectionally orthogonal to all factors' betas because

$$\beta_t' w_{\alpha,t} = \Gamma_\beta' Z_t' Z_t \left(Z_t' Z_t\right)^{-1} \Gamma_\alpha = \Gamma_\beta' \Gamma_\alpha = 0.$$

Therefore the pure-alpha portfolio has no factor risk. We do not estimate an unrestricted IPCA model and hence do not construct such pure-alpha portfolios,[20] but explain this to motivate the non-systematic portfolios we do build from the modified model.

Naively using the strategy in (11) with the modified model of (9) does not ensure the portfolio avoids factor risk. The unrestricted model's orthogonality condition is no longer even well-defined because Γ_β has row dimension L while Γ_α has a different row dimension L_ζ.[21] Since ζ and Z are distinct, if we naively used (11) (i.e. swap out Z for ζ) we see that

$$\beta_t' \zeta_t \left(\zeta_t' \zeta_t\right)^{-1} \Gamma_\alpha = \Gamma_\beta' Z_t' \zeta_t \left(\zeta_t' \zeta_t\right)^{-1} \Gamma_\alpha \neq 0,$$

unless $Z_t' \zeta_t = 0$.

Therefore, in the modified model of (9) we need a new construction in order to arrive at a portfolio with no factor risk. We call this a "beta-neutral" portfolio:

$$w_{\alpha\perp\beta,t} = \left(I - \beta_t(\beta_t'\beta_t)^{-1}\beta_t'\right)\alpha_t = \left(I - Z_t\Gamma_\beta(\Gamma_\beta' Z_t' Z_t \Gamma_\beta)^{-1}\Gamma_\beta' Z_t'\right)\zeta_t\Gamma_\alpha. \tag{12}$$

Clearly it is the case that $\beta_t' w_{\alpha\perp\beta,t} = 0$ for every t, regardless of the value of Γ_α.[22] Interpreting $w_{\alpha\perp\beta,t}$: in order to construct a portfolio with no factor risk, one only uses the part of ζ_t that is cross-sectionally orthogonal to the factor betas. In the case where ζ contains ESG measures, those measures can deliver mispricing only to the extent they are cross-sectionally orthogonal

[20] In unreported results we found no interesting role for ESG information in pure-alpha portfolios coming from unrestricted model estimates.

[21] For example, suppose $L_\zeta = 1$ (as will be the case in Section 4): the only way the scalar Γ_α could be said to be "orthogonal" to Γ_β in any meaningful sense is when $\Gamma_\alpha = 0$ and vector multiplication "becomes" scalar multiplication; for $L_\zeta > 1$ we cannot even use this.

[22] This construction differs from Chini and Rubin (2022) who instead orthogonalize the ζ_t with respect to Z_t—we only orthogonalize with respect to $Z_t\Gamma_\beta$.

with the instrumented betas. Beta-neutral portfolios are constructed to answer the question: are there profitable ESG-based portfolios that *are not compensation for aggregate risk*?

3.7 Missing ESG Values and Estimation Details

For the screened portfolios, the missing ESG scores are not a problem: the negative and positive screens handle missing ESG in different, but apparent ways. For the responsible-investing models and the ESG-integration in the model, however, an observed ESG score is necessary. For the responsible-investing models, the portfolio formulae (7) and (8) are defined only for nonmissing ESG scores. For the ESG-integration, we require nonmissing observations to estimate the model.

Of the many ways of imputing missing ESG values, we consider mainly two. Our primary way imputes missing data with the median ESG score each month, i.e. zero. A value of zero implies that a missing ESG score necessarily contributes nothing to $\beta_{n,t}$ or $\alpha_{n,t}$. It also implies that, given that we don't see information about the firm, we assume its ESG score is average. The second way, used in robustness analysis, is born from the idea that missing ESG information could actually be a negative ESG signal about the firm—imputing an average ESG score might be far too positive. In the absence of evidence that the firm is performing well at ESG, it may instead be safer to assume it is doing poorly. Therefore, one could also consider the imputation of missing ESG scores with a value of -0.5, the worst value possible on our transformed scale. We will see in Section 4.4 that imputing a zero, where it matters at all, provides *conservative* results relative to imputing -0.5.

Although the ESG screens discussed above are implemented without imputing ESG values, note they are equivalent to screens in data where we have imputed. If a missing value is imputed to be zero, then the firm will be zeroed-out by a positive screen if $p_{screen} > 50$ and zeroed-out by a negative screen if $p_{screen} \geq 50$. If a missing value is imputed to be -0.5, then the firm will be zeroed-out in any negative or positive screen.

We include a constant as an instrument alongside the firm characteristics and use the five-factor model as our benchmark, following Kelly et al. (2019). For ease of interpretation, we rescale all portfolios' annualized volatility to 10%. Of course, this has no effect on t-statistics; it serves to put the various portfolios' means on the same footing, and allows the reader to easily see the annualized Sharpe ratio via dividing the mean return by 10. We present t-statistics for mean returns from Newey and West (1987) with three lags—the three lags are for robustness, and little is changed using instead White (1980) standard errors—and in the text report Sharpe ratio t-statistics coming from Lo (2003).

All of our results come from in-sample estimation. Kelly et al. (2019) and Kelly et al. (forthcoming) have shown that IPCA parameters are quite stable due to the great deal of dimension reduction employed, with limited impact of in- versus out-of-sample estimation. Moreover, out-of-sample analysis would be hindered by the short sample of some of the ESG measures we consider. Finally, our focus is really on the comparative static exercise of including versus ignoring ESG scores: in-sample results give ESG measures the best chance of providing predictive information, giving ESG information a lower bar for significance.

4 Results

In this section, we present our main results. These benchmark results reflect several empirical choices. First, given that Figure 2 shows that ESG measures are most widely available for large firms, we focus our attention on results for firms above the NYSE median equity market capitalization each month. Second, we use the median p_{50} ESG score to be our screening threshold. Third, where necessary, we impute missing ESG values as zero (the median). Fourth, we use each data provider's topline ESG score. Later in this section, we discuss the robustness of the results with respect to changes to percentile thresholds, different target ESG levels for the responsible investing portfolios, alternate measures of systematic risk, and varied information sets. Across the analyses, we focus our attention on results from the more recent 2010–2020 sample period.

4.1 Screened portfolios

In Table I we report annualized mean returns for the unadjusted and negatively-screened tangency portfolios, as well as (non-annualized) skewness and excess kurtosis. In the case of negatively-screened portfolios, we view two sample periods (2000–2020 and 2010–2020) as informative to report. After all, the negatively-screened tangency portfolios could represent what savvy investors achieve when they optimally weight assets *until* an ESG mandate tells then to zero-out a position. The Asset4, KLD, MSCI, RepRisk, and Trucost measures begin prior to 2010, so we report performance based on these scores for the 2000-2020 period in Panel A. Panel B reports the performance for 2010–2020, with the addition of the S&P and Sustainalytics measures since all seven measures are observed after 2010.[23] Though in Table I we start chronologically with results for 2000-2020, we do this only for completeness. We view the more recent 2010-2020 sample period of much more importance because ESG interest has increased substantially in recent years (Giglio et al., forthcoming).

In Table I, the top line of Panel A reports that the tangency portfolio has a highly significant annualized mean return of 16.68% ($t = 7.47$) on the full 2000–2020 sample. This implies a Sharpe ratio of 1.67, which itself has a significant t-statistic of 2.15. Therefore, the tangency portfolio is quite profitable with statistically- and economically significant average returns.[24] For comparison, over this sample, the volatility-scaled market portfolio delivers an annualized 4.16% mean return ($t = 1.84$), and hence a Sharpe ratio of 0.42. Thus, these tangency portfolios provide economically- and statistically significant improvements over what a passive market strategy earns. There is a bit of negative skewness and excess kurtosis, as the

[23] But note the caveat, mentioned above and below, that S&P data is missing a lot.

[24] This number is comparable but modestly below the performance Kelly et al. (2019) report for their large-firm sub-sample for the period 1970–2014.

Table I
Negatively-screened tangency portfolios

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, for tangency- and negatively-screened portfolio returns. Panel A reports results from the 2000–2020 sample period, while Panel B uses the more recent 2010–2020 sample period. The screening threshold is the median p_{50}, and the ESG measure is the topline ESG score. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Panel A: 2000–2020				
Tangency	16.68	(7.47)	−0.46	2.09
	Panel I: Screened, long-and-short			
Asset4	15.96	(6.52)	−0.28	2.54
KLD	17.04	(7.33)	−0.04	1.38
MSCI	15.98	(7.08)	−0.22	1.89
RepRisk	16.42	(7.07)	−0.24	2.62
Trucost	17.14	(7.17)	−0.10	2.60
	Panel II: Screened, long-only			
Asset4	12.96	(5.42)	0.11	0.44
KLD	13.11	(5.43)	0.04	0.73
MSCI	14.46	(6.24)	−0.03	0.86
RepRisk	14.19	(6.03)	0.08	0.54
Trucost	12.85	(5.32)	0.16	0.38
Panel B: 2010–2020				
Tangency	19.97	(7.18)	−0.28	0.78
	Panel I: Screened, long-and-short			
Asset4	19.47	(6.89)	−0.39	0.62
KLD	18.31	(6.93)	−0.43	0.11
MSCI	20.24	(7.62)	−0.33	0.40
RepRisk	22.18	(7.87)	−0.35	0.67
S&P	20.35	(7.15)	−0.31	0.73
Sustainalytics	20.89	(7.58)	−0.31	0.54
Trucost	17.49	(6.77)	−0.54	0.56
	Panel II: Screened, long-only			
Asset4	14.64	(4.51)	−0.26	1.27
KLD	16.12	(4.96)	−0.16	1.61
MSCI	15.89	(5.11)	−0.26	1.41
RepRisk	14.35	(4.56)	−0.22	1.18
S&P	19.17	(6.47)	−0.34	0.93
Sustainalytics	16.76	(5.37)	−0.31	1.61
Trucost	14.34	(4.62)	−0.22	1.46

market portfolio also exhibits, because this sample period includes the Great Recession of 2007–2009.

Panel A.I shows that negative screens of both long and short positions have little effect on portfolios' mean returns across several ESG score providers, which hover around 16–17%. For KLD and Trucost, these statistics actually increase, although insignificantly so.[25] Hence, we see that a variety of negative screens have small effects on systematic portfolio performance—ESG investing objectives exhibit low or no cost.

Panel A.II shows that negatively-screening only the long leg can have larger effects. The mean returns hover around 13–14%—so no screened portfolio has statistically-significantly different mean returns than the tangency portfolio, though the economic significance is nontrivial.[26] Overall, while the effects in Panel A.II are mildly larger than in A.I, the take-away remains that ESG investing broadly has modest costs.

Turning to the more recent sample period, the top line of Panel B reports that the tangency portfolio obtains a higher mean return of about 19.98% ($t = 7.18$). Some of this improvement admittedly comes from the prevailing bull market: but the tangency portfolios doubles what the volatility-scaled market delivers—an annualized 9.59% mean return ($t = 3.74$).[27]

Of course, our research question is on the impact of implementing ESG screens, and the 2010–2020 sample does not qualitatively change the answer. Panel B.I shows that little is changed by negatively screening long-and-short positions, using most ESG providers, with mean returns hovering around 17–22%. In fact, screening with MSCI, RepRisk, S&P or Sustainalytics leads to improvements, though they are statistically insignificant. The impact of screening using Trucost scores is a bit more notable, as the mean return falls to 17.49%,

[25] This is possible because we are not optimizing under an ESG constraint.

[26] Meanwhile, the excess kurtosis drops and skewness increases, sometimes becoming positive. For completeness we continue to report skewness and kurtosis so the reader can see that nothing drastic changes, but do not focus our main discussion on them.

[27] The excess kurtosis, of these portfolios and indeed the market, drop on the more recent period because we exclude the Great Recession.

Table II
Positively-screened tangency portfolios

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, for tangency- and positively-screened portfolio returns. The sample period is 2010–2020. The screening threshold is the median p_{50}, and the ESG measure is the topline ESG score. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Tangency	19.97	(7.18)	−0.28	0.78
Panel A: Screened				
Asset4	15.29	(5.94)	−0.45	0.00
KLD	16.43	(6.44)	−0.28	0.19
MSCI	16.77	(6.19)	−0.38	0.58
RepRisk	17.84	(6.00)	−0.09	0.41
S&P	7.75	(2.81)	−0.39	2.22
Sustainalytics	14.19	(5.57)	−0.44	−0.11
Trucost	14.60	(5.75)	−0.56	0.90

but the decrease is statistically insignificant. Panel B.II shows that long-only screens lead to larger reductions. The decreases for the RepRisk and Trucost screens are statistically significant (t-statistics just above 2) and for the Asset4 screen is marginally so, all about 5.5 percentage points less than the tangency portfolio's. Thus, long-only screening has larger effects overall.

Turning now to positively-screened systematic portfolios, recall that these can only include firms with observed ESG information. Therefore, the full sample results are not so relevant, in our view, because ESG information was so sparse before 2010. Accordingly, we focus on the 2010–2020 sample that is more salient for positively-screened portfolios, as well as the responsible-investing models and ESG-integration results described further below. To help the reader, we repeat the tangency portfolio results in the following tables even though it is identical to the top line of Table I Panel B.

Table II shows that positive screens reduce mean returns by more moderate amounts, to range 14–18% across all but one ESG provider. The drops in Asset4-, Sustainalytics-, and Trucost screened portfolios are economically, if not statistically, significant. For the S&P

score, the reduction is significant, huge, and falls below the market average return—however this is primarily driven by the more modest coverage in the S&P ESG data and makes the obvious point that positive screens are problematic when the criteria are largely missing.

Tables I and II suggest that optimal portfolios can be negatively screened to achieve a reasonable ESG mandate, with little reduction in average returns or overall profitability. However, the particular ESG criteria *does* matter. Negatively screening long-only positions has larger effects, adding a return-based distinction for this type of screen that adds to Pedersen et al. (2020)'s point that overall ESG-performance can be hindered by long-only screens. Positive screens have broadly larger effects still, which can be statistically significant. Nevertheless, there exist several approaches that implement an ESG mandate and achieve mean returns that are economically- and statistically identical to the unadjusted portfolio: ESG investing can cost nothing.

4.2 Responsible-investing models

We next consider the performance of optimal portfolios coming from the responsible-investing models of Pedersen et al. (2020) and Pastor et al. (2021). For the former, we take $\bar{s}$=0.25: this means that the portfolio-weighted average ESG score is the 75^{th} percentile. For the latter, we take $d = 0.001$: this means that the ESG-induced spread in effective expected return between the best and worst ESG stock is 1.2% per annum.[28] For these responsible-investing models, as well as the ESG-integration results discussed in the next subsection, we require nonmissing ESG scores. We impute values of zero, i.e. the median. In the top line of Table III we report results for the Markowitz portfolio implied by our estimated model, which differs only slightly relative to the the tangency portfolio's performance.[29]

[28] That is, if the best and worst ESG stock had the same expected return apart from ESG information, the *effective* expected return used by the responsible-investing objective would be 10 basis points per month higher for the best ESG stock.

[29] We later show that the Markowitz portfolio, and hence each responsible-investing portfolio, implies unreasonably large and variable weights, whereas the tangency portfolio does not.

Table III
Responsible-investing optimal portfolios

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, for responsible-investing models' optimal portfolio returns. The sample period is 2010–2020. The ESG measure is the topline ESG score, and we impute a value of zero for all missing ESG observations. In the "Markowitz" line we report performance of the Markowitz portfolio implied by our factor model estimates. For the Pedersen et al. (2020) model we set $\bar{s} = 0.25$; for the Pastor et al. (2021) model we set $d = 0.001$. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Markowitz	20.85	(7.66)	−0.32	0.17
Panel A: Pedersen et al. (2020)				
Asset4	20.42	(7.52)	−0.33	0.17
KLD	20.53	(7.56)	−0.27	0.10
MSCI	20.40	(7.43)	−0.31	0.22
RepRisk	20.91	(7.75)	−0.30	0.17
S&P	20.86	(5.53)	−0.17	−0.08
Sustainalytics	20.32	(7.33)	−0.32	0.13
Trucost	20.79	(7.63)	−0.23	0.02
Panel B: Pastor et al. (2021)				
Asset4	15.83	(5.75)	−0.29	0.23
KLD	17.54	(5.96)	0.05	−0.13
MSCI	18.33	(6.33)	−0.23	0.27
RepRisk	19.79	(7.76)	−0.22	0.14
S&P	19.98	(7.30)	−0.39	0.04
Sustainalytics	16.63	(5.99)	−0.22	−0.07
Trucost	16.16	(5.57)	0.06	−0.22

Table III Panel A reports that the Pedersen et al. (2020) optimal portfolios' performance is virtually unchanged from the tangency's performance. These results obtain even with an imposed average ESG score at the 75^{th} percentile, which we shall later see is higher than the tangency portfolio's average ESG performance. Notably, this is true across all the ESG data providers.

Panel B shows, instead, that different ESG scores lead to very different outcomes across the optimal portfolios from Pastor et al. (2021)'s responsible-investing model. For RepRisk and S&P, the mean returns are virtually unchanged from those of the tangency's. But MSCI scores result in a drop to 18.3%, KLD scores a drop to 17.5%, Sustainalytics scores a drop to 16.6%, Trucost scores a drop to 16.2%, and Asset4 scores a drop to 15.8%. These are increasingly economically, if not statistically, significant.

Therefore, responsible-investing models *can* deliver improved ESG performance without sacrificing returns, but the particulars matter. The Pedersen et al. (2020) model portfolios have mean returns that are virtually unaffected by the choice of ESG data provider. On the other hand, the Pastor et al. (2021) model portfolios are more sensitive to exactly which ESG score is used. These results are qualitatively similar to the screened-portfolio results.

4.3 ESG integration

Now we integrate ESG information into the estimation of our model and the ensuing portfolio creation. As discussed, we implement this in two ways. First, we include ESG information amongst other characteristics in our estimation of beta, and ask whether this significantly enhances the tangency portfolio's performance. Second, we include ESG information in our estimation of alpha to test whether this creates a significantly profitable (beta-neutral) portfolio that is orthogonal to factor risk.

Table IV Panel A reports that integrating ESG measures into beta has virtually no effect on estimates of aggregate risk. Those ESG-integrated tangency portfolios' mean returns are indistinguishable from the tangency portfolio without ESG information. This is even true when we integrate *all the measures* into beta, as the last line of Panel A shows. The take-away is: ESG scores tell us nothing about systematic risk that is not already known in the other firm characteristics.

Turning to beta-neutral portfolios, the line above Panel B labeled "Tangency" reports what *should* be the beta-neutral portfolio's performance if expected returns are the result of only aggregate risk: a zero average return. Panel B shows that individual ESG measures do not uniformly define profitable non-systematic strategies—but distinctions emerge. The

Table IV
ESG integration

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, from IPCA models that include ESG information. The sample period is 2010–2020. The ESG measure is the topline ESG score, and we impute a value of zero for all missing ESG observations. Panel A includes the ESG measure as a characteristic instrumenting beta and reports the tangency portfolio's performance. Panel B uses the ESG measure in alpha and reports the beta-neutral portfolio's performance. Above Panel A in the "Tangency" line is the performance of the tangency portfolio; above Panel B in the "Tangency" line is the theoretical performance of the beta-neutral portfolio if the asset-pricing model were true and expected returns are the result of only aggregate risk exposure. In Panels A and B, the line labeled "All" uses all seven measures in beta or alpha, respectively. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Tangency	19.97	(7.18)	−0.28	0.78
Panel A: ESG-integrated tangency				
Asset4	20.15	(7.25)	−0.26	0.77
KLD	19.99	(7.18)	−0.28	0.80
MSCI	19.99	(7.18)	−0.28	0.78
RepRisk	20.02	(7.21)	−0.28	0.77
S&P	19.95	(7.18)	−0.28	0.78
Sustainalytics	20.07	(7.21)	−0.27	0.79
Trucost	20.23	(7.20)	−0.16	0.61
All	20.22	(7.23)	−0.19	0.58
Tangency	0			
Panel B: Beta-neutral				
Asset4	2.55	(0.96)	−0.36	0.17
KLD	4.73	(1.34)	0.26	−0.18
MSCI	3.88	(1.13)	0.86	1.08
RepRisk	2.41	(0.88)	0.03	0.44
S&P	0.19	(0.06)	−0.45	2.75
Sustainalytics	2.68	(0.87)	−0.01	0.16
Trucost	5.70	(1.89)	0.41	−0.16
All	6.70	(2.17)	0.26	−0.32

beta-netural portfolios for almost all providers have statistically insignificant mean returns. However,the Trucost beta-neutral portfolio yields a marginally significant annualized mean return of 5.70% ($t = 1.89$).

Notably, integrating all seven ESG topline scores creates a beta-neutral portfolio with a significant average return. The last line of Panel B shows that the mean return is 6.70%, significant at the 5% level ($t = 2.17$). Therefore, we find that ESG information *can collectively* reveal significant non-systematic returns.[30] Admittedly, market returns over this period were high and the estimated alpha is less than what the market earned over the period. However, the beta-neutral portfolio is constructed to be orthogonal to aggregate risk, so this result reflects a surprisingly anomalous average return coming from ESG information.

4.4 Robustness

In this subsection, we perform specification analysis for our results. We present variations for each of the screening, responsible-investing models, and ESG-integration approaches. Further, we explicitly connect our results to the recent related literature by considering the ability of ESG sub-components (E dimension, Trucost GHG emissions) and a combination of ESG scores (an IV approach) to predict returns in our conditional framework.

4.4.1 Screened portfolios

Table V evaluates the screened-portfolio results' robustness to changing the screening threshold to the 75th percentile p_{75}, instead of the median as before, on the 2010–2020 sample. Table V Panel A is remarkably similar to Table I Panel B.I for most ESG providers, showing that negative-screening is little affected by the increased ESG threshold, when both long and short positions are screened.

In Table V Panel B, there is a much larger effect of negatively screening at the higher p_{75} threshold when applied only to long positions. All measures except S&P see an economically and statistically large drop in average returns. But the result for S&P is driven by missing values that imply the screen has done very little—so we take little signal from the S&P result. Of more interest are differences between providers like Trucost and MSCI, whose coverage

[30]If we increase the number of latent factors to ten, the beta-neutral average return is 7.79% ($t = 2.45$)—hence the result does not come from restricting the dimension of the factor space.

Table V
Robustness: screened tangency portfolios

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, for tangency- , negatively-screened, and positively-screened portfolio returns. Uses the 2010–2020 sample period. The screening threshold is the 75th percentile p_{75}, and the ESG measure is the topline ESG score. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Tangency	19.97	(7.18)	−0.28	0.78
Panel A: Negatively screened, long-and-short				
Asset4	19.62	(6.56)	−0.20	0.40
KLD	15.81	(5.91)	−0.35	0.19
MSCI	19.70	(7.32)	−0.21	0.73
RepRisk	22.82	(7.88)	−0.57	1.60
S&P	20.58	(7.18)	−0.20	0.51
Sustainalytics	21.26	(7.59)	−0.31	1.18
Trucost	18.29	(6.39)	−0.70	0.90
Panel B: Negatively screened, long-only				
Asset4	7.57	(2.34)	−0.13	0.48
KLD	7.58	(2.26)	0.01	0.20
MSCI	10.89	(3.44)	−0.21	0.51
RepRisk	9.18	(2.98)	−0.15	0.29
S&P	17.67	(5.86)	−0.37	0.68
Sustainalytics	12.24	(3.83)	−0.24	0.89
Trucost	5.42	(1.76)	−0.01	0.35
Panel C: Positively screened				
Asset4	13.57	(5.19)	−0.55	0.28
KLD	15.97	(6.37)	−0.18	0.25
MSCI	14.73	(5.63)	−0.53	1.13
RepRisk	11.01	(3.15)	−0.61	7.33
S&P	7.00	(2.63)	−0.44	2.35
Sustainalytics	13.01	(5.13)	−0.52	0.26
Trucost	13.11	(4.63)	−0.79	0.93

is reasonably similar: the former now has an insignificant average return of 5.42% ($t = 1.76$) while the latter's average return of 10.89% stays significant ($t = 3.44$). Turning now to the positive screens in Panel C, we see a divergence in the effect of the higher threshold. All the providers see lower returns, but the RepRisk screens are reduced by a larger amount, from 17.8% to 11%. On the whole, Panel C echoes the conclusions of Tables I and II that positive screens can be more costly than negative screens.

Table V says that negatively-screened portfolios can continue to deliver competitive returns even at a higher screening threshold. But the particulars definitely matter, as some providers or long-only screening have much larger effects. Positive screens on average have larger deleterious effects from the higher threshold.

The online appendix reports further analysis that we summarize here. Lowering the threshold to p_{25} does little for long-and-short negative screens, but broadly improves the long-only negative- and positive screen results. Using a best-in-class (industry adjusted) measure does not change the story much.[31] Using the E, S, or G subcomponent indices with the median threshold also leads to only small effects.

4.4.2 Responsible-investing optimal portfolios

Table VI considers the robustness of our results for responsible-investing models. For Pedersen et al. (2020)'s model, we change the portfolio's average ESG score $\bar{s}$ to 0.4, the 90th percentile of ESG performance. Panel A shows that this has very little effect on performance (comparing to Table III Panel A). Hence, altering $\bar{s}$ has only small effects on the portfolio performance: why? The reason stems from Pedersen et al. (2020)'s equilibrium value of π.[32] In our empirical exercise, π has a very small absolute value across different $\bar{s}$ choices. For instance, using Trucost data the average π for $\bar{s} = 0.25$ is 0.0003, while for $\bar{s} = 0.40$ it is 0.0004. Thus, w_{PFP_t} from (7) differs little from $\Sigma^{-1}\mu$ across choices of $\bar{s}$. This feature of the Pedersen et al. (2020) model allows responsible-investing portfolios to attain various targets with minimal deviations from the Markowitz portfolio weights, hence the very similar performance.

Table VI shows that Pastor et al. (2021)'s optimal portfolios are quite sensitive to a higher ESG preference parameter—Panel B doubles d to 0.002, meaning that the ESG-induced

[31] We do not do this for MSCI because its topline score is already industry-adjusted.

[32] Given by $(c_{1,\mu}\bar{s} - c_{s,\mu})/(c_{ss} - 2c_{1,s}\bar{s} + c_{11}\bar{s}^2)$ for $c_{1,\mu} = \iota'\Sigma^{-1}\mu$, $c_{s,\mu} = s'\Sigma^{-1}\mu$, $c_{ss} = s'\Sigma^{-1}s$, $c_{1,s} = \iota'\Sigma^{-1}s$, and $c_{11} = \iota'\Sigma^{-1}\iota$.

Table VI
Robustness: responsible-investing optimal portfolios

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, for responsible-investing models' optimal portfolio returns. The sample period is 2010–2020. The ESG measure is the topline ESG score, and we impute a value of zero for all missing ESG observations. For the Pedersen et al. (2020) model we set $\bar{s} = 0.40$; for the Pastor et al. (2021) model we set $d = 0.002$. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Markowitz	20.85	(7.66)	−0.32	0.17
Panel A: Pedersen et al. (2020)				
Asset4	19.93	(7.31)	−0.32	0.19
KLD	20.25	(7.38)	−0.22	0.08
MSCI	20.01	(7.22)	−0.29	0.26
RepRisk	20.86	(7.76)	−0.29	0.17
S&P	20.30	(5.32)	−0.15	−0.10
Sustainalytics	19.69	(7.05)	−0.30	0.12
Trucost	20.55	(7.49)	−0.17	−0.00
Panel B: Pastor et al. (2021)				
Asset4	10.12	(3.54)	−0.28	0.60
KLD	13.43	(4.25)	0.17	−0.10
MSCI	15.08	(4.89)	−0.03	0.26
RepRisk	16.52	(6.54)	−0.13	0.23
S&P	17.80	(6.39)	−0.36	−0.07
Sustainalytics	11.45	(3.93)	−0.14	0.03
Trucost	11.78	(3.93)	0.15	−0.25

spread in effective expected return between the best and worst ESG stock is 2.4% per annum. Changes to the portfolio using several of the ESG provider metrics now lead to significant drops in optimal portfolio performance, especially for Asset4, KLD, Sustainalytics, and Trucost.[33]

4.4.3 ESG integration

Finally, we explore the robustness of our results on ESG integration in the IPCA model. We first consider if ESG information gives us significant incremental information on beta,

[33] In the online appendix, we report negligible differences from imputing −0.5 instead of zero, or using E, S, or G subcomponent indices.

relative to other firm characteristics. Given the strong similarity amongst the lines of Table IV Panel A, we only report results for all the measures included together. There are two broad directions we take in Table VII: exploring the impact of alternate measures of aggregate risks, and exploring the impact of altering the set of non-ESG information. The top lines of Table VII report tangency portfolio results without ESG information. The first line labeled "Tangency" is identical to what we have seen before. The following lines are new. The "FF5C" line uses a six-factor model combining Fama and French (2015) and Carhart (1997), and reports results from the tangency portfolio coming from instrumented exposures thereupon.[34] The label "slow non-ESG" indicates that we reduce the non-ESG information set to the "slow" firm characteristics with low time-series volatility listed in Section 2.1.

Prior to including ESG information, the six FF5C factors with instrumented betas yields an annualized mean return of 17.64%, about two percentage points below what what the five latent factors deliver. If we instead estimate latent-factor IPCA using slow non-ESG information, the tangency portfolio's mean return is 17.72%; the instrumented-FF5C betas deliver a mean return of 15.52%. These results make a couple points. First, the latent factor model is closer to the mean-variance-efficient frontier than the FF5C factor model, thus a superior depiction of aggregate risk, if not by much on this sample. Second, a reduction of instrumenting information has effects that are visible but modest.

Table VII Panel A now includes all topline ESG measures into estimation of beta. We saw in Table IV Panel A that this led to no change when using the full instrumental information set. Comparing the lines labeled "Tangency, slow non-ESG", we see the conclusion holds as well for the smaller instrumental information set: there is only a slight insignificant improvement. Using the FF5C factor model instead changes nothing qualitative; this is visible by comparing the "FF5C" line pair and the "FF5C, slow non-ESG" line pair. Thus, we find a great deal of

[34]That is, we use Mkt-RF, SMB, HML, CMW, RMA, and MOM as the factors. But the result we report is different than the tangency portfolio of the FF5C factors themselves (which could not integrate ESG information). On the 2010–2020 sample, that turns out to yield a volatility-scaled annualized mean return of 13.88%, which is lower than what Table VII reports is obtained using instrumented exposures.

Table VII
Robustness: ESG integration

Notes – Annualized mean, skewness, and excess kurtosis of the monthly returns, from IPCA models that include ESG information. The sample period is 2010–2020. Unless otherwise stated, the ESG measure is the topline ESG score, and we impute a value of zero for all missing ESG observations. Panel A includes the all the ESG measures as characteristics instrumenting beta and reports the tangency portfolio's performance. Panel B uses the ESG measure in alpha and reports the beta-neutral portfolio's performance. "Tangency" means that five latent factors have been estimated. "FF5C" is the Fama and French (2015) and Carhart (1997) six-factor model, while "FF3C" is the Carhart (1997) four-factor model. The label "slow non-ESG" means that non-ESG characteristics are restricted to the slow (low time-series volatility) characteristics listed in Section 2.1. "E" means we use the environmental subcomponent index instead of the topline index. "All" means all seven data providers, "All*" excludes RepRisk, "All+" excludes Trucost, and "All*+" excludes RepRisk and Trucost; "E*" means we use the E subcomponent for the five data providers except Trucost and RepRisk, the former for whom we take their topline score as previously defined in Section 2.2; "MSCI IV" and "MSCI IV, E" use the MSCI score as instrumented by the KLD score, for the topline and E subcomponent index, respectively. In parentheses are t-statistics for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

ESG measure	Mean		Skewness	Kurtosis
Tangency	19.97	(7.18)	−0.28	0.78
Tangency, slow non-ESG	17.72	(6.99)	−0.43	0.21
FF5C	17.64	(6.40)	−0.15	−0.46
FF5C, slow non-ESG	15.52	(5.69)	−0.34	−0.10
Panel A: ESG-integrated tangency				
Tangency, slow non-ESG	18.22	(7.02)	−0.38	0.27
FF5C	17.77	(6.40)	−0.14	−0.47
FF5C, slow non-ESG	15.83	(5.76)	−0.34	−0.19
Tangency	0			
Panel B: Beta-neutral				
Asset4, E	1.73	(0.65)	−0.31	0.05
KLD, E	4.55	(1.47)	0.56	0.68
MSCI, E	7.49	(2.19)	0.11	0.99
S&P, E	1.56	(0.58)	0.40	2.63
Sustainalytics, E	0.84	(0.31)	−0.19	−0.27
MSCI, E, FF3C	6.76	(1.90)	0.13	0.80
All*+, E	9.34	(2.71)	0.19	0.71
All*, E*	8.71	(2.62)	0.26	−0.25
MSCI IV	6.50	(2.00)	0.29	−0.42
MSCI IV, E	7.76	(2.41)	0.01	−0.30
All+, S	6.69	(2.04)	0.05	−0.46
All+, G	5.74	(1.85)	−0.01	−0.20

robustness to this result: ESG measures do not give significant information about systematic risk exposures.[35]

Panel B turns to the question of whether ESG information instead defines profitable mispricing over our sample. This further allows us to directly connect our results to the recent literature, which considers the ability of E (MSCI) scores (Pastor et al., 2022) and attempts at reducing noise in ESG scores (Berg et al., 2021) for predicting returns. We extend these findings by integrating them in our explicitly conditional asset pricing framework.

Specifically, we consider extensions of the analysis presented in Panel B of Table IV, which showed that ESG information *collectively* and Trucost GHG emissions provide significant information about conditional mispricing. Primarily we ask: does the topline ESG score obfuscate useful information that is present in one of the reported subcomponent indices? At this point, it is useful to consider a qualitative difference that exists between Trucost and the other data providers. Trucost measures carbon emissions which are, of course, an environmental object. Since Trucost had the highest beta-neutral return in Table IV, it is natural to ask if this resulted from its distinction as a focused E measure.

To this end, we calculate beta-neutral portfolios using the other ESG providers' E component measure.[36] The lines in Panel B with "E" in the label provide the results. Remarkably, the only ESG data provider for whom this has an significant impact is MSCI. In contrast to the topline score, the MSCI environmental score defines a beta-neutral portfolio with an economically and statistically significant return of 7.49% per annum ($t = 2.19$)—larger than what Trucost delivered in Table IV. The results say that specific environmental scores can be used to construct a profitable portfolio that is orthogonal to systematic risk.

In fact, this resembles what Pastor et al. (2022) find: a portfolio constructed using MSCI's E index delivers a significant alpha with respect to the Carhart (1997) four-factor model.

[35]For the ESG-integrated systematic portfolios, we find virtually no effect from the different specifications detailed next for beta-neutral portfolios—so we refrain from a discussion.

[36]As mentioned before, RepRisk does not have an E subcomponent index.

On a sample period of November 2012 to December 2020, just a bit shorter than our 2010–2020 sample, Pastor et al. (2022) find an annualized alpha of 5.66% ($t = 2.14$): we call this alpha *unconditional* because it comes from a regression of portfolio returns on factors with static slopes, which is of course standard practice. Kelly et al. (2019) contrast these with conditional alphas, coming from a model like IPCA that estimates conditional betas. To more closely connect with Pastor et al. (2022), in line "MSCI, E, FF3C" we use the Carhart (1997) four factors and find a conditional alpha of 6.76% per annum that is significant at the 10% level. Therefore, our results enhance Pastor et al. (2022)'s unconditional alpha result, in fact finding a *larger* conditional alpha.

Going further, are there gains from combining the environmental scores? Yes. In line "All*+, E" we combine the five environmental index scores from the providers other than RepRisk and Trucost and obtain a beta-neutral annual average return of 9.34% ($t = 2.71$). If we combine those five environmental subcomponents with Trucost's topline score, in line "All*, E*" we see a similar average return of 8.71% ($t = 2.62$).[37]

There are gains from combining ESG information—this aligns with the point in Berg et al. (2021), who use ESG scores as instrumental variables for one another and find stronger return predictability. To more closely connect to their work, we now consider using an IV-version of the MSCI topline and E indices to define a beta-neutral portfolio. For these, we run a first-stage regression of the MSCI score on the KLD scores of the same type (i.e. topline and E).[38] In the line "MSCI IV" we find that an approach like Berg et al. (2021) delivers improvements—the beta-neutral average return is almost doubled to 6.5% ($t = 2.00$) and is now significant. In the line "MSCI IV, E" we see that the instrumental variables approach adds to the raw MSCI E index, as the beta-neutral average return is a little bigger.

Taken altogether, Tables IV and VII provide evidence from a conditional framework that

[37]As in footnote 30, for both of these results we checked against a ten factor model, and the average returns were barely changed—the mispricing appears robust to the larger factor-space dimension.

[38]We choose KLD simply because it and MSCI are often nonmissing for the same firm-months; a more sophisticated pruned-IV approach like Berg et al. (2021)'s may yield further improvements.

specific ESG information can provide significant beta-neutral profits. The MSCI E score (and to a lesser extent the Trucost score) does so, which supports Pastor et al. (2022)'s results—other E scores do not. Pooling topline ESG or E scores can also work well—either via a α_t specification with $L_\zeta > 1$, or by a instrumental variables approach for MSCI indices following Berg et al. (2021). Therefore, we essentially report return predictability along the lines of what Berg et al. (2021) found, and enhance their finding by showing it results from mispricing with respect to a conditional asset-pricing model.

How about other subcomponents? The online appendix reports that using individual S or G subcomponent does not produce significant conditional alpha. However, combining the S subcomponents gives a 6.69% annualized mean return that is just significant ($t = 2.04$), and the combined G subcomponents is marginally significant ($t = 1.85$) at 5.74%.[39]

The online appendix reports the effects of instead imputing -0.5 (embodying the assumption that missing ESG information indicates *bad* ESG performance): now the KLD and MSCI toplines give marginally significant average returns, and the combination is half a percentage point greater at 7.20% ($t = 2.29$) than the combination before in Table IV. For the E subcomponent index, the -0.5 is even more impactful. Now MSCI E defines a beta-neutral return of 9.12% ($t = 2.61$), about 1.5 percentage points greater than before in Table VII, and only a little below the 9.34% ($t = 2.71$) average return of the combination of the E subcomponents (excluding Trucost). Broadly speaking, we view this as suggesting the results in Table VII are somewhat conservative, especially for E subcomponents. The alternate -0.5 imputation has small results for S and G or their combinations.

Therefore, we find a special role in alpha for E subcomponent indices. Several individual E subcomponents create profitable beta-neutral portfolios on their own; combining the E scores leads to further gains. Meanwhile, S and G subcomponent indices do not deliver significant alpha on their own, but combining the S information yields a significant average return.

[39] Trucost is excluded from these combinations.

5 Discussion

We elaborate on our main results in two ways. First, we report some salient properties of tilted systematic portfolios that deliver an ESG mandate with modest effects on return performance, thus providing more detail on trade-offs between ESG implementation and return performance. Second, we reconcile the profitable mispricing results from ESG-integration with the screening implications that ESG tilts do not significantly alter portfolio performance.

5.1 Properties of ESG-tilted portfolios

In the following discussion we focus on the benchmark-specification portfolios from Section 4 that employ the MSCI topline ESG score. In particular, we show results from the negatively-screened tangency portfolios using the median p_{50} threshold, and the Pastor et al. (2021) optimal portfolio using the ESG taste parameter $d = 0.001$: the long-and-short screened portfolio has an average return of 20.2%, the long-only screened portfolio an average return of 15.9%, and the responsible-investing portfolio an average return of 18.3% (Tables I and III). We consider properties of the portfolio weights themselves, as well as the portfolio-weighted average ESG score $\bar{s}$ (where missing ESG is imputed to be zero). Recall that s has been rank-demeaned to live on the interval $[-0.5, 0.5]$: therefore $\bar{s}$ values can be interpreted as percentiles of the cross-sectional ESG score distribution, and to ease exposition in this section we report it in those terms.

Figure 3 Panel A shows the sum of portfolio weights over time.[40] Looking first at the tangency portfolio weight (gray solid line), we see that this total portfolio weight consistently hovers around 0.9, meaning that an investor invests 90% of their weight in these large stocks. The long-and-short negatively-screened portfolio weight (black solid line) and Pastor et al. (2021)

[40]Of course, this is driven by our assumptions about risk aversion and volatility scaling.

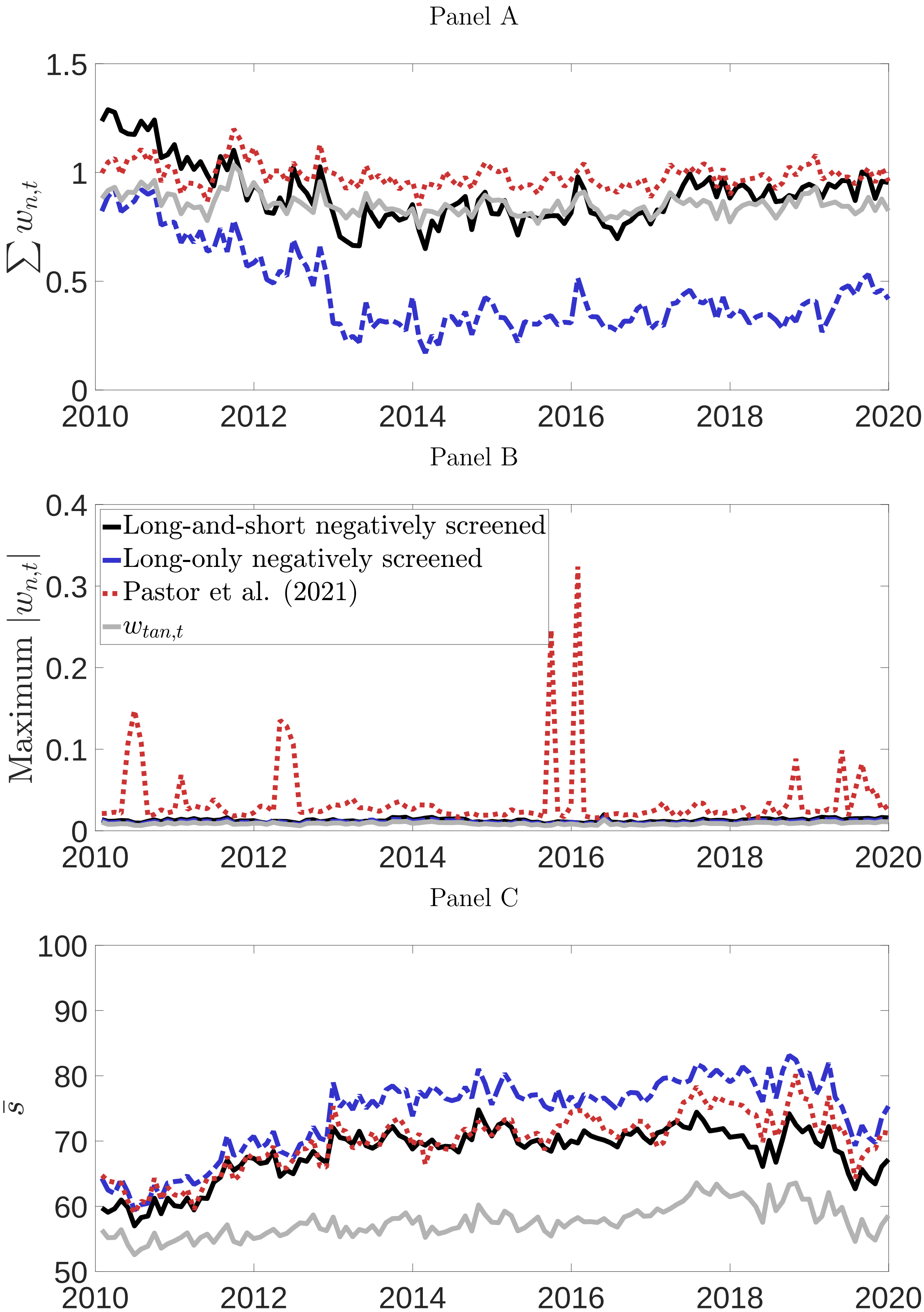

Figure 3: Portfolio weights and ESG performance

Notes – Time-series of portfolio properties for chosen portfolios using the MSCI topline score and median p_{50} threshold or $d = 0.001$ ESG taste parameter. Panel A reports the sum of portfolio weights each period. Panel B reports the maximum absolute weight each period. Panel C reports the portfolio-weighted average ESG score where missing ESG is imputed to be zero, translated to the cross-sectional percentile for exposition. The solid black line is for the negatively-screened, long-and-short, portfolio; blue dash-dotted line for the negatively-screen, long-only, portfolio; red dotted line for the Pastor et al. (2021) optimal portfolio; and, the gray solid line for the unadjusted tangency portfolio. Reproduced by permission of MSCI Research LLC portfolio weight (red dotted line) also hover around 90–100%. Meanwhile, the long-only negatively-screened portfolio weight (blue dash-dotted line) hovers around the much lower 40%, which makes sense given that we have zeroed out only long positions. By similar logic, note that the long-only negatively-screened weight is always below both the long-and-short and original tangency portfolio weights, as it must be.[41] But in the first few years, when MSCI coverage was lower, the long-only portfolio weight is only a little below that of the tangency's, while the long-and-short portfolio weight is higher, indicating that early in the sample bad ESG performance tended to be associated with lower expected return stocks.

Panel B shows the maximum absolute weight on any single stock. Strikingly, the Pastor et al. (2021) portfolio can sometimes place very large weight on single names: the maximum weight magnitude shoots above 10% several times, and once to above 30%! This is due to the fact that the Pastor et al. (2021) portfolio tilts the Markowitz portfolio, which itself (not shown) also exhibits these large weights. Meanwhile, the tangency portfolio and its screened versions never place more than 2% of wealth in a single stock, demonstrating an advantage of the dimension-reduction naturally employed in the estimated factor portfolios and betas thereupon.

Panel C shows the portfolio-weighted average ESG score $\bar{s}$, a key measure of the portfolio's overall *ESG* performance.[42] Once MSCI coverage expands in 2013, we see that all the tilted portfolios exhibit substantially better ESG performance than the unadjusted tangency portfolio. The latter has an average $\bar{s}$ of 58 (58th percentile). The long-and-short negatively-screened portfolio is always higher with an average 68, and the Pastor et al. (2021) optimal portfolio is nearby with an average $\bar{s}$ of 70. The long-only negatively-screened portfolio is higher still with an average $\bar{s}$ of 75.

A few observations are in order. First, the unadjusted tangency portfolio's ESG performance

[41] Technically, if they all had the same volatility; in this empirical exercise the volatilities are similar.

[42] Indeed, this is exactly an input into the Pedersen et al. (2020) model: hence its $\bar{s}$ (not shown) is constant at a chosen level, by design.

is surprisingly middle-of-the road: its profitable positions are not associated with bad-ESG stocks and so its performance is above-median.[43] Second, the nature of Pastor et al. (2021)'s model ensures that its ESG performance is improved by the taste for ESG: its better ESG performance does not come as a surprise, but instead is a feature of that portfolio design.

Finally, Figure 3 Panel C explains in more detail the trade-offs that exist between responsible-investing mandates of different types. The long-only screen has modestly better ESG performance (75 versus 68)—so in this case we do not see materialized the possibility that shorting leads to better ESG performance, pointed out by Pedersen et al. (2020). Meanwhile, recall that the long-and-short screen achieved a 20.2% average return while the long-only screen achieved a 15.9%. Furthermore, there is no real difference in the ESG performance of the long-and-short and Pastor et al. (2021) portfolios, but the latter obtains a lower 18.3% average return. Hence, we see meaningful trade-offs between ESG- and return performance; additionally, the specifics by which ESG performance is improved matters too. One can imagine many managers for whom the long-only screen would not be worth the sacrificed return for the gained ESG performance, but of course any individual investor's choice in the matter would depend on the strength of their ESG preference. Nevertheless, the point is that Figure 3 enriches the story coming from Table I: the choice of long-and-short screening delivers a return benefit for an ESG cost. It is potentially useful for future research to explore the trade-offs between return performance, ESG goals, and portfolio-tilting specifications.

5.2 ESG disagreement

We have shown that we can use ESG measures to tilt well-performing portfolios without a significant reduction in performance. The tilts downweight bad-ESG stocks to achieve an ESG-investing mandate. But if every investor does this, what is the equilibrium effect? Won't the stock's price fall, expected return rise, and ESG begin to predict returns?

[43] Of course, well-performing firms might have increased resources to devote to ESG-related efforts, and institutions might demand coverage for particular firms.

For the preponderance of data considered, it turns out the answer is no. The extensiveness of our empirical analyses provides a hint as to how this can be the case—there is no *single* way to "do ESG." We are not the first to document that different ESG measures disagree. Among others, Berg et al. (2022), Avramov et al. (2021), Christensen et al. (2021), and Gibson et al. (2021) document relatively low correlations across the ESG ratings from different data providers and examine the implications for stock returns and ESG-alpha. Furthermore, even within the same data provider, the top-level ESG score is an aggregation of several subcomponents that investors may weight differently. In addition, there could be differences between whether or not one industry-adjusts or otherwise normalizes scores. Hence, even if all ESG investors in the market were to use the same ESG data source, one might still observe significant differences in how they implement ESG mandates.

If different investors use different ESG measures or implementations to invest, what are the equilibrium expected returns? Pastor et al. (2021) provides an answer.[44] In this theory, investor i forms the portfolio

$$w_{i,PST} = \Sigma^{-1}(\mu + d_i \tilde{g}_i)$$

using return moments μ, Σ, the scalar ESG taste parameter $d_i \geq 0$, and the agent-specific ESG-measure vector $\tilde{g}_i$.[45] If $\tilde{g}_i = 0$, then the investment problem is completely unaffected by ESG concerns and the investor does not distinguish firms' ESG scores from one another. Regardless of what g_i are, market clearing implies their equation 6

$$\mu = \Sigma w_{mkt,PST} - \bar{d}g$$

where $w_{mkt,PST}$ is the market portfolio, $\bar{d} = \int_i \omega_i d_i di$ is a wealth-weighted average of the

[44] See their footnote 4 and its proof in their appendix.
[45] We set relative risk aversion to 1 and abstract from t subscripts, for simplicity.

non-negative ESG tastes d_i across agents using wealth-weights ω_i, and

$$g = (1/\bar{d}) \int_i \omega_i d_i \tilde{g}_i di$$

is a wealth- and ESG-taste-weighted average of the investors' ESG measures $\tilde{g}_i$. If we just had $\mu = \Sigma w_{mkt,PST}$ then we would be in the ordinary CAPM world and ESG tastes would not affect expected returns. Clearly $\bar{d} \geq 0$ with inequality if any wealth-weighted mass of agents have positive ESG taste. Therefore, a generic way for expected returns to be unaffected by ESG concerns, even if agents have them, is if $g = 0$.

Pastor et al. (2021) provide the decomposition

$$g = E_\omega(\tilde{g}_i) + Cov_\omega(d_i/\bar{d}, \tilde{g}_i)$$

using wealth-weighted expectation and covariance, and note it is plausible to assume the covariance is zero. Obviously if every $\tilde{g}_i = 0$ then $E_\omega(\tilde{g}_i) = 0$, but this is not necessary. Instead suppose that every investor perceives differences between firms' ESG scores and invests accordingly, but that their $\tilde{g}_i$ differ. If $E_\omega(\tilde{g}_i) = 0$, we are saying that the wealth-weighted *average* ESG score does not distinguish between firms. In this case, equilibrium expected returns are unaffected by ESG concerns, even when all agents have them.

Empirical support for this mechanism comes from comparing what different ESG measures and providers say about the same firms. A simple way to make this comparison is to consider the rank correlation between measures. A correlation of 1 tells us that the two measures completely agree on firms' ESG ranking. On the other hand, a correlation of 0 tells us that the two measures' rankings have no relationship to one another, as though their agreement is random. Each month, we calculate the rank correlations and fit a kernel density to these correlations to clean up things pictorially.[46] Figure 4 shows a surface plot of these densities

[46] We use a Gaussian kernel estimated on 100 equally-spaced points on $[-1, 1]$.

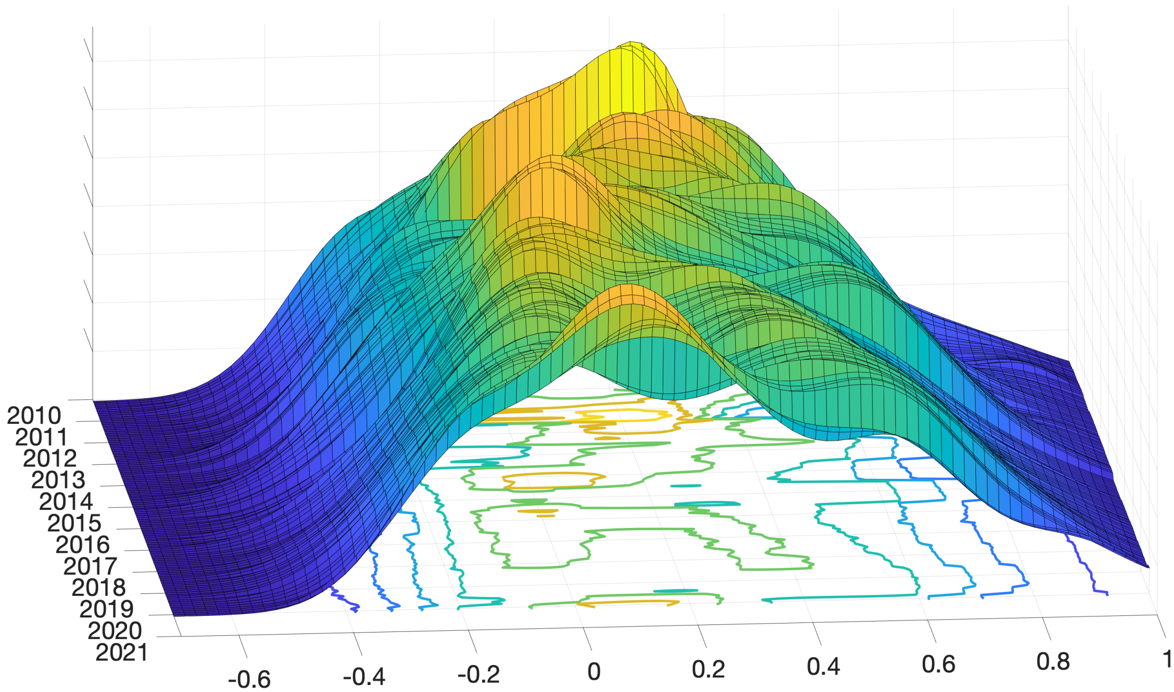

Figure 4: Densities of cross-sectional rank correlations

Notes – Cross-sectional kernel density estimates of ESG measures' rank correlations.

over the 2010–2020 sample period. The most striking feature of this three-dimensional surface is how the peak hovers near 0.[47]

Does this imply that available ESG information is noise? Our beta-neutral portfolio results provide evidence against this interpretation. From Table IV Panel B, we observe weakly significant alphas for the Trucost measure and significant alpha when information across providers is combined. Trucost is perhaps the most objectively quantifiable of our ESG measures, providing information on green house gasses scaled by revenue. Insofar as investors may differ in interpretations for what is important in ESG implementation, it is intuitive that emissions might prove both focal and unambiguously measurable.

[47]This is additionally visible in the bottom plane of the figure where a contour plot shows that yellow and green contours (the highest density values) straddle zero.

This supposition is underscored in Table VII, which provides beta-neutral results for the 'E' score for each provider and various ways of combining information across providers for E, S, and G separately. The observation that combining all information together for the subcomponent indices reliably produces (at least weakly) significant alphas is evidence of mispricing for each pillar category over the last decade. We see that the E component of MSCI, one of the most widely used providers, also exhibits mispricing.

The evidence is broadly consistent with investors using multiple inputs to their ESG approaches. Indeed, from conversations with corporate CFOs, it is clear that firms communicate with large shareholders about their particular preferences for ESG disclosures. Moreover, the SustainAbility survey of large institutional managers indicates that most investors employ their own scoring methodologies using underlying data from multiple providers, rather than relying on topline information. Therefore, there is a large variety of different ways to measure and implement ESG, and they do not agree—this can imply that equilibrium expected returns need not be affected by investors' ESG preferences.

Separately, there are further related issues involving ESG implementation. For instance, Brandon et al. (2021) find that institutional investors in the US do not have better ESG scores even when they say they take ESG into account. This reality could add a layer of cheap talk wherein investors need not commit to acting on stated ESG goals—yet another mechanism by which stated ESG preferences could fail to affect equilibrium prices. Why might investors wish to highlight their "responsible" portfolio construction, regardless of actually acting upon it? Riedl and Smeets (2017) and Bauer et al. (2021) find that social preferences and social signaling explain ESG adoption, not financial considerations. In fact, their results say that investors expect lower returns and higher management fees, and are thus willing to forgo financial performance—hence asset managers may wish to signal ESG concerns to attract a clientele with lower fee-price elasticity. Consistent with this, Hartzmark and Sussman (2019) find that sustainability causes outflows from low-sustainability (in our context, think bad-ESG) funds, and inflows to high-sustainability funds—and of course, increases in assets-under-management increase fee revenue.

Thus, it seems clear that professional portfolio managers have incentives to advertise good ESG performance.[48] At the same time, there is no definitive rule for how to measure ESG characteristics—as opposed to, say, accounting information produced under generally-accepted accounting principles and subject to regulation by the SEC. It seems natural that in such an environment one might *expect* many ESG methods and measure-providers to flourish, and perhaps exist to cater to different portfolio-managers whose underlying pre-ESG weights are different.[49] Broadly speaking, a lack of uniform consensus (or regulation) on definitive ESG measurement is a straightforward channel explaining the low cost of ESG investing for much of our paper.

6 Conclusion

Using a conditional factor model with a wealth of firm information to examine the return predictability for a host of ESG characteristics, we find that systematic portfolio weights can be adjusted to improve the portfolio's ESG performance without sacrificing profits. Naturally, eliminating "bad" ESG firms from an optimal portfolio is less costly than selecting only "good" firms given that not all firms are covered. Results from recent taste-oriented responsible-investing models are qualitatively similar, though the cost to performance can vary with the choice of information provider and model.

When we integrate ESG information in the conditional factor model, we find that ESG measures do not provide independently significant information about a firm's (financial) risk

[48]And, more recently, even bad: multiple anti-ESG funds have emerged, underscoring values-catering by fund managers.

[49]From an investor protection perspective, there may be a greater impetus for ESG transparency in advertised retail-facing funds—but it isn't clear that convergence across data providers is optimal, and more research is called for.

exposures. And, when we constrain ESG information to be orthogonal to aggregate risk, we find evidence that some information can earn alpha, and particularly so when information is combined.

Overall, these findings highlight a degree of disagreement in ESG measures and investor focus. In the model of Pastor et al. (2021), ESG implementation can have little to no cost for investors if ESG-minded investors each measure ESG performance differently. We provide evidence that common ESG measures disagree and also note that multiple providers offer a sufficiently rich set of component information such that investors may focus on different information.

It is, therefore, unclear if disagreement in ESG scoring is problematic, as noted in speeches by policymakers at the SEC[50] and the UN Climate Change Conference (COP26). If market-competition or regulation increases the coordination of ESG measurement and implementation, theory suggests that ESG scores would begin to significantly predict returns. Nevertheless, making that empirical statement will require controlling for the wealth of other conditioning information that is available to investors, just as in this paper.

[50]See for example the *ESG Subcommittee Update Report* to the SEC Asset Management Advisory Committee, May 27, 2020.

References

Aragon, G. O., Y. Jiang, J. Joenväärä, and C. I. Tiu (2020): "Socially Responsible Investments: Costs and Benefits for University Endowment Funds," Working Paper, Available at SSRN 3446252.

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli (2021): "Sustainable investing with ESG rating uncertainty," *Journal of Financial Economics*.

Baker, M., D. Bergstresser, G. Serafiem, and J. Wurgler (2018): "Financing the response to climate change: the pricing and ownership of U.S. green bonds," NBER Working Paper Series.

Bauer, R., T. Ruof, and P. Smeets (2021): "Get real! Individuals prefer more sustainable investments," *The Review of Financial Studies*, 34, 3976–4043.

Berg, F., K. Fabisik, and Z. Sautner (2020): "Rewriting history II: The (un) predictable past of ESG ratings," ECGI Working Paper.

Berg, F., J. F. Kölbel, A. Pavlova, and R. Rigobon (2021): "ESG Confusion and Stock Returns: Tackling the Problem of Noise," Working Paper, Available at SSRN 3941514.

Berg, F., J. F. Kölbel, and R. Rigobon (2022): "Aggregate Confusion: The Divergence of ESG Ratings*," *Review of Finance*, rfac033.

Bialkowski, J. and L. T. Starks (2016): "SRI Funds: Investor Demand, Exogenous Shocks and ESG Profiles," Working Paper, University of Texas at Austin and University of Canterbury.

Bolton, P. and M. Kacperczyk (2020): "Do investors care about carbon risk?" NBER Working Paper Series.

Brandon, R. G., S. Glossner, P. Krueger, P. Matos, and T. Steffen (2021): "Do Responsible Investors Invest Responsibly?" ECGI Working Paper.

Carhart, M. M. (1997): "On the Persistence of Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Ceccarelli, M., S. Ramelli, and A. F. Wagner (2021): "Low-carbon mutual funds," Swiss Finance Institute Research Paper.

Chini, E. and M. Rubin (2022): "Time-varying Environmental Betas and Latent Green Factors," Working Paper, EDHC Business School.

Christensen, D. M., G. Serafeim, and A. Sikochi (2021): "Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings," *The Accounting Review*, tAR-2019-0506.

Di Giuli, A. and L. Kostovetsky (2014): "Are red or blue companies more likely to go green? Politics and corporate social responsibility," *Journal of Financial Economics*, 111, 158–180.

Dimson, E., P. Marsh, and M. Staunton (2020): "Exclusionary screening," *The Journal of Impact and ESG Investing*, 1, 66–75.

Dyck, A., K. V. Lins, L. Roth, and H. F. Wagner (2019): "Do institutional investors drive corporate social responsibility? International evidence," *Journal of Financial Economics*, 131, 693–714.

Edmans, A. (2011): "Does the stock market fully value intangibles? Employee satisfaction and equity prices," *Journal of Financial Economics*, 101, 621–640.

Fabozzi, F. J., K. C. Ma, and B. J. Oliphant (2008): "Sin Stock Returns," *Journal of Portfolio Management*, 35, 82–94.

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

——— (2007): "Disagreement, tastes, and asset prices," *Journal of Financial Economics*, 83, 667–689.

——— (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Gibson, R., P. Krueger, and P. S. Schmidt (2021): "ESG Rating Disagreement and Stock Returns," *Financial Analysts Journal*, 77, 104–127.

Giglio, S., B. Kelly, and J. Stroebel (forthcoming): "Climate Finance," *Annual Review of Financial Economics*.

Glossner, S. (2021): "Repeat Offenders: ESG Incident Recidivism and Investor Underreacton," Working Paper, University of Viriginia.

Gorgen, M., A. Jacob, M. Nerlinger, R. Riordan, M. Rohleder, and M. Wilkens (2020): "Carbon Risk," Working Paper, Unversity of Augsburg, Queen's University.

Hartzmark, S. M. and A. B. Sussman (2019): "Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows," *Journal of Finance*, 74, 2789–2837.

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

Hou, K., C. Xue, and L. Zhang (2015): "Digesting Anomalies: An Investment Approach," *Review of Financial Studies*, 28, 650–705.

Huang, W., G. A. Karolyi, and A. Kwan (2021): "Paying Attention to ESG Matters," Working Paper.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (forthcoming): "Is there a replication crisis in finance?" *Journal of Finance*.

Kelly, B. T., T. Moskowitz, and S. Pruitt (2021): "Understanding Momentum and Reversal," *Journal of Financial Economics*, 140, 726–743.

Kelly, B. T., D. Palhares, and S. Pruitt (forthcoming): "Modeling Corporate Bond Returns," *Journal of Finance*.

Kelly, B. T., S. Pruitt, and Y. Su (2019): "Characteristics are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 134, 501–524.

——— (2020): "Instrumented Principal Components Analysis," Working Paper, Yale University, Arizona State University, and Johns Hopkins University.

Kim, S. and A. Yoon (2020): "Analyzing Active Managers' Commitment to ESG: Evidence from United Nations Principles for Responsible Investment," Working Paper, Available at SSRN 3555984.

Lo, A. W. (2003): "The Statistics of Sharpe Ratios," *Financial Analysts Journal*, 58, 36–52.

Luo, H. A. and R. J. Balvers (2017): "Social Screens and Systematic Investor Boycott Risk," *Journal of Financial and Quantitative Analysis*, 52, 365–399.

Newey, W. and K. West (1987): "A Simple Positive Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 8, 703–705.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2021): "Sustainable investing in equilibrium," *Journal of Financial Economics*, 142, 550–571.

——— (2022): "Dissecting Green Returns," Working Paper, University of Chicago and University of Pennsylvania.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski (2020): "Responsible investing: The ESG-efficient frontier," *Journal of Financial Economics*, 142.

Riedl, A. and P. Smeets (2017): "Why Do Investors Hold Socially Responsible Mutual Funds?" *Journal of Finance*, 72, 2505–2549.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–38.

Zerbib, O. D. (2019): "The effect of pro-environmental preferences on bond prices: Evidence from green bonds," *Journal of Banking & Finance*, 98, 39–60.

——— (2020): "A Sustainable Capital Asset Pricing Model (S-CAPM): Evidence from Green Investing and Sin Stock Exclusion," Working Paper, Boston University.