The Cost of ESG Investing

Laura Lindsey
Arizona State University

Seth Pruitt
Arizona State University

Christoph Schiller
Arizona State University

November 18, 2021

Abstract

Even against increasing interest in socially responsible investing mandates, we find that implementing ESG strategies can cost nothing. Modifying optimal portfolio weights to achieve an ESG-investing tilt negligibly affects portfolio performance across a broad range of ESG measures and thresholds. This is because those ESG measures do not provide information about future stock performance, either in relation to risk or mispricing, beyond what is provided by other observable firm characteristics. That the stock market does not reflect significant equilibrium pricing of ESG information is rationalized in a model of responsible investing wherein investors differ in selection of ESG-related criteria to weight their portfolios.

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

1 Introduction

Over the past two decades, the amount of investment linked to 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 of doing so on portfolio efficiency and performance are unclear. Economic theory generally argues that, all else equal, high-ESG firms should have lower expected returns as socially-oriented investors require less compensation for holding high-ESG firms (e.g. Fama and French, 2007; Pedersen et al., 2020; Pastor et al., 2021b). While Fabozzi et al. (2008), Hong and Kacperczyk (2009), Bolton and Kacperczyk (2020), and Pastor et al. (2021a) empirically document higher risk-adjusted returns for "sin" stocks and high-carbon-emissions firms, Edmans (2011) and Glossner (2021), among others, find that firms with higher employee satisfaction and fewer ESG-related controversies 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).

The aim of this paper is to study the costs of implementing an ESG-investing mandate, and additionally investigate whether or not ESG ratings identify systematic risk exposures or exploitable mispricing. In contrast to previous research, which has primarily relied on long-short portfolio sorts following Fama and French (1993), we consider ESG ratings as firm characteristics 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 our estimates of firms' risk exposures, and also detect what predictive information is present in ESG scores that is distinct from other firm characteristics. Our contribution is twofold: first, we demonstrate that the cost of following an ESG mandate is potentially quite low, as portfolios with ESG screens perform similarly to the unconstrained tangency portfolio implied by our model. This finding can be explained by our second result: ESG measures from commonly-used data providers neither provide information about systematic risk exposure nor exploitable mispricing.

First, we use scores from any of four major ESG data providers to tilt optimal portfolios toward satisfying a range of ESG mandates.[1] We construct the tilts in two steps: in step one we use the other valuable firm information to define a profitable portfolio; in step two we tilt the portfolio to downweight bad-ESG firms (and possibly upweight good-ESG firms). Some of these tilts are simple screens that drop bad-ESG firms, while others are optimal portfolios from responsible-investing models in Pedersen et al. (2020) and Pastor et al. (2021b). We primarily focus on ESG tilts corresponding to negative and exclusionary screening, as this form of ESG investment approach represents the most common ESG mandate globally (Dimson et al., 2020).[2] In all the tilts we consider, the pre-tilt portfolio comes from the conditional factor model and is systematic.[3] Many of these portfolios can be overlaid with ESG mandates without substantially affecting profitability. Therefore we show that the cost of ESG investing can be quite low. Pedersen et al. (2020) states that "many investors 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"— our results show that this hope can be realized.

We next broaden our focus to ask: how can ESG tilts have zero cost? To answer this, we investigate the role of ESG characteristics in determining either alpha or beta. We begin by including ESG measures along with other firm characteristics and estimate instrumented betas on aggregate factors via IPCA. We find no evidence that ESG scores drive factor exposures.[4] Next, we allow the ESG characteristics to instead drive alpha, defined as predictable returns that are orthogonal to aggregate risk exposures, either on their own or alongside other characteristics: regardless, we find no evidence of significant profits. Therefore, we conclude that ESG measures do not give significant information about either systematic risk-exposures or exploitable mispricing. This result closely matches the main message in Pastor et al. (2021a), who document that green-returns have occurred mostly due to random shocks to ESG concerns and find little scope for an ESG premium in expected returns. Our conditional-model results strengthen this conclusion, suggesting that green-returns were not significantly different than what one would have predicted on the basis of non-ESG firm

[1]Data employed include KLD (now MSCI ESG KLD STATS), Asset4 (now Refinitiv ESG), Sustainalytics (now Morningstar), and RepRisk.

[2]Consequently, our results are less closely related to the recent literature on ESG impact and activist investing (see e.g. Dimson et al., 2015, 2021; Hoepner et al., 2021).

[3]More specifically: the screens are based on the tangency portfolios implied by the model. The Pedersen et al. (2020) and Pastor et al. (2021b) portfolios tilt the Markowitz portfolio, and we take the covariance matrix and mean vector from the model estimates that impose there is no alpha.

[4]Our focus is on exposures to aggregate risk, and it is there that we find no role for ESG information. Therefore our results can coexist with the results of Engle et al. (2020), who find a significant role for hedging specific climate risks.

information.

An important aspect of our work is the large extent of ESG information we consider. Our main results are robust to using data from different ESG providers and ESG index subcomponents. They are robust to different subsamples, for instance focusing only on the recent sample where ESG coverage is richer and ESG concerns, perhaps, have become more salient. The results are similar whether or not we industry-adjust the measures. The results are similar for various choices of missing-value imputation, or when restricted to firms with nonmissing ESG data. And while different tilting criteria (e.g. what constitutes a good- or bad-ESG firm, what is the desired ESG level of the portfolio, etc.) can drive significant changes to portfolio performance (as one would expect!), there is a wide range of reasonable choices that yield minimal deviations from the profitability of non-tilted portfolios.

How can investors care about the ESG performance of firms, tilt their portfolios to reflect ESG scores, 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. (2021b) and propose a simple solution: investors use varying ESG measures. As noted above, our empirical analysis is quite extensive across different ESG data providers and particular ESG scores, and our main results are robust across them all. This is a testament to the fact that there are many ways to "do ESG". If investors do not agree on a single definition of ESG measurement, equilibrium pricing need not reflect their ESG concerns even if all investors act on them. We provide empirical support that ESG measures disagree, supporting the explanation.[5] This part of our analysis is closely aligned with recent research documenting substantial disagreements across ESG data providers and the resulting asset pricing implications (Berg et al., 2020b; Avramov et al., 2021; Christensen et al., 2021; Gibson et al., 2021).

Despite extensive research, there is widespread disagreement in the literature on the return predictability of ESG characteristics. The lack of return predictability in our model 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 longer literature on so-called "sin" stocks has found a premium for firms in industries like alcohol or tobacco.[6] In a similar vein, Zerbib (2020) uses the holdings of "green" institutional investors to show that excluded firms have a significantly

[5] An alternate interpretation that the measures available to researchers are simply noise.

[6] 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.

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) find a positive "greenium" for green bonds relative to similar non-green municipal 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.

Our paper contributes 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) and Kelly et al. (forthcoming). It is crucial to have the best-possible depiction of systematic risks when we evaluate how firms' differing 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, and this pertains even when we use the same factor model 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 empirical impact. Third, we take into account a large set of other firm characteristics, meaning that we control for a substantial amount of the conditioning information investors have at their disposal *already* in addition to ESG scores. Fourth, we use data from four major ESG providers (and evaluate both aggregate and subcomponent performance) in our empirical analysis, making our conclusions broad.

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 factor model estimation, how systematic and non-systematic strategies are formed, and how ESG-mandate tilts are implemented. Section 4 presents the results when ESG is included in estimation or instead used as a tilt, shows robustness of those results, discusses properties of some ESG-tilted portfolios, connects to the existing empirical literature, and discusses our results within the context of an equilibrium model. Section 5 concludes. The Online Appendix includes additional results and robustness tests.

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 attention to a subset of these that are "slow", 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. Table 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.

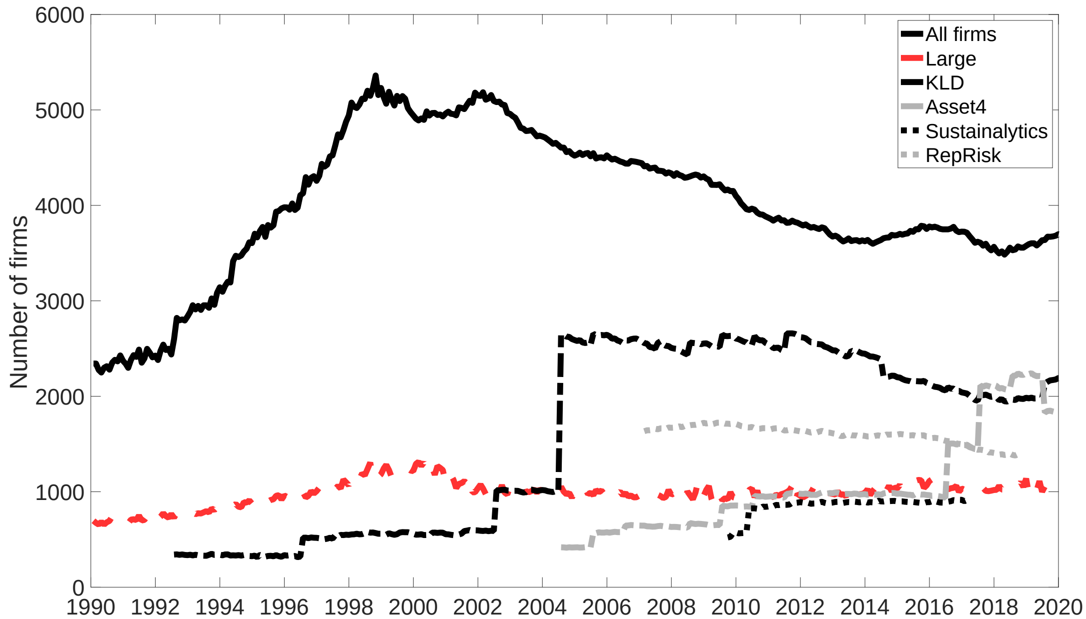

Figure 1: Available observations

The number of large firm observations is plotted in Table 1 as the dashed red line.

2.2 ESG characteristics

We obtain data on firm-level ESG scores from four major data providers commonly used by investors and in the academic literature (see e.g. Berg et al., 2020b; Huang et al., 2021). Our first data source is MSCI ESG KLD STATS (KLD), which is available from 1992 to 2018. For each covered firm, KLD evaluates "strengths" and "concerns" across the following six dimensions of ESG: environmental impact, community relations, product characteristics, employee relations, diversity, and governance. KLD has changed its methodology and underlying data items several times since 1992. We follow Akey et al. (2021) in accounting for these changes and construct time-consistent scores by identifying data items that covered the same issues but changed names over time and retaining only data items with continuous coverage since 1992. The score for each category and the overall score are then calculated as the number of strengths minus the number of concerns. We summarize community relations, product characteristics, employee relations, and diversity as the "social" category, as is standard in the literature.

Second, we construct ESG scores using data from Asset4 (now Refinitiv ESG).[7] Asset4 coverage starts in 2003 and includes ESG information based on over 250 key performance indicators and over 750 individual data points, across three pillars: 'E' (emissions, resource use, product innovation), 'S' (product responsibility, community, human rights, diversity and opportunity, employment quality, health & safety, training and development) and 'G' (board functions, board structure, compensation policy, vision and strategy, shareholder rights). The ESG 'pillar scores' reported by Asset4 are constructed by comparing firms' ESG measures to peers and weighting by materiality. To avoid comparability issues and put our ESG scores on equal footing, we instead construct E, S, and G scores and an overall ESG score by aggregating over the corresponding raw data items (equal-weighting) as in Dyck et al. (2019).[8] This also helps us address concerns about changes in the Asset4 data aggregation methodology throughout our sample period (Berg et al., 2020a).

Third, we obtain ESG scores from Sustainalytics (now 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 overall ESG score for the sample period from 2009 to 2017.

Fourth, we obtain ESG data from RepRisk, which is available to us for the sample period from 2007 to 2019. RepRisk analysts 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. 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.

The ESG measures from the four data providers are reported at different frequencies. In particular, KLD and Asset4 ESG scores are reported annually, while 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 definitively time. In fact, this issue is quite similar

[7]Our Asset4 data was downloaded in 2019 and is therefore unaffected by the recent changes due to backfilling documented by Berg et al. (2020a).

[8]Following Dyck et al. (2019), for questions where a "yes" answer is associated with better ESG performance (positive direction), we translate the Y/N items into 0 (N) and 1 (Y), and the answers to double Y/N questions into 0 (NN), 0.5 (YN or NY), and 1 (YY). For questions with a negative direction we use a reversed coding scheme. We normalize numerical data items to be between 0 and 1.

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 KLD or Asset4 score 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 (black dashed-dot line) are available starting in 1992 for a small number of firms, with noticeable increases in coverage in 1996, 2002, and particularly 2004. Asset4 measures (gray dashed-dot line) start in 2004, again for a small number of firms, and with a noticeable increase in coverage in 2016. RepRisk (gray dotted line) starts in 2007 with relatively large coverage that gradually declines over time. Sustainalytics (black dotted line) starts in 2009 with a small number of firms, bumps up in 2010, and remains steady. 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 available firms 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 in Panel A. For about the first ten 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 increases in coverage in 1996 and 2002 do little to change these facts, but 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 the percentiles of all firms, showing us the ESG coverage is better for larger firms.[9]

[9]The top gray and blue dotted lines always lie on top of each other, for all providers. This says that all

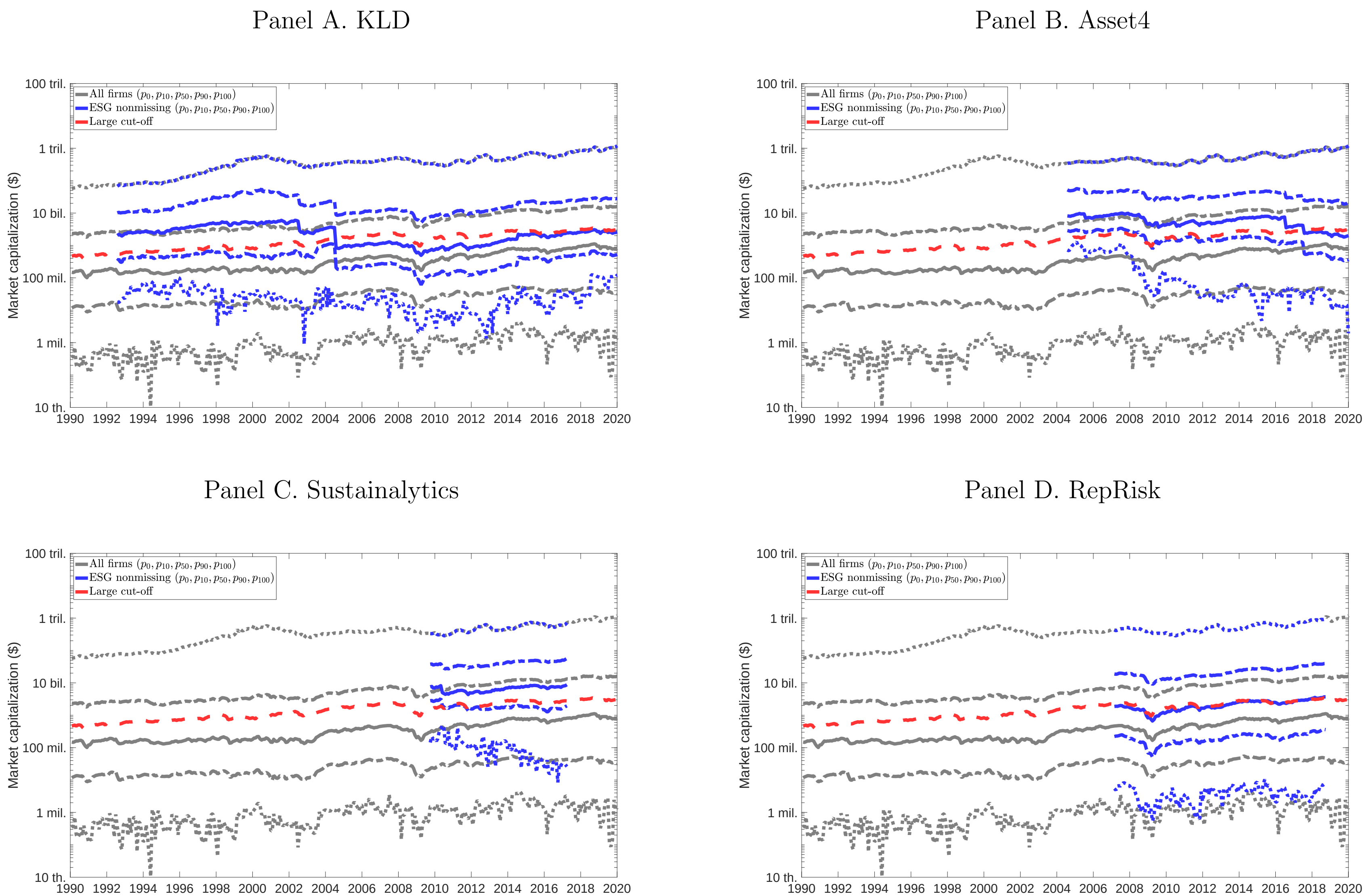

Figure 2: Firm size and ESG availability

The remaining panels show that this feature is also true 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 above the NYSE median for the most part. 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. And for RepRisk, half of its firms are above the NYSE median and its distribution is consistently skewed towards large firms.

The broad takeaway here is that firms which 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 results 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. Nevertheless, sensitivity analysis will confirm that our main results are robust to instead using the panel of all firms.

3 Model and Portfolio Construction

In this section we briefly describe IPCA and model-implied investment strategies. We then discuss various ways of using ESG ratings within the IPCA model, as well as using model-implied ESG implementation strategies as a tilt.

3.1 Basic and modified IPCA models

The basic IPCA model is

$$r_{n,t+1} = \alpha_{n,t} + \beta_{n,t}' f_{t+1} + \varepsilon_{n,t+1}, \quad \text{where } \alpha_{n,t} = \Gamma_\alpha' z_{n,t} \text{ and } \beta_{n,t} = \Gamma_\beta' z_{n,t}, \tag{1}$$

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. Following Kelly et al. (2019), we refer to (1) as an *unrestricted* model, and a *restricted* model is one where we impose $\Gamma_\alpha = 0$. 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

providers have always covered the largest firms in our data.

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, so we can concisely state the first-order conditions upon which one iterates until convergence to the least-squares estimates:

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

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

where for the restricted model we impose $\Gamma_\alpha = 0$ and define $\tilde{\Gamma} \equiv \Gamma_\beta$ and $\tilde{f}_{t+1} \equiv f_{t+1}$, while for the unrestricted model we instead allow a nonzero Γ_α and define $\tilde{\Gamma} \equiv [\Gamma_\alpha, \Gamma_\beta]$ and $\tilde{f}_{t+1} \equiv \left[1, f'_{t+1}\right]'$.[10] Following Kelly et al. (2019), in the unrestricted model we impose the identification assumption $\Gamma'_\alpha \Gamma_\beta = 0$ meaning that risk loadings "explain as much of the asset's mean returns as possible". Then $\Gamma_\alpha \neq 0$ means that there is a predictable part of returns unrelated to aggregate risk, which we refer to as mispricing.

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.[11] 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 a return's *covariance with aggregate risk*. Hence, we construct a *systematic investment* when we base the portfolio weights on $\beta_{n,t}$ even when characteristics are part of the picture.

On the other hand, when we allow $\Gamma_\alpha \neq 0$ then we are admitting a non-systematic investment strategy to be formed. This is due to two reasons. First, we follow Kelly et al. (2019) and impose that Γ_α and Γ_β are orthogonal: thus, $\alpha_{n,t}$ contains characteristic information that is orthogonal to systematic exposure. Second, the timing of (1) means that $\alpha_{n,t}$ is known before the return $r_{n,t+1}$. Therefore, $\alpha_{n,t}$ represents an anomaly: a predictable return that is not compensation to aggregate risk. Kelly et al. (2019) find that Γ_α is statistically insignificant in stocks (and Kelly et al., forthcoming, finds the same in bonds), therefore argue that the restricted model best fits the data. But, those papers do not include ESG measures or consider an ESG-investing mandate. Moreover, using the Jensen et al. (forthcoming) data

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

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

allows us to look at stock performance in recent years (the data in Kelly et al., 2019, ends in 2014), which could be important if ESG-investing's salience has increased over time.

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 $N \times K$ matrix β_t, and the $K \times N$ 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):

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

3.3 Using ESG as a tilt

There are several terms we could use to convey this section's main idea: tilt, overlay, adjustment, screen etc. The main idea is that ESG measures are not used in the estimated model, but instead to achieve an ESG mandate. In all of the ESG tilts we consider, we constrain our attention to systematic portfolios.

Our first method of implementing an ESG mandate is to screen firms based on ESG scores using a simple two-step approach. First, we find the tangency portfolio weights utilizing all non-ESG-related information via IPCA. Second, we tilt the tangency-portfolio weights using an ESG measure—downweighting stocks with "bad" ESG information, and retaining stocks with "good" ESG information. The result is final portfolio weights. By doing the two steps separately, we accomplish two goals. First, we use profit maximizing weights that are familiar and easy to describe—see Section 3.2. This is useful because we want to use transparent machinery so as to make the ESG-investing costs clear. Second, we can consider a variety of choices for step two and see how they impact profitability while achieving an ESG-investing mandate.

When we tilt the investment strategy using ESG scores, we confront the reality that the scores can be missing for many firms. We consider two main ways of dealing with this issue. First, we include all of the available firms and adjust weights only for those with nonmissing ESG measures that satisfy some criteria. Alternatively, we restrict attention only to firms with nonmissing ESG measures. In addition, we must take a stand on what defines an acceptable level of ESG performance. In our empirical analyses, we consider different percentile cut-offs (p_{75}, p_{50}, p_{25}) to gauge the effects of more- or less-stringent ESG mandates.

Two of our tilts include all of the available firms. In the first screen we exclude any firm with an unacceptable ESG measure, defined as below the stated percentile. Of course, this requires that we actually observe a firm's ESG score, and therefore any firm with missing ESG is unaffected. This tilt accomplishes the mandate: Avoid taking any position in firms with unacceptable ESG measures. The second screen is similar to the first, but further only excludes stocks that would otherwise be in the long-leg of the portfolio. This tilt accomplishes the mandate: Avoid going long in firms with unacceptable ESG measures.

The third screen restricts attention only to firms with nonmissing ESG measures. We only include the firm in the portfolio if its ESG score is acceptable, defined as equal-to or above the stated percentile. Hence, both firms with unacceptable ESG scores and firms with missing ESG scores are excluded. This tilt accomplishes the mandate: Take positions only in firms with acceptable ESG measures.

In addition to screened tangency portfolios, we construct the optimal portfolios derived by Pedersen et al. (2020) and Pastor et al. (2021b) (for which we use the shorthand PFP and PST, respectively)—these take into account firms' expected returns, covariances, and ESG information. They are similar to the tilts we just defined, because they can be seen as tilts to the usual Markowitz weights. 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.[12] 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. (2021b) derive optimal weights as

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

where the scalar $d \geq 0$ is the investor's "ESG taste."[13] Assume for the moment that "bad" ESG is denoted by negative values in s: the PST weight reduces the *effective* expected return for bad-ESG stocks. While (7) and (8) are clearly similar, some differences emerge when we use them in Section 4.4 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. We use the factor decomposition of stocks' covariance matrix, for simplicity assuming the idiosyncratic return covariance matrix Σ_ε is diagonal: hence, $\Sigma_t = \beta_t S \beta_t' + \Sigma_\varepsilon$.[14]

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. (2021b) 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 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—we consider different values of d in our

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

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

[14] 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 observations.

results below.

The portfolio formulae are defined only for nonmissing ESG scores. Therefore we consider two ways of imputing missing ESG values. The first imputes missing data with the average/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. Our second imputation value 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, we also consider the imputation of missing ESG scores with a value of -0.5, the worst value possible on our transformed scale.

Returning to the ESG screens we discussed first, note that we can accomplish them without having imputed missing data, but they are equivalent to screens in data where we have. If a missing value is imputed to be -0.5, then that firm would be zeroed-out in the third screen. If a missing value is imputed to be 0, then that firm is not zeroed out by any of the screens in the p_{25} and p_{50} cut-offs.

3.4 Including ESG in the model

Instead, we might suppose that ESG measures provide useful information for either $\alpha_{n,t}$ or $\beta_{n,t}$, or both, and therefore affect 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 into the firm characteristics $z_{n,t}$ and estimates Γ_β (and Γ_α) as in Kelly et al. (2019). In this way, ESG is treated like any other firm characteristic and the estimated model is just what was presented in Section 3.1.

The second route uses ESG and firm characteristics differentially in $\alpha_{n,t}$ and $\beta_{n,t}$. 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 following that $\alpha_{n,t} = \Gamma'_\alpha \zeta_{n,t}$ and $\beta_{n,t} = \Gamma'_\beta z_{n,t}$ and so the modified model is

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

This embodies the idea that ESG measures tell us about return-predicting mispricing, while other firm characteristics tell us about aggregate risk exposure. This model is a modified version of the unrestricted model above, and because $z_{n,t}$ and $\zeta_{n,t}$ are different from each other we obtain modified versions of (2) and (3) along with a new first-order condition for

Γ_α:

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

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

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

for ζ_t the $N_t \times L_\zeta$ matrix stacking $\zeta_{n,t}'$. Note that we no longer need an identification assumption for Γ_α and Γ_β so long as there does not exist Q such that $\zeta_t = Z_t Q$ for all t.[15]

As Section 2 noted, there are many firm-month observations that have no ESG data available. There is not a single obvious solution for dealing with the missing data; we therefore employ two broad alternatives. The first alternative is to restrict attention to only those firm-months where ESG data is available. This essentially treats ESG like the other characteristics, and requires we observe everything in order for the firm-month to be in our data. The second alternative is to impute missing values to either 0 or -0.5, as previously discussed. We report all of these alternatives in our results below.

3.5 Non-systematic strategies

In the basic model (1), Kelly et al. (2019) define a non-systematic investment strategy called a "pure-alpha portfolio":

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

Recall that we impose $\Gamma_\alpha' \Gamma_\beta = 0$: therefore 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.

However, naively using the strategy in (11) with the modified model of (9) does not ensure the portfolio avoids factor risk. Section 3.1 noted we need no identification condition to sep-

[15]This follows because we are no longer able to subtract a constant $K \times 1$ vector ξ from the latent factor time series and add $\Gamma_\beta \xi$ to Γ_α while leaving the model's fitted values unchanged.

arately identify Γ_β and Γ_α in (9), therefore the estimator does not impose any orthogonality between the parameters—indeed, orthogonality is not even well-defined because Γ_β has row dimension L while Γ_α has a different row dimension L_ζ.[16] And 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. As it is distinct from the pure-alpha portfolio construction but achieving the same aim, we call this a "beta-neutral" portfolio:

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

Clearly it is the case that $\beta_t' w_{\alpha\perp\beta,t} = 0$ for every t, regardless of the value of Γ_α. 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 with the instrumented betas.

Both pure-alpha and beta-neutral portfolios are constructed to answer the same question: are there predictable returns that *are not compensation for aggregate risk*? Of course, our main focus is on the importance of ESG measures. In pure-alpha portfolios, we will see if including ESG information affects whatever alpha is present from *all* the characteristics together. In beta-neutral portfolios, we will see if ESG delivers an alpha on its own.

3.6 Estimation details

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 the Sharpe ratios or t-statistics, and only serves to put the various portfolios' means on the same footing. 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

[16]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.

of dimension reduction employed, with limited impact of in- versus out-of-sample estimation. Moreover, 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.

4 Results

In this section we present our main results. We start by using ESG scores to tilt systematic investment strategies. Then we consider the role of ESG measures in model estimation. We consider the robustness of our results to many alternative specifications, and point to comprehensive further analysis in the appendix. Then we consider the properties of well-performing ESG-tilted portfolios. Finally, we connect our empirical results to the empirical literature, and also to a model of responsible investing.

4.1 ESG as a tilt

We assume the investor cares about an ESG-investing mandate for an exogenous reason, regardless of whether or not ESG measures help predict returns. Nevertheless, the investor retains a profit motive and uses the information at hand to form a profitable investment strategy. Our question becomes: how costly is it to adjust a profitable investment strategy to adhere to an ESG-investing mandate?

Given that Figure 2 shows that ESG measures are most widely available for large firms, we focus our main attention on results for large firms. Our main results are in Table I. The takeaway is that following an ESG mandate can have a very low cost, perhaps none at all, in terms of investment performance. The top row reports the original tangency portfolio's performance from step 1. This portfolio is based on a five-factor IPCA model restricted to large firms. The annualized Sharpe ratio is 1.46 ($t = 2.30$) and explains 31% of individual stock returns. These R^2 and tangency-portfolio Sharpe ratios are broadly in line with what Kelly et al. (2019) report for large firms on their longer sample using different characteristics and a different size cut-off. In the first column across a multitude of rows, we see that many ESG-tilted strategies have essentially the same performance.

In Panel B we consider the KLD scores. If we simply zero-out firms with a below-median (p_{50}) ESG score, the tilted-tangency portfolio yields an annualized Sharpe ratio of 1.52 that remains significant at the 5% level ($t = 2.39$), slightly higher than the original tangency

Table I
ESG as a tilt

Notes – Annualized Sharpe ratio and mean, and excess kurtosis and skewness of the monthly returns, for tangency portfolio and tilted portfolio returns. In parentheses are t-statistics: for Sharpe ratios from Lo (2003), and for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

	SR		Mean		Kurtosis	Skewness
Panel A						
Large	1.46	(2.30)	14.58	(7.29)	1.96	0.18
Panel B: KLD						
Large, zero-out $w_{tan,t}$ below p_{25} ESG	1.48	(2.34)	14.79	(7.35)	2.36	0.46
Large, zero-out $w_{tan,t}$ below p_{50} ESG	1.52	(2.39)	15.15	(7.52)	3.86	0.76
Large, zero-out $w_{tan,t}$ below p_{75} ESG	1.39	(2.20)	13.90	(6.48)	6.24	1.10
Large, zero-out $w_{tan,t}$ below p_{25} ESG in long-leg	1.43	(2.25)	14.26	(7.06)	2.21	0.39
Large, zero-out $w_{tan,t}$ below p_{50} ESG in long-leg	1.25	(1.97)	12.49	(6.17)	2.76	0.19
Large, zero-out $w_{tan,t}$ below p_{75} ESG in long-leg	0.78	(1.23)	7.75	(3.78)	1.73	−0.00
Large, zero-out $w_{tan,t}$ not-above p_{25} ESG	1.07	(1.69)	10.67	(6.08)	2.31	−0.16
Large, zero-out $w_{tan,t}$ not-above p_{50} ESG	1.14	(1.81)	11.41	(6.71)	1.99	0.09
Large, zero-out $w_{tan,t}$ not-above p_{75} ESG	1.01	(1.59)	10.05	(5.68)	2.83	−0.39
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0$	1.49	(2.25)	14.87	(7.25)	1.94	−0.03
Large, PFP optimal, missing ESG as 0, $\bar{s} = -0.25$	1.46	(2.20)	14.58	(7.08)	2.03	−0.01
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0.25$	1.49	(2.25)	14.86	(7.26)	1.87	−0.05
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = 0$	1.51	(2.28)	15.08	(7.44)	1.81	0.04
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = -0.25$	1.49	(2.26)	14.92	(7.29)	1.91	−0.01
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = 0.25$	1.51	(2.28)	15.04	(7.47)	1.73	0.08
Large, PST optimal, missing ESG as 0, $d = 0.01$	0.35	(0.56)	3.51	(1.85)	1.91	−0.29
Large, PST optimal, missing ESG as 0, $d = 0.001$	1.36	(2.15)	13.60	(7.11)	1.12	−0.16
Large, PST optimal, missing ESG as 0, $d = 0.0001$	1.49	(2.35)	14.89	(7.71)	1.83	−0.02
Large, PST optimal, missing ESG as −0.5, $d = 0.01$	0.17	(0.22)	1.70	(0.76)	0.25	0.05
Large, PST optimal, missing ESG as −0.5, $d = 0.001$	1.26	(2.00)	12.63	(6.95)	1.16	0.15
Large, PST optimal, missing ESG as −0.5, $d = 0.0001$	1.50	(2.36)	14.97	(7.81)	1.74	0.04
Panel C: Asset4						
Large, zero-out $w_{tan,t}$ below p_{25} ESG	1.39	(2.19)	13.84	(6.70)	2.70	0.03
Large, zero-out $w_{tan,t}$ below p_{50} ESG	1.34	(2.12)	13.39	(6.29)	3.05	0.28
Large, zero-out $w_{tan,t}$ below p_{75} ESG	1.31	(2.06)	13.04	(5.99)	3.77	0.67
Large, zero-out $w_{tan,t}$ below p_{25} ESG in long-leg	1.40	(2.21)	13.98	(6.79)	2.33	0.37
Large, zero-out $w_{tan,t}$ below p_{50} ESG in long-leg	1.22	(1.93)	12.20	(5.84)	2.38	0.47
Large, zero-out $w_{tan,t}$ below p_{75} ESG in long-leg	0.96	(1.52)	9.62	(4.63)	1.75	0.23
Large, zero-out $w_{tan,t}$ not-above p_{25} ESG	0.68	(1.08)	6.83	(3.54)	6.61	0.19
Large, zero-out $w_{tan,t}$ not-above p_{50} ESG	0.59	(0.93)	5.85	(2.96)	7.47	0.25
Large, zero-out $w_{tan,t}$ not-above p_{75} ESG	0.51	(0.81)	5.14	(2.60)	6.00	−0.07
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0$	1.18	(1.34)	11.74	(4.50)	1.51	−0.43
Large, PFP optimal, missing ESG as 0, $\bar{s} = -0.25$	1.16	(1.31)	11.53	(4.39)	1.43	−0.43
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0.25$	1.17	(1.33)	11.71	(4.50)	1.68	−0.45
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = 0$	1.19	(1.35)	11.85	(4.54)	1.68	−0.47
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = -0.25$	1.16	(1.32)	11.60	(4.44)	1.56	−0.45
Large, PFP optimal, missing ESG as −0.5, $\bar{s} = 0.25$	1.20	(1.36)	11.94	(4.56)	1.84	−0.49
Large, PST optimal, missing ESG as 0, $d = 0.01$	0.36	(0.56)	3.55	(1.89)	3.88	−0.26
Large, PST optimal, missing ESG as 0, $d = 0.001$	1.36	(2.14)	13.54	(7.13)	1.59	−0.14
Large, PST optimal, missing ESG as 0, $d = 0.0001$	1.48	(2.34)	14.81	(7.68)	1.91	−0.02
Large, PST optimal, missing ESG as −0.5, $d = 0.01$	0.52	(0.58)	5.15	(1.83)	0.31	−0.20
Large, PST optimal, missing ESG as −0.5, $d = 0.001$	1.37	(2.17)	13.70	(7.01)	1.99	−0.25

Continued on next page

	SR		Mean		Kurtosis	Skewness
Large, PST optimal, missing ESG as -0.5, $d = 0.0001$	1.49	(2.35)	14.87	(7.69)	1.95	-0.03
Panel D: Sustainalytics						
Large, zero-out $w_{tan,t}$ below p_{25} ESG	1.42	(2.25)	14.22	(7.04)	2.04	0.19
Large, zero-out $w_{tan,t}$ below p_{50} ESG	1.37	(2.17)	13.71	(6.65)	2.32	0.23
Large, zero-out $w_{tan,t}$ below p_{75} ESG	1.33	(2.10)	13.31	(6.34)	2.70	0.30
Large, zero-out $w_{tan,t}$ below p_{25} ESG in long-leg	1.41	(2.22)	14.07	(6.90)	2.24	0.19
Large, zero-out $w_{tan,t}$ below p_{50} ESG in long-leg	1.31	(2.07)	13.06	(6.28)	2.36	0.24
Large, zero-out $w_{tan,t}$ below p_{75} ESG in long-leg	1.17	(1.85)	11.72	(5.59)	1.91	0.25
Large, zero-out $w_{tan,t}$ not-above p_{25} ESG	0.74	(1.17)	7.36	(3.84)	17.90	2.74
Large, zero-out $w_{tan,t}$ not-above p_{50} ESG	0.65	(1.02)	6.45	(3.40)	14.03	2.21
Large, zero-out $w_{tan,t}$ not-above p_{75} ESG	0.62	(0.97)	6.15	(3.05)	9.29	1.77
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0$	1.86	(1.47)	18.49	(6.23)	0.75	0.17
Large, PFP optimal, missing ESG as 0, $\bar{s} = -0.25$	1.86	(1.47)	18.48	(6.12)	0.78	0.16
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0.25$	1.83	(1.45)	18.24	(6.23)	0.68	0.19
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = 0$	1.87	(1.48)	18.56	(6.34)	0.71	0.17
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = -0.25$	1.86	(1.47)	18.53	(6.21)	0.72	0.13
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = 0.25$	1.85	(1.47)	18.45	(6.40)	0.68	0.20
Large, PST optimal, missing ESG as 0, $d = 0.01$	0.48	(0.76)	4.82	(2.59)	6.46	-0.23
Large, PST optimal, missing ESG as 0, $d = 0.001$	1.42	(2.24)	14.20	(7.45)	1.53	0.01
Large, PST optimal, missing ESG as 0, $d = 0.0001$	1.48	(2.34)	14.83	(7.68)	1.88	-0.01
Large, PST optimal, missing ESG as -0.5, $d = 0.01$	0.16	(0.13)	1.63	(0.41)	0.35	0.02
Large, PST optimal, missing ESG as -0.5, $d = 0.001$	1.30	(2.05)	12.97	(6.68)	1.67	0.02
Large, PST optimal, missing ESG as -0.5, $d = 0.0001$	1.48	(2.33)	14.74	(7.63)	1.91	-0.01
Panel E: RepRisk						
Large, zero-out $w_{tan,t}$ below p_{25} ESG	1.53	(2.42)	15.31	(7.63)	2.21	0.45
Large, zero-out $w_{tan,t}$ below p_{50} ESG	1.51	(2.38)	15.06	(7.33)	2.75	0.60
Large, zero-out $w_{tan,t}$ below p_{75} ESG	1.46	(2.31)	14.59	(6.99)	2.93	0.66
Large, zero-out $w_{tan,t}$ below p_{25} ESG in long-leg	1.50	(2.37)	15.01	(7.45)	2.20	0.45
Large, zero-out $w_{tan,t}$ below p_{50} ESG in long-leg	1.37	(2.17)	13.72	(6.61)	2.47	0.46
Large, zero-out $w_{tan,t}$ below p_{75} ESG in long-leg	1.26	(1.98)	12.55	(5.99)	2.17	0.44
Large, zero-out $w_{tan,t}$ not-above p_{25} ESG	0.65	(1.03)	6.53	(3.57)	8.57	0.86
Large, zero-out $w_{tan,t}$ not-above p_{50} ESG	0.62	(0.98)	6.17	(3.36)	7.03	0.35
Large, zero-out $w_{tan,t}$ not-above p_{75} ESG	0.50	(0.79)	5.01	(2.64)	9.99	0.35
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0$	1.16	(1.14)	11.58	(3.87)	1.54	-0.50
Large, PFP optimal, missing ESG as 0, $\bar{s} = -0.25$	1.13	(1.11)	11.29	(3.75)	1.64	-0.54
Large, PFP optimal, missing ESG as 0, $\bar{s} = 0.25$	1.17	(1.15)	11.66	(3.90)	1.47	-0.48
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = 0$	1.17	(1.15)	11.68	(3.90)	1.52	-0.49
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = -0.25$	1.15	(1.13)	11.46	(3.82)	1.61	-0.52
Large, PFP optimal, missing ESG as -0.5, $\bar{s} = 0.25$	1.18	(1.16)	11.78	(3.94)	1.44	-0.47
Large, PST optimal, missing ESG as 0, $d = 0.01$	0.68	(0.91)	6.78	(2.61)	9.92	-0.79
Large, PST optimal, missing ESG as 0, $d = 0.001$	1.47	(2.31)	14.65	(7.65)	1.09	0.03
Large, PST optimal, missing ESG as 0, $d = 0.0001$	1.50	(2.36)	14.93	(7.73)	1.84	0.00
Large, PST optimal, missing ESG as -0.5, $d = 0.01$	-0.28	(-0.12)	-2.78	(-0.33)	-0.71	-0.22
Large, PST optimal, missing ESG as -0.5, $d = 0.001$	1.36	(2.14)	13.55	(6.86)	0.90	0.02
Large, PST optimal, missing ESG as -0.5, $d = 0.0001$	1.49	(2.35)	14.90	(7.71)	1.81	0.01

portfolio. If we instead zero-out below-median ESG only in the long-leg, the tilted-tangency portfolio has a mildly lower Sharpe of 1.25 that remains significant ($t = 1.97$). Making the screen less-stringent at p_{25} has little effect when applied to both legs, but results in a smaller drop (to 1.43, $t = 2.25$) when applied only to the long-leg. A more-stringent p_{75} screen deteriorates performance only mildly when applied to both legs (to 1.39, $t = 2.20$) but is more deleterious when applied only to the long-leg (0.78, $t = 1.23$). Hence, the portfolio is not immune to *whatever* change implemented—of course, we wouldn't expect it to be. But simple, reasonable screens deliver portfolios that lose little-to-none of the original tangency portfolio's profits. In the remaining panels, zeroing-out below-median ESG leads to tilted-tangency portfolios with significant Sharpe ratios of 1.34, 1.37, and 1.51 for the Asset4, Sustainalytics, and RepRisk ESG providers, respectively. Hence, the low costs of an ESG-mandate tilt are common across the various measures.

If instead we *keep* only those firms with above median ESG (labeled "zero-out $w_{tan,t}$ not-above p_{50}"), we see worse outcomes. This is the case particularly for the ESG providers with fewer observations (Asset4, Sustainalytics, and RepRisk). For these providers, the Sharpe ratios more than halve and become insignificant. We generally see the portfolio's kurtosis and skewness rise. This follows because the portfolio contains a much smaller number of stocks than the original tangency portfolio, and so picks up the fatter tails and positive skewness inherent to individual stock returns (see for instance Bessembinder, 2018). Therefore, it costs more to restrict non-zero portfolio weights only to those firms with observable and acceptable ESG scores—it is instead preferable to screen a firm only if we can see its ESG score is unacceptable.

Now we turn to the model-implied strategies. Recall that these require nonmissing values, and therefore we report results for both imputed values, -0.5 and 0. Considering first the optimal strategies of Pedersen et al. (2020) in Panel A, we see virtually no cost to tilting to achieve an ESG mandate. The results are surprisingly insensitive to the choice of imputed value or average ESG level $\bar{s}$: Sharpe ratios remain between 1.46–1.51, insignificantly different than the tangency portfolio's performance.[17] It turns out that the performance using KLD scores is not uniform across the other score providers. In Panels C and E, the Asset 4 and RepRisk scores lead to noticeable declines in the Sharpe ratios; but in Panel D, the Sustainalytics scores lead to an increase. What is true is that for all the non-KLD scores, the portfolios' Sharpe ratios are no longer statistically significant, while the average returns remain so—reflecting low power of Lo (2003) t-statistics for short time series.

Next, examining the optimal strategies of Pastor et al. (2021b), we see greater sensitivity to the d parameter across the board. To gain intuition as to why, suppose $d = 0.01$. If a firm were to drop from the best ESG score to the worst, the ESG-investor in Pastor et al. (2021b)

[17] Technically speaking, the portfolio being tilted is not the tangency portfolio *per se* but the Markowitz portfolio of individual firms.

acts as if its expected return has dropped by 1% per month. Hence, $d = 0.01$ is a strong "ESG taste," and with it the tilted portfolios are harmed, especially when missing values are imputed to be -0.5. But as d gets smaller, the ESG taste weakens and at $d = 0.0001$ we get the Markowitz portfolio essentially untouched, and therefore profitability is good.

The main takeaway from this section is that there are numerous ways to tilt a profitable systematic portfolio with an ESG mandate and sacrifice nothing: Sharpe ratios can remain unchanged and average returns stay high and statistically significant.

4.2 ESG in the model

The previous section shows that a variety of systematic portfolios can be ESG-tilted with little effect on their performance. How can this be? Table II reports the section's main results—ESG measures do not significantly predict returns. More specifically, we find no significant role for ESG measures in driving beta or alpha.

To build up this conclusion, let us start in Panel A with the first row labeled "Large, 5-factor restricted". As before, this is a five-factor IPCA model which has an annualized Sharpe ratio of 1.46 ($t = 2.30$) and explains 31% of individual stock returns. Relative to the Kelly et al. (2019) benchmark sample of all firms, the R^2 is higher and Sharpe ratios lower for large firms only.

Do the characteristics yield profits apart from what they tell us about beta? The second row ("Large, 5-factor unrestricted") says no. This pure-alpha portfolio uses characteristic information to form a portfolio that is orthogonal to market-beta exposure—it receives insignificant profits, with a negligible average return of 1.8% per annum ($t = 1.01$) and annualized Sharpe ratio of 0.18 ($t = 0.29$). This is an indication that characteristics do not predict large-firm returns apart from risk exposures, which the five-factor IPCA captures.

Now consider what happens when we add the ESG characteristic: these are the rows with labels starting "Large, missing ESG as..." in Panels B-E. The main point is that these rows are essentially unchanged from what the first two rows of Panel A report. Including the ESG measures does very little to the tangency portfolios' profitability, as Sharpe ratios remain significant and range between 1.36 to 1.58. Including ESG measures does very little to the pure-alpha portfolios' profitability, as Sharpe ratios remain insignificant and range between -0.38 to 0.38. Hence, against the backdrop of other firm characteristics, ESG tells us nothing about either beta or alpha.

Table II
ESG in the model

Notes – Total R^2, annualized Sharpe ratio and mean for tangency portfolios and pure-alpha portfolio returns. In parentheses are t-statistics: for Sharpe ratios from Lo (2003), and for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility. Total R^2 and mean return in percentage.

	R^2	Factor				Pure-alpha			
		SR		Mean		SR		Mean	
Panel A									
Large, 5-factor restricted	31.0	1.46	(2.30)	14.57	(7.28)				
Large, 5-factor unrestricted	31.1					0.18	(0.29)	1.82	(1.01)
Panel B: KLD									
Large, missing ESG as 0, 5-factor restricted	31.1	1.41	(2.23)	14.13	(7.17)				
Large, missing ESG as 0, 5-factor unrestricted	31.2					−0.08	(−0.11)	−0.75	(−0.37)
Large, missing ESG as −0.5, 5-factor restricted	31.1	1.36	(2.15)	13.62	(6.97)				
Large, missing ESG as −0.5, 5-factor unrestricted	31.2					0.19	(0.28)	1.85	(0.98)
Large, ESG nonmissing, 5-factor restricted	32.8	1.16	(1.76)	11.59	(6.43)				
Large, ESG nonmissing, 5-factor unrestricted	32.9					0.24	(0.36)	2.40	(1.27)
Large, ESG nonmissing, ESG included, 5-factor restricted	32.9	1.16	(1.75)	11.55	(6.39)				
Large, ESG nonmissing, ESG included, 5-factor unrestricted	33.0					0.16	(0.25)	1.62	(0.85)
Panel C: Asset4									
Large, missing ESG as 0, 5-factor restricted	31.0	1.48	(2.33)	14.76	(7.37)				
Large, missing ESG as 0, 5-factor unrestricted	31.1					0.12	(0.13)	1.16	(0.45)
Large, missing ESG as −0.5, 5-factor restricted	31.0	1.47	(2.32)	14.68	(7.28)				
Large, missing ESG as −0.5, 5-factor unrestricted	31.1					−0.07	(−0.08)	−0.69	(−0.27)
Large, ESG nonmissing, 5-factor restricted	35.2	1.33	(1.51)	13.23	(5.77)				
Large, ESG nonmissing, 5-factor unrestricted	35.2					0.32	(0.37)	3.20	(1.28)
Large, ESG nonmissing, ESG included, 5-factor restricted	35.2	1.31	(1.49)	13.09	(5.67)				
Large, ESG nonmissing, ESG included, 5-factor unrestricted	35.3					0.34	(0.39)	3.37	(1.36)
Panel D: Sustainalytics									
Large, missing ESG as 0, 5-factor restricted	31.0	1.47	(2.32)	14.71	(7.32)				
Large, missing ESG as 0, 5-factor unrestricted	31.1					0.38	(0.30)	3.76	(1.12)
Large, missing ESG as −0.5, 5-factor restricted	31.0	1.47	(2.32)	14.69	(7.31)				
Large, missing ESG as −0.5, 5-factor unrestricted	31.1					−0.10	(−0.08)	−1.00	(−0.28)
Large, ESG nonmissing, 5-factor restricted	35.9	1.90	(1.50)	18.91	(6.60)				
Large, ESG nonmissing, 5-factor unrestricted	36.0					0.37	(0.30)	3.69	(1.04)
Large, ESG nonmissing, ESG included, 5-factor restricted	36.0	1.89	(1.50)	18.82	(6.59)				
Large, ESG nonmissing, ESG included, 5-factor unrestricted	36.1					0.37	(0.30)	3.71	(1.05)
Panel E: RepRisk									
Large, missing ESG as 0, 5-factor restricted	31.0	1.46	(2.31)	14.63	(7.30)				
Large, missing ESG as 0, 5-factor unrestricted	31.1					0.24	(0.23)	2.36	(0.77)
Large, missing ESG as −0.5, 5-factor restricted	31.0	1.58	(2.49)	15.76	(8.65)				
Large, missing ESG as −0.5, 5-factor unrestricted	31.1					−0.38	(−0.38)	−3.81	(−1.33)
Large, ESG nonmissing, 5-factor restricted	35.8	1.51	(1.48)	15.01	(5.97)				
Large, ESG nonmissing, 5-factor unrestricted	35.9					−0.30	(−0.30)	−3.00	(−1.04)
Large, ESG nonmissing, ESG included, 5-factor restricted	35.8	1.51	(1.48)	15.03	(5.97)				
Large, ESG nonmissing, ESG included, 5-factor unrestricted	35.9					−0.30	(−0.30)	−2.99	(−1.04)

Moreover, this remains true if we restrict the sample to large firms with nonmissing ESG information, so as to abstract from data imputation. Including ESG scores barely moves the R^2s, Factor Sharpe ratios, and Pure-alpha Sharpe ratios when we compare a label beginning "Large, ESG nonmissing, 5-factor'...' to a comparative label beginning "Large, ESG non-missing, ESG included, 5-factor..." within Panels B-E. Relative to the rows discussed above, the particular numbers do change somewhat (particularly for Sustainalytics, which has a much shorter sample period)—but the point is that including ESG measures do nothing to change the tangency portfolios' or pure-alpha portfolios' average returns.

Table III
ESG in the model only as alpha

Notes – Total and predictive R^2 in percentage, and annualized Sharpe ratio and mean of beta-neutral portfolio. Row labeled "Baseline" is the 5-factor restricted model not using the ESG characteristic. Subsequent rows use the ESG characteristics only for $\alpha_{n,t}$ using equations 2.1, 3.1, and 10. In parentheses are t-statistics: for Sharpe ratios from Lo (2003), and for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

	Total R^2		Predictive R^2		Beta-neutral							
					Sharpe ratio				Mean			
Baseline	31.04		0.39									
	Missing ESG as											
	0	−0.5	0	−0.5	0		−0.5		0		−0.5	
KLD	31.04	31.04	0.39	0.39	0.20	(0.32)	0.10	(0.17)	2.03	(1.04)	1.09	(0.59)
Asset4	31.04	31.04	0.39	0.39	0.06	(0.09)	0.10	(0.16)	0.60	(0.33)	1.02	(0.52)
Sustainalytics	31.04	31.04	0.39	0.40	0.03	(0.05)	−0.32	(−0.50)	0.34	(0.19)	−3.16	(−1.59)
RepRisk	31.04	31.04	0.38	0.38	0.20	(0.32)	0.15	(0.24)	2.01	(1.03)	1.50	(0.74)

Table III reports the results when we instead constrain the ESG measure to give us $\alpha_{n,t}$ and the remaining characteristics to give us $\beta_{n,t}$, from the modified model in (2.1), (3.1), and (10). The obvious takeaway from the first two columns is that ESG alphas do *absolutely nothing* to change the overall fit of the model. The next two columns show us that ESG alphas provide *essentially no* predictive content. This is further emphasized by the final two columns, that show that pure-alpha portfolios based only on ESG scores give insignificant profits.[18]

Taken all together, the results cast doubt on the idea that ESG scores are useful for *creating* profitable portfolio strategies. We find no evidence that they define alpha with respect to successful asset-pricing factors. And we find no role for ESG scores in determining firms' beta. This helps explain why Section 4.1 found that ESG tilts can have no cost.

4.3 Robustness

We now consider several robustness checks of the main results of the previous two sections.

For the sake of exposition, in Table IV we focus on the KLD data and select specifications, and relegate to the appendix comprehensive tables of other specifications and other ESG

[18]In fact, the pure-alpha portfolio if anything yields negative returns when we impute missing values as −0.5. This is because we predominantly impute values for small firms. When we estimate a positive Γ_α, as we do here, this means that the short-leg of the pure-alpha portfolio has a lot of small firms—the negative average return is picking up a size effect.

Table IV
ESG in the model, Robustness

Notes – Total R^2, annualized Sharpe ratio and mean for tangency portfolios and pure-alpha portfolio returns. In parentheses are t-statistics: for Sharpe ratios from Lo (2003), and for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility. Total R^2 and mean return in percentage.

	R^2	Factor			
		SR		Mean	
Large, FF5C restricted	28.6	1.14	(1.80)	11.38	(6.37)
Large, missing ESG as −0.5, FF5C restricted	28.6	1.14	(1.79)	11.34	(6.35)
All firms, 5-factor restricted	16.4	4.08	(6.28)	40.75	(16.35)
All firms, missing ESG as −0.5, 5-factor restricted	16.4	4.08	(6.28)	40.76	(16.34)
All firms, FF5C restricted	13.7	3.51	(5.45)	35.08	(15.57)
All firms, missing ESG as −0.5, FF5C restricted	13.7	3.49	(5.41)	34.84	(15.46)
Large, Total ind. adj., missing ESG as −0.5, 5-factor restricted	31.1	1.41	(2.22)	14.07	(7.11)
Large, E, missing ESG as −0.5, 5-factor restricted	31.1	1.39	(2.19)	13.84	(7.00)
Large, S, missing ESG as −0.5, 5-factor restricted	31.1	1.38	(2.18)	13.81	(7.04)
Large, G, missing ESG as −0.5, 5-factor restricted	31.1	1.46	(2.31)	14.60	(7.28)
Large, 2010-, 5-factor restricted	33.0	1.98	(1.80)	19.72	(7.04)
Large, 2010-, missing ESG as −0.5, 5-factor restricted	33.1	1.98	(1.81)	19.75	(7.04)
Large, 2010-, FF5C restricted	30.5	1.74	(1.59)	17.32	(6.21)
Large, 2010-, missing ESG as −0.5, FF5C restricted	30.5	1.74	(1.59)	17.33	(6.22)
Large, Slow, 5-factor restricted	28.1	1.10	(1.74)	11.03	(6.12)
Large, Slow, missing ESG as −0.5, 5-factor restricted	28.1	1.19	(1.88)	11.92	(6.56)
Large, Slow, FF5C restricted	26.0	0.65	(1.03)	6.51	(3.64)
Large, Slow, missing ESG as −0.5, FF5C restricted	26.0	0.65	(1.03)	6.47	(3.62)
All firms, Slow, 5-factor restricted	13.5	3.54	(5.48)	35.31	(15.08)
All firms, Slow, missing ESG as −0.5, 5-factor restricted	13.5	3.53	(5.48)	35.28	(15.08)
All firms, Slow, FF5C restricted	10.9	2.99	(4.66)	29.85	(14.49)
All firms, Slow, missing ESG as −0.5, FF5C restricted	10.9	2.98	(4.65)	29.79	(14.51)

data providers. We focus on the KLD data simply because it has the most observations over our entire sample period.

Given the prominence of the five Fama and French (2015) factors and a momentum factor (Carhart, 1997), the first two rows repeat the analysis of Table II Panel A using those six factors as the systematic-risk space. This model, which we call FF5C, implies a tangency portfolio with a lower Sharpe ratio than the five IPCA factors, and it is not significant at the 5% level—this highlights the impact of estimating the aggregate risk space, as opposed to taking it as given by the FF5C observable factors. Nevertheless, average returns are significant and just under 12% per annum when using the FF5C model, and Section 4.2's main point is robust: the estimates are basically unchanged when we include the ESG score.

The next four rows consider the sample of all firms. The 5-factor tangency portfolio is more profitable when it can also invest in small firms, with an annualized Sharpe ratio of

4.08 ($t = 6.28$); likewise the FF5C tangency portfolio is more profitable at 3.51 ($t = 5.45$). Nonetheless, the main point holds: the Sharpe ratios and R^2 are essentially unchanged by including the ESG measure.

Moving back to the large firm sample, we find that different varieties of the ESG score do not change things. We consider a best-in-class version of the Total ESG score, defined as an industry-adjusted measure for each firm, for each of the Fama-French 49 industries (such a measure is considered by Pedersen et al., 2020). We also consider the separate environmental ("E"), social ("S"), and governance ("G") subcomponents. Nothing appreciably changes. Next we consider starting the sample in 2010 ("2010-"). Adding the ESG measure does nothing to change the estimates of either the five-factor IPCA or six-factor FF5C models.

Finally, we consider the effect of restricting the non-ESG firm characteristics to a subset of 17 "slow" characteristics listed in Section 2. The main point continues to manifest, even against the backdrop of this reduced conditioning information, and we once again see that including the ESG characteristic does little-to-nothing to change the estimates.

The conclusion from Table IV is that ESG does not impact the estimated model in virtually any way. This is true across different models for systematic risk, samples of firms, particular ESG measures or subcomponents, or when restricted to only the most-recent sample. The appendix contains many more tables illustrating that this conclusion is true for different ESG providers and other specifications in addition to what has been presented here. Therefore we conclude that Section 4.2's main result is very robust.

Turning now to the robustness of Section 4.1's main result, Table V considers some select specifications, again using only the KLD ESG scores.

Consistent with what we saw in Table IV, the tangency portfolio is much more profitable when allowed to invest in smaller firms, as the first row shows. However, the main point persists: there are numerous ways to screen the tangency portfolio with only a small effect on performance—that is evident in the next three rows. Once again, our simple screens change very little when zeroing-out only those firms we can see have unacceptable ESG scores. The optimal portfolios of PFP and PST (in the next four rows) can have a larger effect than we saw in large firms. Nonetheless, the tilted portfolios still obtain Sharpe ratios of 2.67-3.26 with average returns of 26.7-32.5 percent.

In the remainder of the table, we continue with the simple screen of dropping firms with ESG scores below median. In the next eight rows we see that industry-adjustment or using ESG subcomponents changes little. Hence, in both the large and all-firm samples, we can

Table V
ESG as a tilt, Robustness

Notes – Annualized Sharpe ratio and mean, and excess kurtosis and skewness of the monthly returns, for tangency portfolios and pure-alpha portfolio returns. In parentheses are t-statistics: for Sharpe ratios from Lo (2003), and for means from Newey and West (1987) with three lags. Portfolios scaled to have 10% annualized volatility.

	SR		Mean		Kurtosis	Skewness
All firms	4.08	(6.28)	40.75	(16.35)	0.89	0.23
All firms, zero-out $w_{tan,t}$ below p_{50} ESG	4.12	(6.33)	41.11	(15.63)	0.84	0.40
All firms, zero-out $w_{tan,t}$ below p_{50} ESG in long-leg	3.92	(6.05)	39.15	(15.38)	0.53	0.25
All firms, zero-out $w_{tan,t}$ not-above p_{50} ESG	1.01	(1.59)	10.07	(5.31)	14.79	−1.34
All firms, PFP optimal, missing ESG as 0, $\bar{s} = 0$	3.26	(4.85)	32.50	(13.68)	2.42	0.12
All firms, PFP optimal, missing ESG as −0.5, $\bar{s} = 0$	3.19	(4.75)	31.82	(14.05)	2.68	0.31
All firms, PST optimal, missing ESG as 0, $d = 0.001$	2.88	(4.50)	28.78	(12.82)	2.46	−0.01
All firms, PST optimal, missing ESG as −0.5, $d = 0.001$	2.67	(4.17)	26.70	(13.19)	2.78	0.32
Large, Total ind. adj., zero-out $w_{tan,t}$ below p_{50} ESG	1.44	(2.27)	14.38	(6.92)	4.56	0.83
Large, E, zero-out $w_{tan,t}$ below p_{50} ESG	1.52	(2.40)	15.18	(7.66)	2.28	0.45
Large, S, zero-out $w_{tan,t}$ below p_{50} ESG	1.55	(2.44)	15.45	(7.74)	3.30	0.62
Large, G, zero-out $w_{tan,t}$ below p_{50} ESG	1.46	(2.31)	14.61	(7.23)	2.09	0.24
All firms, Total ind. adj., zero-out $w_{tan,t}$ below p_{50} ESG	4.01	(6.17)	40.00	(14.67)	0.80	0.45
All firms, E, zero-out $w_{tan,t}$ below p_{50} ESG	4.14	(6.37)	41.39	(16.31)	0.92	0.28
All firms, S, zero-out $w_{tan,t}$ below p_{50} ESG	4.07	(6.27)	40.65	(15.56)	0.84	0.39
All firms, G, zero-out $w_{tan,t}$ below p_{50} ESG	4.11	(6.32)	41.03	(16.25)	0.88	0.26
Large, 2010-	1.98	(1.80)	19.72	(7.04)	0.82	−0.30
Large, 2010-, zero-out $w_{tan,t}$ below p_{50} ESG	1.73	(1.58)	17.24	(7.15)	0.09	−0.43
All firms, 2010-	2.89	(2.61)	28.81	(10.07)	1.39	−0.14
All firms, 2010-, zero-out $w_{tan,t}$ below p_{50} ESG	2.87	(2.59)	28.58	(10.16)	2.47	0.26

use a variety of particular ESG subcomponents or industry adjustments with little effect. The final four rows (with "2010-" in the label) show that the main conclusion holds when we start the sample in 2010, for both the large and all-firm sample. Of particular note for this exercise is the all-firm sample, since in the latter half of our data the KLD coverage of smaller firms is much greater. Amongst all firms, there is almost no effect at all of following an ESG-mandate screen.

The message from Table V is that ESG tilts can have essentially no cost. This is true across firm samples and ESG providers. The appendix contains many more tables illustrating that this conclusion is true for different ESG providers and other specifications in addition to what has been presented here. Therefore, as with Section 4.2's, we conclude that Section 4.1's main result is very robust. There are many "low-cost" portfolios that are profitable and deliver on an ESG mandate.

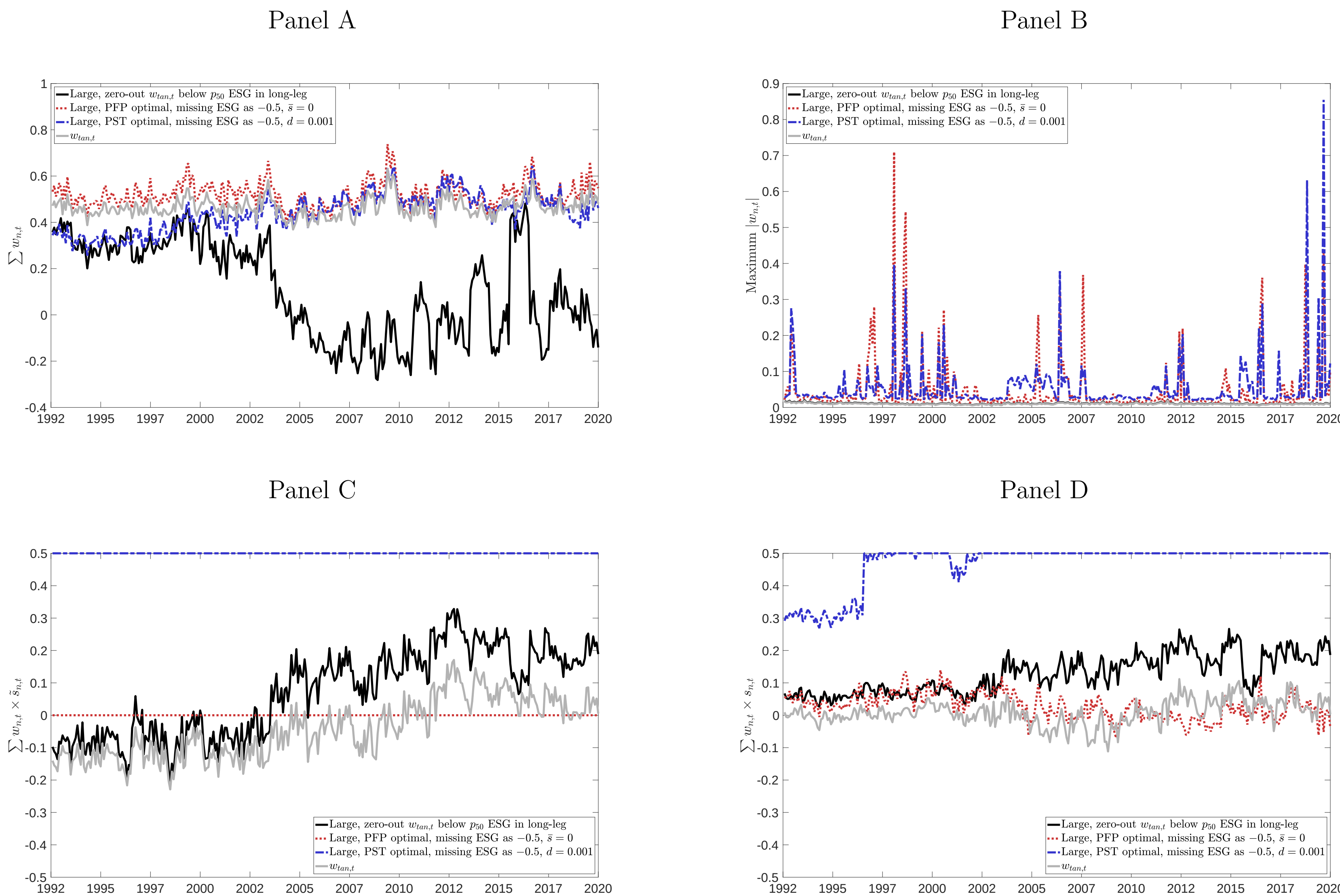

Figure 3: Properties of ESG-tilted portfolios

Notes – Time-series of portfolio properties for three well-performing portfolios from Table I Panel B (KLD). 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 measure where missing ESG is set to -0.5, which we denote $\tilde{s}$. Panel D reports the portfolio-weighted-average ESG measure only for stocks with nonmissing ESG, which we denote s.

4.4 Properties of ESG-tilted portfolios

Figure 3 plots additional information for three ESG-tilted portfolios that performed well in Table I Panel B (KLD). The gray solid lines are for the original tangency portfolio, whose annualized Sharpe ratio is 1.46 ($t = 2.30$) and mean return is 14.58% ($t = 7.29$). The black solid lines are for the tangency portfolio that screens out firms in the long-leg with an observed ESG score below the median—this portfolio has an annualized Sharpe ratio of 1.25 ($t = 1.97$) and annualized mean return of 12.49% ($t = 6.17$). The red dotted lines are for the Pedersen et al. (2020) optimal portfolio where missing ESG are imputed to be -0.5 and the average ESG level is set as $\bar{s} = 0$—this portfolio has an annualized Sharpe ratio of 1.51 ($t = 2.28$) and annualized mean return of 15.08% ($t = 7.44$). Finally, the blue dashed-dot lines are for the Pastor et al. (2021b) optimal portfolio where missing ESG are imputed to be -0.5 and the ESG taste parameter is set as $d = 0.001$—this portfolio has an annualized Sharpe ratio of 1.26 ($t = 2.00$) and annualized mean return of 12.63% ($t = 6.95$).

Panel A reports the sum of the portfolio weights each period. These values hover around 0.4–0.5 for the original tangency portfolio, as well as the PFP and PST portfolios. Recall that a zero-cost spread portfolio has $\sum w_{n,t} = 0$ while a long-only portfolio has $\sum w_{n,t} = 1$. So these systematic strategies are net long over the whole period at reasonable values.[19] Meanwhile, the screened tangency portfolio has a lower net weight, consistent with our having zeroed out bad-ESG stocks in the long-leg.

Panel B reports the maximum absolute weight taken in any one firm. Striking differences are apparent between the screened tangency portfolio and the ESG-optimal portfolios. In multiple months, the PFP and PST optimal portfolios take very large positions in a stock, and so the red and blue lines shoot up. Indeed, the maximum position size for the PFP portfolio is 0.71 and for the PST portfolio is 0.85, coming from the underlying Markowitz weights themselves. On the other hand, the IPCA (screened or original) tangency portfolio at most every puts a weight of 0.02 on any one stock, implying a much more stable diversification throughout—a benefit of constructing the factor-model-implied systematic strategy.

The last two panels report the portfolio-weighted-average ESG score for each portfolio.[20] In Panel C we use the ESG score $\tilde{s}$ where missing values have been imputed to be -0.5. Consider first the black line in Panel C showing the ESG performance of the screened tangency portfolio—it starts out hovering around -0.25 for the first decade, and then hovers around 0.1 for the last decade. This reflects how the ESG data availability interacts with a screen

[19] Recall that this was a necessary condition for the PFP portfolio.

[20] For visibility, we constrain the portfolio-weighted average to lie in $[-0.5, 0.5]$; we discuss this more below.

only of those firms that are observed to have bad ESG. Hence, at the beginning of the sample there are many firms for whom ESG was missing and are not screened, and in Panel C we impute $\tilde{s}_{n,t} = -0.5$ in this case. So the screened tangency portfolio's ESG performance is only modest at the beginning. As more ESG data comes on line, more bad-ESG firms are screened, and now the portfolio's weighted-average ESG score is better than median ESG. Meanwhile, recall from Table V that in data since 2010, the performance of this very portfolio was little affected by the ESG screen (the Sharpe ratio insignificantly moved from 1.98 to 1.73)—indicating that this increasing ESG performance does not coincide with a decrease in profitability. Comparing to the original tangency portfolio's gray line, the ESG-screen has effectively improved the portfolio-weighted average ESG measure—and this improvement widens with the expansion in KLD coverage around 2004.

Now consider the red dotted line of the PFP portfolio that is exactly 0 every period. This is just as expected—we fed in exactly this $\tilde{s}$ each month to form the PFP portfolio and targeted $\bar{s} = 0$, and so PFP's formulae ensure that the weighted-average ESG score zero. Meanwhile, the PST portfolio's weighted-average ESG score is not pinned down by their formulae. It turns out that PST's average ESG hovers above 0.5 (for readability we plot this as 0.5 on the figure). This might seem surprising at first, since the ESG scores live on the interval $[-0.5, 0.5]$, but it reflects how the ESG taste operates. From the investor's perspective, a firm with a negative ESG score has a lower *effective* expected return. All else equal, a lower effective expected return would have a lower weight in a portfolio; in fact, the investor might even want to short it. This means there is positive correlation between ESG scores and portfolio weights. Therefore we see both large positive weights on positive ESG scores, and large negative weights on negative ESG scores, and each leg's weight can be greater than one—implying that the overall portfolio-weighted ESG score is high. The large difference between PFP and PST portfolio-weighted-average ESG scores is a feature of their different modeling assumptions for how investors incorporate ESG into their investment decision.

Panel D instead uses the ESG score s where we do no imputation: hence these portfolio-weighted averages are taken only over the firms with nonmissing ESG. In terms of what one can actually observe, the screened tangency portfolio (black line) has better ESG performance at- or a little above median ESG, with larger improvement over the original tangency portfolio (gray line).[21] This ESG performance is superior to the PFP-optimal portfolio's, which also hovers around zero but now varies because s is different from the $\tilde{s}$ the PFP formulae used to find weights. Meanwhile, the PST-optimal portfolio continues to have very good ESG

[21] If these are only firms with observed ESG scores, and we have thrown out any firm with an ESG score below 0, how can the weighted average be negative? Because of short portfolio weights.

performance which is notably above what PFP or the screened-tangency provide.

Hence, there is variation amongst the portfolio strategies in the literature incorporating ESG concerns. The PFP and PST portfolios sometimes see large individual positions. The PST-optimal portfolio's ESG performance is always strikingly strong. The PFP-optimal portfolio's ESG performance can be made to be constant. The screened-tangency portfolio's ESG performance varies around reasonably good levels. We note that ESG-tilted portfolios can have similarly strong performance while exhibiting notable differences.

4.5 Relationship to Empirical Literature

There is a large literature on what the data has to say about socially-responsible investing.[22] In this section we draw a connection between our results and other recent work in public equities. It is clear from our introduction's literature survey (which could not possibly be exhaustive) that the full picture of ESG characteristics' impact on returns is complicated. Our novel contribution is using the IPCA methodology that a) allows us to capture systematic risk more accurately relative to the existing literature (Kelly et al., 2019, forthcoming) and b) allows for ESG measures and other firm characteristics to drive cross-sectional and time-series variation in alphas, betas, or both, while simultaneously controlling for a host of other firm characteristics that investors may be considering. This is important in our setting, as a model that inaccurately captures the mean-variance-efficient frontier may lead one to falsely attribute risk-based effects to alpha.

With these key advantages in mind, we connect our results in greater detail to the recent work by Pastor et al. (2021a) in the following. Pastor et al. (2021a) construct a "green" factor as the value-weighted portfolio sorted by the MSCI E score. On data since November 2012, they find that this portfolio has a significant unconditional alpha against the Fama and French (1993) three-factor model, and is robust to instead using the Carhart (1997) or Fama and French (2015) models. We do not have access to the MSCI data, but note that MSCI purchased RiskMetrics (including KLD) in 2010 and currently operates both products. Hence, we construct a green factor from our data by using the KLD E industry-adjusted

[22]Outside of public equity markets there is further evidence of differential average returns driven by ESG concerns. Baker et al. (2018) and Zerbib (2019) find that green municipal bonds are issued at a premium to similar bonds, thereby giving a lower yield-to-maturity, especially for externally-certified green bonds. Barber et al. (2021) find that impact venture-capital funds receive 4.7 percentage points lower internal rates of return, on average, which indicates that investors are willing to trade lower returns for their preference for impact—furthermore they find that this effect varies significantly across different regulatory frameworks. Giglio et al. (forthcoming) survey the literature's results in public equities, municipal bonds, private equity, housing, and other asset classes.

Table VI
Unconditional alpha from regressions

Notes – Regressions of spread portfolio using tercile portfolios sorted by the KLD E industry-adjusted measure, from sample of all firms. In parentheses are t-statistics using Newey and West (1987) with three lags. Intercept reported in annualized percentage, R^2 reported in percentage.

	Intercept		Mkt-RF		SMB		HML		RMW		CMA		Mom		$R^2(\%)$
FF3	3.11	(2.49)	0.01	(0.16)	−0.41	(−10.44)	−0.00	(−0.08)							56.0
FF5C	2.88	(2.43)	0.01	(0.22)	−0.43	(−11.65)	0.15	(2.64)	−0.06	(−0.75)	−0.23	(−2.93)	0.08	(2.39)	63.4

Table VII
Conditional alpha from beta-neutral portfolios

Notes – Beta-neutral portfolio annualized mean returns and Sharpe ratio, using the all-firm sample and the KLD E ind. adj. measure, using the three-factor FF3 or six-factor FF5C conditional models (using instrumented betas), in data since Nov 2012. In parentheses are t-statistics: for means from Newey and West (1987) with three lags, and for Sharpe ratios from Lo (2003).

	Mean		SR	
	Panel A: FF3			
Missing ESG as 0	3.29	(0.97)	0.33	(0.26)
Missing ESG as −0.5	−2.77	(−0.85)	−0.28	(−0.22)
ESG nonmissing	2.14	(0.62)	0.22	(0.17)
	Panel B: FF5C			
Missing ESG as 0	−0.92	(−0.27)	−0.09	(−0.07)
Missing ESG as −0.5	−1.56	(−0.47)	−0.15	(−0.12)
ESG nonmissing	0.15	(0.04)	0.02	(0.01)

score on the sample of all firms. Our green factor has a significant average return of 8.57% per annum ($t = 2.30$). In the first row of Table VI, we report that our green factor has a significant unconditional alpha of 3.11% per annum ($t = 2.49$), which is insignificantly different from the 4.56% reported in Pastor et al. (2021a) for their green factor using MSCI data. In the second row we confirm that result is basically unchanged in the six-factor FF5C model. Thus, we support the finding in Pastor et al. (2021a) that there is a significant unconditional alpha for a green factor.

Is a green alpha reliably significant in the *conditional* model? Table VII says no. We use the modified model so that ESG drives only alpha as represented by the returns on the beta-neutral portfolio. Furthermore, we restrict attention to the three-factor FF3 and six-factor FF5C models to make a clean comparison to the results in Table VI. The beta-neutral portfolio's mean return is insignificant across both ESG-imputations or data restricted to only firms with nonmissing ESG, for both factor models. Since we are not estimating aggregate factors, these results say that our estimates of conditional betas are the key reason the ESG score no longer reliably delivers beta-neutral profits.

Nevertheless, it is worth noting that our conclusion fits nicely with the main message in Pastor et al. (2021a). There the authors find that green-returns have occurred mostly due to random shocks to ESG concerns—they find little scope for an ESG premium in expected returns. Our conditional-model results strengthen this conclusion, suggesting that green-returns were not significantly different than what one would have predicted on the basis of non-ESG information about firm risk exposures.

4.6 Relationship to Theory

We have shown that ESG measures don't reliably predict returns, and therefore we can use them 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?

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. (2020b), 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 scores.

If different investors use different ESG measures to invest, what are the equilibrium expected returns? Pastor et al. (2021b) provides an answer.[23] 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 $\tilde{g}_i$.[24] 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

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

[23]See their footnote 4 and its proof in their appendix.

[24]We set relative risk aversion to 1 and abstract from t subscripts, for simplicity.

where $w_{mkt,PST}$ is the market portfolio, $\bar{d} = \int_i \omega_i d_i di$ is a wealth-weighted average of the 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'd 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. (2021b) 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.[25] Figure 4 shows a surface plot of these densities over all the years of our data. The most striking feature of this three-dimensional surface is how the peak hovers over 0.[26] The figure says that the modal agreement between measures is essentially random. Therefore there is a large variety of different ways to measure ESG scores, and they do not agree—this can imply that equilibrium expected returns are unaffected by investors' ESG preferences.

Stepping outside the model, there are further related issues involving ESG measurement.

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

[26] 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.

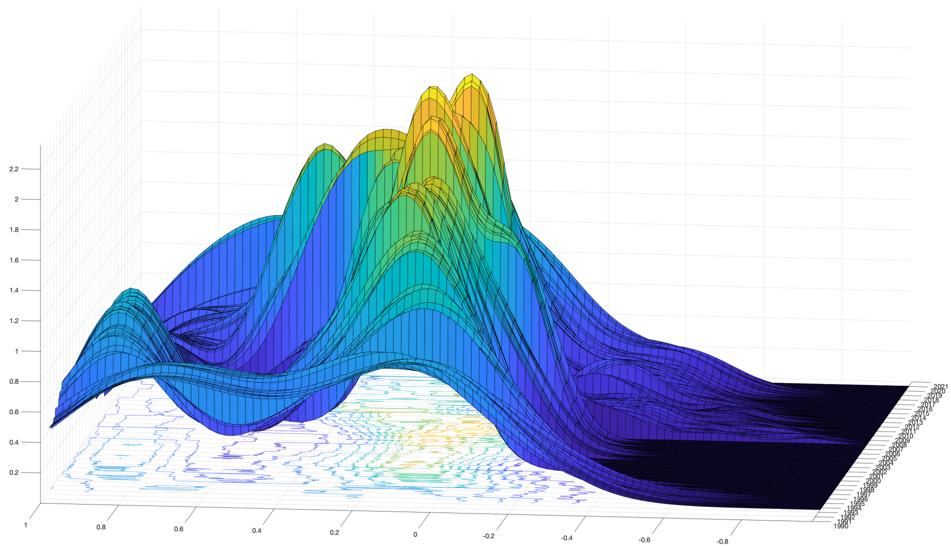

Figure 4: Densities of cross-sectional rank correlations

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

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. 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 measures and measure-providers to flourish, and perhaps exist to cater to different portfolio-managers whose underlying (i.e. absent ESG concerns) weights are different. And even if investors act as promised, the plethora of ESG metrics can lead to negligible equilibrium effects. Broadly speaking, a lack of uniform consensus (or regulation) on definitive ESG measurement is a straightforward channel explaining our results.

5 Conclusion

Using a conditional factor model with a wealth of firm information to examine the return predictability of ESG characteristics we find several notable results. First, systematic portfolio weights can be adjusted to improve the portfolio's ESG performance while sacrificing negligible profits. Second, ESG measures do not provide independently significant information about firm risk exposures. Third, ESG measures do not provide significant alpha, either alongside other characteristics or on their own. This suggests that ESG investing has essentially no cost for systematic investors. How can this occur in equilibrium? In the model of Pastor et al. (2021b), the result obtains if ESG-minded investors each measure ESG performance differently. We provide evidence that ESG measures disagree sufficiently that this may be the case.

Broadly speaking, our paper highlights the importance of widespread disagreement in measuring ESG performance, as revealed by several firms that do so. We expect our results to continue to hold in the future if this state-of-play remains. Perhaps it will, if measuring ESG goals is so difficult a process that reasonable attempts can largely disagree. On the other hand, recent speeches by policymakers at the SEC[27] and the UN Climate Change Conference (COP26) have noted that disagreement in ESG scoring could be problematic. If market-competition or regulation increases the coordination of ESG measurement, theory suggests that ESG scores may 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 we have done in this paper.

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

References

Akey, P., S. Lewellen, I. Liskovich, and C. Schiller (2021): "Hacking corporate reputations," Rotman School of Management Working Paper, Number 3143740.

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.

Barber, B. M., A. Morse, and A. Yasuda (2021): "Impact investing," *Journal of Financial Economics*, 139, 162–185.

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 (2020a): "Rewriting history II: The (un) predictable past of ESG ratings," ECGI Working Paper.

Berg, F., J. F. Kölbel, and R. Rigobon (2020b): "Aggregate Confusion: The Divergence of ESG Ratings," Working Paper, Available at SSRN 3438533.

Bessembinder, H. (2018): "Do stocks outperform Treasury bills?" *Journal of Financial Economics*, 129, 440–457.

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.

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.

Dimson, E., O. Karakaş, and X. Li (2015): "Active ownership," *The Review of Financial Studies*, 28, 3225–3268.

——— (2021): "Coordinated engagements," ECGI Working Paper.

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.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel (2020): "Hedging Climate Change News," *Review of Financial Studies*, 33, 1184–1216.

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.

Hoepner, A. G., I. Oikonomou, Z. Sautner, L. T. Starks, and X. Zhou (2021): "ESG shareholder engagement and downside risk," ECGI Working Paper No. 671/2020.

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 (2021a): "Dissecting Green Returns," Working Paper, University of Chicago and University of Pennsylvania.

——— (2021b): "Sustainable investing in equilibrium," *Journal of Financial Economics*, 142, 550–571.

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.

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.