Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows

Finance Working Paper N° 565/2018
April 2019

Samuel M. Hartzmark
University of Chicago

Abigail B. Sussman
University of Chicago

ECGI Working Paper Series in Finance

Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows

Working Paper N° 565/2018

April 2019

Samuel M. Hartzmark
Abigail B. Sussman

We are grateful to Jonathan Berk, Anat Bracha, Alex Edmans, Max Farrell, Mariassunta Giannetti, Matti Keloharju, Karl Lins, Vikas Mehrotra, Sanjog Misra, Giovanna Nicodano, Jacopo Ponticelli, Antonino Rizzo, Brad Shapiro, David Solomon, Kelly Shue, Paul Smeets, Oleg Urminksy, and Eric Zwick and seminar participants at Aalto, Emory, Cambridge, Chicago Booth, Warwick, London School of Economics, Imperial College, Northwestern, Bernstein Quantitative Finance Conference, Boulder Summer Conference on Financial Decision Making, UVA Darden Symposium on Mutual Funds and ETFs, Harvard Global Corporate Governance Colloquia, Development Bank of Japan Conference, Texas Finance Festival, European Finance Association Conference, Global Research Alliance for Sustainable Finance Conference, Swedish House of Finance Conference on Sustainable Finance for comments. We thank Halley Bayer, Nicholas Herzog, and Nathaniel Posner for excellent research assistance. We thank Ray Sin, Steve Wendel, and Sara Newcomb at Morningstar for providing the data. This work is supported by the True North Communications, Inc. Faculty Research Fund at the University of Chicago Booth School of Business. Sussman is a member of the Morningstar Behavioral Science Advisory Board. The authors have nothing further to disclose with respect to the Journal of Finance's Submission Guidelines and Conict of Interest Disclosure Policy.

Abstract

Examining a shock to the salience of the sustainability of the US mutual fund market, we present causal evidence that investors marketwide value sustainability. Being categorized as low sustainability resulted in net outflows of more than $12 billion while being categorized as high sustainability led to net inflows of more than $24 billion. Experimental evidence suggests that sustainability is viewed as positively predicting future performance, but we do not find evidence that high sustainability funds outperform low sustainability funds. The evidence is consistent with positive affect influencing expectations of sustainable fund performance and non-pecuniary motives influencing investment decisions.

Keywords: Sustainability, Behavioral Finance, Fund Flows, Mutual Funds, Salience, Rank

JEL Classifications: D03, G02, G12, G23

Samuel M. Hartzmark*
Associate Professor of Finance
University of Chicago, Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637, United States
phone: +1 773 834 7123
e-mail: Samuel.Hartzmark@chicagobooth.edu

Abigail B. Sussman
Associate Professor of Marketing
University of Chicago, Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637, United States
phone: +1 773 834 2030
e-mail: Abigail.Sussman@chicagobooth.edu

*Corresponding Author

Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows*

Samuel M. Hartzmark
University of Chicago
Booth School of Business

Abigail B. Sussman
University of Chicago
Booth School of Business

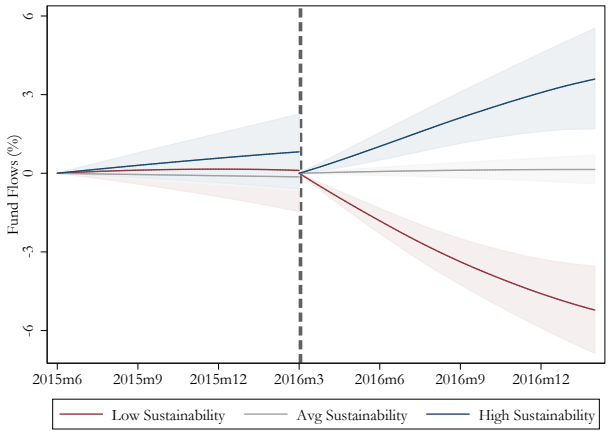
March 25, 2019

Abstract

Examining a shock to the salience of the sustainability of the US mutual fund market, we present causal evidence that investors marketwide value sustainability. Being categorized as low sustainability resulted in net outflows of more than $12 billion while being categorized as high sustainability led to net inflows of more than $24 billion. Experimental evidence suggests that sustainability is viewed as positively predicting future performance, but we do not find evidence that high sustainability funds outperform low sustainability funds. The evidence is consistent with positive affect influencing expectations of sustainable fund performance and non-pecuniary motives influencing investment decisions.

*We are grateful to Jonathan Berk, Anat Bracha, Alex Edmans, Max Farrell, Mariassunta Giannetti, Matti Keloharju, Karl Lins, Vikas Mehrotra, Sanjog Misra, Giovanna Nicodano, Jacopo Ponticelli, Antonino Rizzo, Brad Shapiro, David Solomon, Kelly Shue, Paul Smeets, Oleg Urminksy, and Eric Zwick and seminar participants at Aalto, Emory, Cambridge, Chicago Booth, Warwick, London School of Economics, Imperial College, Northwestern, Bernstein Quantitative Finance Conference, Boulder Summer Conference on Financial Decision Making, UVA Darden Symposium on Mutual Funds and ETFs, Harvard Global Corporate Governance Colloquia, Development Bank of Japan Conference, Texas Finance Festival, European Finance Association Conference, Global Research Alliance for Sustainable Finance Conference, Swedish House of Finance Conference on Sustainable Finance for comments. We thank Halley Bayer, Nicholas Herzog, and Nathaniel Posner for excellent research assistance. We thank Ray Sin, Steve Wendel, and Sara Newcomb at Morningstar for providing the data. This work is supported by the True North Communications, Inc. Faculty Research Fund at the University of Chicago Booth School of Business. Sussman is a member of the Morningstar Behavioral Science Advisory Board. The authors have nothing further to disclose with respect to the Journal of Finance's Submission Guidelines and Conflict of Interest Disclosure Policy.

Figure 1



Cumulative fund flows in percent by sustainability rating for 9 months before and 11 months after rating publication (denoted by the dashed vertical line). Estimates accumulated from local linear plot of monthly flows after removing year by month fixed effects. Shaded areas indicate the 90% confidence interval.

As firms invest more resources in sustainable and socially responsible endeavors, it is important to know whether such investments reflect investor's preferences marketwide. Some investors will believe that an increase in resources directed towards sustainability is costly and belies the primary goal of maximizing profits. Others will believe that a well run company should care about the environment or that companies should act for reasons beyond simple value maximization. Others still will value such an investment not because they inherently care about the environment, but because they view it as a sound way to maximize profit. And finally, some investors will be unaware that a firm is investing in sustainability or will not care. While surely the market contains examples of each of these investors, it remains unclear which type represents the average investor and thus it is unclear whether investments in sustainability are consistent with what investors want. Put simply, do investors collectively view sustainability as a positive, negative, or neutral attribute of a company?

This paper demonstrates that the universe of mutual fund investors in the US collectively put a positive value on sustainability by providing causal evidence that marketwide demand for funds varies as a function of their sustainability ratings. Directly addressing this question is difficult in most settings, as it is unclear how to identify the preferences of the average investor. Analysis of investment products with an explicit sustainability focus only reflects the preferences of the subset

of investors holding those products, and does not speak to the average preferences of investors in the entire market. Furthermore, market outcomes related to firm attributes, such as sustainability, are usually viewed in equilibrium where analysis is by necessity indirect.

We circumvent these challenges by examining a novel natural experiment where the salience of the sustainability of over $8 trillion of mutual fund assets experienced a large shock. Sustainability went from being difficult to understand to being clearly displayed and touted by one of the leading financial research websites, Morningstar. In March of 2016, Morningstar first published sustainability ratings where more than 20,000 mutual funds were ranked on a percentile basis and given a globe rating based on their holdings. The worst 10% of funds were rated one globe (low sustainability) while the best 10% were rated five globes (high sustainability). Prior to the publication, there was not an easy way for investors to judge the sustainability of most mutual funds without considerable effort.

Figure 1 illustrates the main finding of the paper: mutual fund investors collectively treat sustainability as a positive fund attribute, allocating more money to funds ranked five globes and less money to funds ranked one globe. Moderate ratings of either two, three, or four globes did not significantly affect fund flows. The dashed vertical line indicates the initial publication of the sustainability ratings. To the left of the line, fund flows after controlling for monthly fixed effects are accumulated over the 9 months prior to the rating publication and to the right of the line flows are accumulated for the 11 months post publication. The navy line represents five globe funds, the maroon line one globe funds, and the gray line those rated in the middle (two to four globe funds). Prior to the rating publication, the funds were receiving similar levels of flows. After the publication, the funds rated highest in sustainability experienced substantial inflows of roughly 4% of fund size over the next 11 months. On the other hand, funds rated lowest in sustainability experienced outflows of about 6% of fund size. Over the 11 months after the sustainability ratings were published, we estimate between 12 and 15 billion dollars in assets left one globe funds and between 24 and 32 billion dollars in assets entered five globe funds as a result of their globe rating.

Our experiment is rare in financial markets in that it examines a large quasi-exogenous shock,

equivalent to approximately 40% of NYSE market cap, that does not directly impact fundamentals. The shock yields easy to understand measures of sustainability by simply repackaging publicly available information in a form that attracts attention and is easy to process. Further, the construction of the measure is based on within-category comparisons that rely on Morningstar's own classification of funds, so it is unlikely to be highly correlated with investment style or other general measures of sustainability.[1] Thus our measured response is to the rating itself, not to new information about fund fundamentals. In addition, examining mutual funds rather than individual stocks allows us to directly observe fund flows. This allows us to avoid focusing on indirect measures, such as prices, which suffer from the joint hypothesis problem that they could be capturing risk.

This shock allows us to identify the causal impact of the globe rating along a variety of margins. If funds were systematically different before the publication of the ratings, then flows could be reflecting this difference. The initial figure suggests this is not the case, as do a variety of robustness checks including a matching exercise on pre-publication characteristics and a placebo test.

The globes are a discrete rating system of five categories, though Morningstar also released each fund's sustainability score and the percentile ranks underlying the ratings. If investors responded to the five globe rating system rather than to other aspects of sustainability, we should find that the globe category itself drove the fund flows. Examining the percentile ranks that underlie the sustainability rating, we find evidence consistent with discontinuities at the extreme globe category edges, but find minimal impact of either the percentiles themselves or the sustainability scores. This suggests that investors focused on the simple globe rating and largely ignored the more detailed sustainability information.

We find strong flow effects from being in the two extreme globe categories (i.e., one or five globe funds) relative to the three categories in the middle, but find insignificant differences across funds receiving two, three, or four globe ratings. This is consistent with prior evidence that investors often focus on discrete rather than continuous measures and that when they do so they focus on extreme

[1] Put another way, *Barron's* noted that funds rated high sustainability by Morningstar were not "whom you'd associate with even a faint whiff of patchouli." http://www.barrons.com/articles/the-top-200-sustainable-mutual-funds-1475903728

outcomes (e.g. Hartzmark 2015; Feenberg et al. 2017).[2] It underscores the general importance of salience on investment decisions (e.g. Bordalo et al. 2012; Bordalo et al. 2013a) as well as the impact of attributes that stand out in consumer choice (Bordalo et al. 2013b). These findings suggest that evaluating information based on extreme ranks reflects a fundamental cognitive process underlying decision making that impacts the market.

The large causal flow response we observe in the short-term allows us to reject both the hypothesis that investors are indifferent to sustainability as well as the hypothesis that they view sustainability as a negative characteristic, but it leaves open the question of which specific aspect of sustainability drove investors to reallocate funds from one globe funds to five globe funds. While we are unable to definitively pinpoint the specific motive, we explore three possibilities. The first is that institutional pressure, either to hold high sustainability stocks or not to hold low sustainability stocks is responsible for the results. We find that fund flows from institutional share classes in response to the globe rating are similar to those from other share classes. This could be evidence that investors in institutional share classes face constraints that force them to behave like other investors, or that their preferences are similar to that of other investors. Since non-institutional share classes display a similar pattern, institutional constraints cannot fully account for the finding.

Another possible explanation is that investors rationally view a rating of high sustainability as a signal of high future returns. We examine whether funds experienced high returns after their high sustainability ratings relative to a variety of benchmarks and find evidence more consistent with the opposite or no relation. While it is difficult to make definitive statements using only 11 months of data, we do not find evidence of high returns for high sustainability funds.

If the results are not driven purely by institutions or a rational belief in higher expected returns,

[2]More broadly, our findings are consistent with literature in psychology and economics that model rank dependent preferences (e.g., cumulative prospect theory; Tversky and Kahneman 1992), and with the corresponding intuition that extreme ranks are the most perceptually salient positions (Diecidue and Wakker 2001; Tversky and Kahneman 1986). See also Quiggin (1982) and Schmeidler (1989) for early rank-dependent models of risk under uncertainty and Weber and Kirsner (1997) for an examination of why people rely on extreme rank in evaluations. Furthermore, it is consistent with existing literature showing that people overweight extreme attributes when making judgments about people (Skowronski and Carlston 1989) or when evaluating outcomes (Sussman and Shafir 2012; Sussman 2017) and make choices to avoid products with attributes ranked in extreme positions when confronted with tradeoffs (Simonson and Tversky 1992; Tversky and Simonson 1993).

then some investors want to hold high sustainability funds and avoid low sustainability investments either due to an irrational belief that there is a positive correlation between future returns and sustainability or for non-pecuniary motives (such as altruism, warm glow or social norms). Unfortunately the data does not allow us to distinguish between these two possibilities, so we run an experiment using MBA students and MTurk participants. We elicit expectations about future performance, risk and investment decisions as a function of globe ratings. We find a strong positive relation between globe ratings and expected future performance and a strong negative relation between globe ratings and expected riskiness. This pattern of an inverse relation between expectations of risk and returns is consistent with judgments based on affect, rather than reason (e.g., Slovic et al. 2004, 2005, 2007; Finucane et al. 2000). We also find some evidence of non-pecuniary motives across both populations. Participants considering environmental or social factors when making their decision invest more money in five globe funds and less money in one globe funds than their performance and risk expectations can account for, while those not considering such factors do not exhibit such a pattern. The results suggest that globe ratings impact expectations of future performance and also lead investors to make choices based on non-pecuniary motivations.

Our paper contributes to the literature that has examined how investors value non-financial aspects of stocks. While other studies have examined how subsets of investors value characteristics of securities, such as whether it is a "sin" (Hong and Kacperczyk 2009), local (Huberman 2001) or offers a certain dividend yield (Harris et al. 2015, Hartzmark and Solomon 2019), our study has the benefit of examining a quasi-exogenous shock which means we can measure how all mutual fund investors collectively value the characteristic, rather than the subset that hold the security. Perhaps most closely related to our paper, Hong and Kacperczyk (2009) find that sin stocks yield higher returns, consistent with investors needing to receive a premium to hold these companies due to social norms. Our paper complements this finding by examining an exogenous shock to a significantly larger portion of the market with a more direct measure of demand.

A recent literature has examined the rapidly growing set of investment products with explicit mandates of social responsibility (e.g. Bialkowski and Starks 2016; Barber et al. 2017; Benson

and Humphrey 2008; Bollen 2007; Geczy et al. 2005; Riedl and Smeets 2017; see Renneboog et al. (2008) for a review). While understanding the preferences underlying such investments represents an important area of research, it is only indicative of the investors selecting into this subset of products (roughly 2% of funds in our sample) and need not be representative of investors or funds marketwide. If a small subset of investors had strong preferences for sustainability while most investors did not directly value sustainability, under standard models (e.g. Berk and Green 2004) we would not find an effect of the ratings on net flows. If the market was largely comprised of the investors of Berk and Green (2004) who solely maximized profits, these investors who did not value sustainability would undo the effects of investors with a preference for sustainability, resulting in zero net flows. On the other hand, if most investors value sustainability, investors seeking only to maximize profit would not be numerous enough to fully offset the flows from the sustainability seeking investors. One interpretation of the flow response we observe is that the relative size of purely profit maximizing investors in the US mutual fund market as a whole is smaller than that of investors with a preference for sustainability. Thus our paper contributes to this literature by examining the preferences for sustainability in the universe of US mutual fund investors into products lacking explicit sustainability goals.

Additionally, our paper contributes to the literature on why firms invest in sustainability, and more broadly to investment in "doing well by doing good."[3] Some sustainable investing is clearly due to agency issues (Cheng et al. 2013) while others have argued that it is consistent with efficient investment, for example by improving morale (Edmans 2011). As emphasized by Hart and Zingales (2017), investments for non-pecuniary "pro-social" reasons, such as sustainability, are something that companies should engage in if they reflect the preferences of their shareholders. While our paper does not break down the fraction of sustainability that is due to agency versus appeasing shareholders, a general demand for sustainability from mutual fund investors suggests that a significant portion of the observed investment in sustainability is not purely due to agency issues.

Finally, the evidence highlights the potential role of emotion in guiding investment decisions.

[3]For recent overviews see: Bénabou and Tirole (2010); Heal (2005); Kitzmueller and Shimshack (2012); Margolis et al. (2009); Christensen et al. (2017); Chowdhry et al. (2017).

Although it may seem surprising that higher globe funds are associated with expectations of both higher returns and lower risk, this pattern is consistent with research on the affect heuristic (e.g., Slovic et al. 2004, 2005, 2007; Finucane et al. 2000), which finds that feelings associated with a given stimulus often take the place of more reasoned analysis and guide subsequent judgments and decisions about the stimulus. While the affect heuristic has been prominent within the psychology literature in discussions of risk evaluations, its role has primarily been examined in laboratory tasks, and it has received minimal attention in the context of financial products.[4] Thus, an additional contribution of the current work is to highlight the consequential role of affect versus analytic thought in financial decision making and financial markets as a whole.

1 Sustainability Ratings

On March 1, 2016 Morningstar launched the Morningstar Sustainability Rating. The company classified more than 20,000 mutual funds, representing over $8 trillion dollars in market value, into a simple rating between one and five globes. The rating system was designed to provide "a reliable, objective way to evaluate how investments are meeting environmental, social, and governance challenges. In short, it helps investors put their money where their values are."[5]

The classification system is based on the underlying holdings of a given mutual fund. Each holding is given a sustainability score based on research of public documents undertaken by the company Sustainalytics. This rating is related to how a firm scores on environmental, social and governance issues (ESG). At the end of each month, Morningstar takes the weighted average of this measure based on holdings to form a mutual fund specific sustainability score.[6] Each fund in a Morningstar category is ranked based on its sustainability score and this ranking serves as the basis of the Morningstar globe ranking. According to the documentation, a fund is given five globes and rated as "High" if it is in the top 10% of funds in the category. It is given four globes and rated as

[4]For exceptions see Hirshleifer and Shumway (2003) examining the role of sun exposure on market movements or Birru (2017) examining risk taking and anomaly predictability based on shifts in mood throughout the week.

[5]http://news.morningstar.com/articlenet/article.aspx?id=745467

[6]Complete details of the methodology can be found at: https://corporate1.morningstar.com/Morningstar-Sustainability-Rating-Methodology-2/

"Above Average" if it is ranked between 10% and 32.5%. It is given three globes and rated "Average" if it is ranked between 32.5% and 67.5%. It is given two globes and rated "Below Average" if it is ranked between 67.5% and 90%. It is given one globe and rated "Low" if it is ranked in the bottom 10% of its fund category.[7] The globe ranking is prominently reported using pictures of one to five globes as well as the descriptive label (e.g., "High") on each fund's Morningstar page. The percentile rank in category and raw sustainability score are displayed in smaller text alongside the rating, see Figure 2 for an example.

While Morningstar's definition of sustainability is a precise formula transforming holdings and ESG ratings into a globe rating, "sustainability" has generally become a popular term that lacks a clear and consistent definition. An investor who wished to understand the details of Morningstar's system could easily do so, but it is likely that a number of investors responded based on their preconceived notion of the meaning of sustainability rather than to the specific details of the rating methodology. Thus it is useful to more precisely understand how investors interpret sustainability.

Therefore, we recruited 482 participants from an online sample and asked them which elements of a company's business practices they believe "sustainability" refers to.[8] The results are reported in Table 2. The dominant answer was that sustainability relates to a company's environmental practices, with 79% of participants including environmental issues in their definition of sustainability. Participants included a number of other aspects as well, but none other garnered more than 50% of responses. In total, participants listed 2.7 items on average, with less consistency in the selection of the additional items.[9] While the meaning of sustainability varied across participants, there was not confusion as to what any given participant's definition was. Only 2% of participants listed that they did not know what was meant when a company's business practices became more sustainable.

[7] A coding error included 11% of the data in the one globe category.

[8] Participants selected as many options as desired from the following list: Corporate Governance, Community, Diversity, Employee Relations, Environment, Human Rights, Products, Other, and I don't know. We chose these options because they are the dimensions by which KLD Research & Analytics, Inc, a leading provider of social investment research, evaluates companies on environmental, social, and governance issues.

[9] e.g., the next most popular item, product quality and safety, was listed by only 48% of people.

2 Data Sources and Summary Statistics

All of the mutual fund data was directly provided to us by Morningstar. This historical dataset is proprietary and is not currently publicly available. The data was anonymized of fund specific identifiers by Morningstar, which means that we are unable to link the data to other publicly available data sources, such as holdings, and cannot update our sample. The sample includes all US domiciled open-end funds with a sustainability rating from Morningstar and is at the monthly frequency. Since Morningstar's ratings were meant to cover the entire universe of such funds, the data runs the gamut of Morningstar categories.[10] The data is provided at the share class level, but the analysis is conducted at the fund level. Fund size (TNA), dollar flows and web traffic are calculated as the sum across share classes, while expense ratios and returns are the mean. Morningstar "star" fund ratings are the rating of the largest share class and fund age is calculated from the inception date of the earliest share class. Morningstar category names sometimes vary slightly within a fund across share classes. We remove these share class specific attributes to form consistent categories within and across funds. We limit the sample to funds with TNA above one million dollars and winsorize continuous variables at the 1% level.

Flows are the main variable of interest in the paper and are measured as monthly dollar flows divided by TNA at the end of the prior month.[11] Flows are noisy and may systematically vary based on characteristics, such as size. To make sure the results are not driven by such properties, we examine a normalized flow variable. To construct this variable, each month we split firms into deciles based on size and assign each fund to percentiles based on flows within each size decile. This normalized flow variable is inoculated from differences in flows by fund size and from outliers.[12]

Table 1 Panel A shows summary statistics for the funds after the publication of the sustainability ratings in March of 2016 through January of 2017. In Table 1 Panel B we show the summary

[10] These range from standard style box categories, such as Large Growth, to an international focus, such as Europe or China. There are also examples of more exotic styles such as long-short strategies and sector specific investments, such as energy or health.

[11] While it would be interesting to examine whether the effect comes from new buyers, or the selling behavior of existing shareholders, our anonymized data limits the analysis to the net effects of both.

[12] We thank an anonymous referee for suggesting this methodology to construct the variable.

statistics prior to the globe publication for each globe ranking, where globe is what each fund was eventually assigned in March 2016. Both one and five globe funds tend to be smaller, which could be due to the sustainability rating becoming less extreme for funds with more diversified holdings. Examining flows, web traffic and Morningstar star ratings, we do not observe systematic differences across funds by globe rating.

In Table 1 Panel C, we examine the same variables during the publication period. Over this period mutual funds experienced outflows of -0.4% per month on average, but the funds rated lowest in sustainability experienced outflows of -0.9%, while flows to funds rated highest in sustainability were nearly zero. Also, examining web visits, we see that the lowest amount of web traffic was received by funds rated one globe, while the highest rated funds in sustainability received substantially more traffic than the other funds. Finally, consistent with the flows, we see that one globe funds shrank while five globe funds grew relative to their pre-publication average.

In Table 1 Panel D we examine the probability of moving to a different globe category. The sample is restricted to the post-publication period, excluding the first month where no switching was possible. In general, if a fund is ranked as a given number of globes, there is a roughly 80% chance that it will have the same rating the next month. Funds that do change categories rarely change more than one category in a given month.

3 Do Investors Value Sustainability?

3.1 Attention to Ratings

Although Morningstar created its sustainability ratings because it believed there would be investor interest in them, one reasonable hypothesis is that they did not receive attention when published and thus had no impact. This could be because investors did not care about the rating, did not know about the rating, or were already aware of the information contained in the rating. The Sustainalytics score for each company was based on publicly available information and the Sustainalytics scores themselves were also publicly available, for example through Bloomberg. Further,

fund holdings were publicly reported. Thus, all of the information used to construct the globe ratings was available before the publication of the ratings. Perhaps investors already understood the information that Morningstar aggregated into a globe rating and the ratings were simply ignored.

We provide evidence based on Google searches that the globe rating system attracted significant attention at its launch, but not prior to its launch. Figure 3 shows the relative interest of monthly Google searches using Google Trends data for "Morningstar star rating" versus "Morningstar sustainability rating."[13] The star rating refers to Morningstar's popular fund rating system. Its search intensity is represented by the navy line. The maroon line represents searches for "Morningstar sustainability rating" while the vertical gray line represents the first publication of those ratings.

There are two notable aspects of Figure 3. First, before their publication, there was no measurable volume of searches for the sustainability ratings. This suggests that their publication was not anticipated, at least not by Google users. Second, subsequent to their publication, there were roughly as many Google searches for the sustainability rating as there were for the star rating. This is consistent with there being significant interest in the sustainability ratings, which were publicized through white papers, traditional marketing campaigns, included as a search filter option for some Morningstar clients, covered by outside media outlets and included on every rated fund's Morningstar web page. The large search volume suggests many investors became aware of the existence of the rating and were likely interested in issues related to sustainable investing.[14]

The paper's focus is on investor's perception of sustainability. For the ratings to provide a valid test of this mechanism, investors cannot have systematically sorted into funds based on their future rating before publication. The search frequency and subsequent findings suggest that the publication of the ratings induced the flow response by investors. While investors did not respond to the ratings before their publication, it is possible that mutual funds predicted their publication

[13]The monthly measure is the average of the weekly searches, where month is assigned based on the month that a given week ends. Google trends normalizes the results of every search to a different scale with the maximum search volume in a week for the term with the highest intensity normalized to 100 at its maximum. The results in Figure 3 are from a search that included both terms so the magnitudes are comparable between the two measures.

[14]Search volume may be elevated in the period directly after the launch of the ratings as a result of media attention surrounding the launch and the ratings system. Our results should be interpreted within this context.

and traded prior to the publication in an attempt to receive a high globe rating.[15] If such behavior was widespread, this would potentially impact the interpretation of the results related to the cause of return predictability (discussed in Section 4.2), but would not change the interpretation of the paper's core results related to fund flows and investor preferences.

3.2 Base Results

Did the publication of the sustainability ratings impact how investors traded mutual funds? To begin answering this question we examine the mutual fund flow reaction to the publication of the ratings. The ability to study flows makes mutual funds an ideal laboratory to examine the revealed preferences of investors. If a fund is generally viewed as more desirable after its rating becomes public, money will flow to it and it will grow. If it is viewed as less desirable than we will see money flow from it and it will shrink. This stands in contrast to individual stocks which have a fixed supply in the short run, and therefore does not allow for such a direct measure of investor response.[16]

In addition, our setting is rare in financial markets in that we examine an event that does not change fundamentals. Studies of socially conscious investing generally focus on fixed firm specific traits. For example, a tobacco company tends to remain a tobacco company, and any change to such a characteristic would represent a large shift in its business. Our study examines a shock to the salience of a characteristic, so while the characteristic is fixed, there is no change to the underlying business by the publication of the fund rating.

When Morningstar published its ratings, it displayed three separate measures of sustainability together on a fund's page as shown in Figure 2. It released a fund's raw sustainability score, the percentile rank of that score within the fund's Morningstar category, and a picture of how

[15]For example, Sustainalytics announced that they had licensed their ratings to be used by Morningstar for sustainability prior to the ratings publication (https://www.sustainalytics.com/press-release/morningstar-to-launch-first-environmental-social-and-governance-esg-scores-for-funds-globally/).

[16]Prior to the ratings publications it was difficult to ascertain a fund's sustainability without considerable effort. An exception to this is the small subset of funds, roughly 2% of our sample, with an explicit sustainability mandate. The Internet Appendix shows no significant flow variation for these funds based on globe ratings. We do not focus on such funds due to the small sample size and because investors had sorted into these funds based on sustainability prior to the Morningstar ratings. For papers examining these funds see Bialkowski and Starks (2016); Benson and Humphrey (2008); Bollen (2007); Geczy et al. (2005).

many globes the fund was rated based on percentile rank cutoffs. If investors want to invest in the most sustainable fund in the market overall, then the raw sustainability score is the most informative measure, but it is difficult to interpret without a significant amount of effort dedicated to understanding its scale. The percentile rank variable yields a continuous measure of within Morningstar category rank available to investors that is easier to interpret than the raw sustainability score and provides more granular detail than the globe rating. If investors want to invest in the most sustainable fund in a given Morningstar category, then the percentile rank is the most informative measure. As shown in Figure 2, the globe rating is given the most space on the sustainability portion of a fund's webpage. It is presented as a large picture of the number of globes along with the corresponding rating label (e.g. High, Average or Low) in a larger font than either of the two measures. However, all of the information needed to determine the globes is included in the percentile rank variable. If investors are paying attention to the available percentile information, there is no need to pay attention to the globe rating. If investors' attention is drawn to the globe rating itself, they may simply examine this salient measure and ignore the underlying percentiles.

In Table 3, we explore the reaction to each sustainability measure and find that it is the globes, rather than the other measures that are the main driver of fund flows. We regress fund flows on each sustainability measure and include Morningstar category by year by month fixed effects to control for time variation by category. In Column 1, we examine the raw sustainability score and the percentile rank in category variables and we see insignificant coefficients on both. In Column 2 we include dummy variables for each globe rating omitting the three globe category. One globe funds, the funds rated worst in terms of sustainability, experienced outflows of roughly -0.44% per month lower than three globe funds, with a t-statistic of -2.80 clustered by month and fund. Five globe funds, those rated highest in terms of sustainability, experienced inflows of 0.30% per month more than three globe funds, with a t-statistic of 1.81. These point estimates indicate that the lowest sustainability funds lost 5.4% of TNA per year while the highest rated funds gained about 3.6% of TNA per year. Below the regression results is the difference between one and five globe funds, of 0.74 per month with the p-value on the test that they are equal, 0.004, underneath. The

globe ratings in the middle – two and four globes – are not statistically distinct from the omitted three globe funds.

The insignificance of the two and four globe funds suggests that investors focus on extreme one and five globe categories. If this is the case then the relevant test is how one and five globe funds compare against those rated in the middle. Column 3 conducts such a test, where two, three and four globe funds comprise the omitted category. One globe funds see outflows of -0.46% per month lower than middle ranked funds with a t-statistic of -3.21 while five globe funds see inflows of 0.28% higher than middle ranked funds with a t-statistic of 1.87.

The prior results may be due to globe ratings systematically varying with other variables associated with flows, so in Column 5 we add a number of controls. We include the prior month's return, the prior 12 month return and the prior 24 month return to control for the fund-flow relation (Chevalier and Ellison 1997). To make sure the globe ratings are not simply capturing fund-flows based on size, we control for the log of fund TNA the prior month. We also add controls for the expense ratio and for log of fund age. There could be a correlation between Morningstar's globe rating and their star ratings, so we also control for the star rating. After including these controls, we find similar effects. In Column 5, one globe funds are associated with outflows of -0.40% with a t-statistic of -3.38, while five globe funds had inflows of 0.33% with a t-statistic of 2.35.

In Column 6 we include all three of the variables and find that investors respond to the coarse globe ratings, not the other two variables.[17] After including the globe rating variables, the coefficients on both the category percentile rank and the raw sustainability score are insignificant. The coefficients on globe ratings are materially unchanged. The one globe variable is negative and significant while the five globe variable is positive and significant. The regression suggests that investors responded to the globe ratings, not the other measures of sustainability. In all specifications the shift in flows is above 0.7% per month moving from one to five globe funds.

In Panel B we examine the normalized flow variable to address the concern that the results are driven by systematic noise over the short sample. If the results are driven by outliers or small firms

[17] One potential concern is the results are due to a high correlation between the raw score and percentile rank. In the internet appendix we repeat the regression with each variable separately and find similar results.

with volatile flows, rather than the sustainability ratings, the results will decrease, or disappear in this specification. If the measure is reducing noise that attenuated the estimates using raw flows, the relation will be stronger in this specification.

The first two columns of Panel B shows the results become stronger using the normalized flow variable. Examining Column 2, which includes the additional controls, one globe funds have flows 4.4 percentiles lower than middle ranked funds with a t-statistic of -4.62 while five globe funds have inflows 3.3 percentiles higher than middle ranked funds with a t-statistic of 3.19. The spread of 7.7 percentiles between one and five globe funds has a p-value of 0.0004. Reducing the noise in flows using this normalization significantly increases the statistical significance of the results, consistent with a strong response by investors based on the globe ratings themselves.

Another concern is that the regressions are driven by small, economically unimportant funds. In columns 3 through 6 we repeat the analysis weighting the regressions based on the log of fund size the prior month. For both measures the results are similar and get slightly stronger in point estimates. For the flow measure including controls, one globe funds underperform middle ranked funds by -0.40% with a t-statistic of -3.41 and five globe funds outperform middle ranked funds by 0.33% with a t-statistic of 2.38. The spread between the two of 0.73% has a p-value of 0.0002. Examining the normalized measures in Column 5 and 6, after including the additional controls one globe funds had outflows of -4.4 percentiles with a t-statistic of -4.71 while five globe funds received inflows of 3.3 percentiles with t-statistics of 3.25. The difference between the two of 7.7 percentiles has a p-value of 0 to four decimal places.

3.3 Within Globe Rating Analysis

The results suggest that investors focus on the extreme globe ratings and largely ignore both the middle globe ratings and the available underlying sustainability information. If so, funds within a globe rating should receive a similar level of flows, regardless of how different they are based on the more detailed sustainability information. Further, investors should treat funds with similar sustainability characteristics that happen to fall on different sides of an ad-hoc globe rating breakpoint

quite differently, leading to discontinuities in flows around the category edges. Finally, these effects should be concentrated in the extreme one and five globe categories, not the three in the middle.

Figure 4 allows us to explore these hypotheses by taking a more detailed look at the relation between fund flows, the globe rating and the underlying percentile ranks. Panel A shows the average fund flow for each percentile rank from 1 through 100 after removing a year by month fixed effect. Panel B repeats the analysis using the normalized flow measure. The dashed vertical lines indicate the globe cutoffs with the globe category listed at the top of the chart. The bars to the extreme left are five globe rated funds while those to the extreme right are one globe funds. Examining each percentile separately limits our power as each bar is populated by roughly 350 observations. Examining the ten percentiles assigned to high sustainability funds (5 globes) nine of the ten point estimates are positive and five of the ten are positive and significant at the 90% level. Examining the 11 percentiles assigned to low sustainability funds (1 globe) all eleven are negative and five of the eleven are negative and significant at the 90% level. Looking at the two, three and four globe categories, there is a mix of positives and negatives throughout, with no discernible pattern. Of these 79 percentile ranks, only seven are significant at the 90% level, less than the ten significant percentiles in the 21 extreme percentile categories. Panel B repeats the analysis with the percentile rank measures and the results are if anything stronger. Six of the five globe percentiles are positive and significant while nine of the one globe percentiles are negative and significant. Across all other percentiles there are seven that are significant. The evidence suggests investors responded to the one and five globe categories, largely ignoring the 2, 3 and 4 globe categories.

While Figure 4 presents evidence suggesting that the extreme globe ratings are largely responsible for the observed flows, it also suggests that percentile ranks were not altogether ignored. The major exception where flows appear to be different based on percentile ranks, but not at globe cutoffs, is the extreme low sustainability funds which received higher outflows when ranked 98th and above. Comparing the flow in percentiles 98 and above to the other one globe funds yields a difference of -0.51 with a t-statistic of 3.08. Examining the normalized measure yields an estimate -7.2 percentiles lower with a t-statistic of -8.37. The effect of being in the top percentiles of high

sustainability funds is more muted. The top 3 percentiles for 5 globes have inflows 0.35 higher with a t-statistic of 3.64, while the normalized measure shows these funds receive inflows 3.4 percentiles higher with a t-statistic of 2.51. Thus it appears that investors again pay attention to the extreme ranked funds by percentile, but only for the most extreme ratings of sustainability.

If investors are responding to the globe ratings, the ad-hoc choice of cutoff will leave similar mutual funds receiving different ratings on either side of the cutoff. We examine this question more formally in Table 4 using regression discontinuity analysis. We use the rank within each category as the running variable. For example, in June of 2016, there were 265 funds ranked in the US based Emerging Market funds category, and the top 26 were ranked as 5 globes. Thus, we look at the break point of the five globe funds ranked just below 26 compared to the lower globe funds with ranks greater than 26 by running discontinuity tests (e.g. Thistlethwaite and Campbell 1960; Imbens and Lemieux 2008 and DiNardo and Lee 2011). Panel A examines flows around one globe breakpoints and Panel B examines flows around five globe breakpoints.

Table 4 presents a series of estimates of the discontinuities surrounding the globe cutoffs. For each estimate, we provide a conventional estimate and an estimate correcting for the bias as described in Calonico et al. (2014). We select bandwidths using the methods described in Calonico et al. (2014) and Calonico et al. (2018) and include monthly fixed effects, using the methods described in Calonico et al. (2018).[18] We present each estimate using uniform bandwidths in odd columns and using separate bandwidths in even columns. In the first two columns of Panel A, the four point estimates based on the simplest specifications (without clustering) range from -0.319 to -0.536 with z-statistics based on the heteroskedasticity-robust nearest neighbor estimator from Calonico et al. (2014) ranging from -2.02 to -2.85. This suggests that moving from a two globe rating to a one globe rating leads to a decrease in flows of roughly 0.4% per month. Examining the same specifications for the five globe discontinuity in Panel B, the four point estimates range from -0.503 to -0.821 with z-statistics ranging from -2.56 to -3.27 which suggests that moving from a five globe rating to a four globe rating leads to a decrease in flows of roughly 0.6% per month. Each of these eight estimates

[18]The internet appendix repeats the analysis on residuals controlling for month by category fixed effects.

is consistent with discontinuities around the one and five globe breakpoints.

A variety of ad-hoc choices are involved in conducting a regression discontinuity test, so we present a number of alternative estimates to show that no particular choice is driving the results. Lee and Card (2008) show that standard errors can be misspecified with a discrete running variable, and they suggest clustering by the running variable to account for such an effect. Following this suggestion, Columns 3 and 4 cluster by category rank (the running variable in our analysis) and find similar estimates that are all significant at the 95% level.[19] We re-run the analysis clustering by fund in columns 5 and 6. For the one globe breakpoint, the point estimates are slightly smaller in absolute value with one insignificant (z-statistic of -1.42), two significant at the 90% level (z-statistics of -1.66 and -1.75) and one significant at the 95% level (z-statistic of -1.98). The five globe estimates are more negative, ranging from -0.477 to -0.807, with one estimate significant at the 90% level (z-statistic of -1.81) and the other three estimates significant at the 95% level with z-statistics more extreme than -2.14. In the final two columns, estimates are clustered by month. Each of the eight estimates is slightly more negative than the other columns with parallel bandwidth specifications, and significant at the 95% level with z-statistics more extreme than -2.00 for one globe funds and more extreme than -3.09 for five globe funds. The internet appendix repeats the analysis using normalized flows, flows removing a category by month fixed effect and normalized flows removing a category by month fixed effect. While there is some variation in the magnitude and statistical significance of the estimates, the evidence is consistent with discontinuities at the one and five globe breakpoints.

The paper presents a variety of evidence suggesting that investors respond to the coarse globe ratings. In Table 3 there is not a significant flow response to the raw sustainability scores, percentile ranks or middle globe categories. In Figure 4 there is no discernible pattern in the flow response to percentiles for the middle ranked globe funds. In Table 4 there is evidence of discontinuities surrounding the one and five globe breakpoints. In the Internet Appendix we present evidence that funds receive more extreme flows in months when they are ranked one or five globes compared to

[19] The choice of clustering influences bandwidth selection and the bias correction so point estimates are not identical under different clustering assumptions in the Calonico et al. (2018) framework.

months when they are not. The exception noted above, is that investors appear particularly averse to the absolute least sustainable funds based on percentile ranks. Taken as a whole, the results emphasize that the formation and display of information as categories can have a significant impact on investor decision making.

3.4 Controlling for Pre-period

The prior section showed that there was a high correlation between globe ratings and flows. One still may be worried though that the prior section simply captured pre-period differences in funds that were not addressed by these specifications. In this section we examine whether the globe ratings were capturing such pre-period effects and find that it is unlikely to be the case.

Figure 1 examines cumulative flows based on globe ratings, both before and after their publication. The globe ratings did not exist before they were published, so for the period before their publication every fund is assigned their first globe rating from March 2016. Raw flows are regressed on year by month fixed effects to control for time trends. The estimates are from a local linear plot are accumulated to form the plot for the 9 months before and 11 months after the rating's publication. Before publication, to the left of the dashed line, there are not significant differences across the groups and the trends are roughly similar. After the publication, we see significant increases in flows to funds rated five globes and significant outflows from funds rated one globe.

Figure 5 examines this further presenting the raw averages for each month along with a version of the local linear plot figure without accumulating the flows. Examining the simple averages during the pre-period in Panel A, there is not a clear relation. Four of the nine pre-period months have higher flows to funds that will be rated one globe than to funds that will be rated five globes with the other five having the opposite pattern. The smoothed local linear plots in Panel B presents evidence consistent with these patterns as there is not a significant difference across globe categories in the pre-period. The confidence intervals for all three categories are overlapping in each month.

After publication, the pattern becomes stronger and less volatile. The gap between the blue dots and the red dots becomes more extreme and the white space between the red and blue lines

becomes significantly greater. Every month post publication the five globe funds have higher inflows than the one globe funds. The results are consistent with flows being impacted by the ratings and the funds being broadly similar before the ratings were published.

We examine this pattern more formally in Table 5 by matching funds based on their characteristics in the period before rating publication. Funds are examined based on the intent to treat, so the globe category they were initially assigned to in March 2016 is assigned for all 11 months subsequent to publication. Funds in an extreme rank are matched to other funds that had the same Morningstar star rating as of the month prior to the rating publication. A nearest neighbor match is used based on flows, size, age and return prior to the ratings. Using this method, the results suggest that one globe funds had outflows of -0.72% (t-statistic of -9.07) which were -6.7 percentiles lower (t-statistic of -11.60) using the normalized measure. Five globe funds had inflows of 0.21% (t-statistic of 2.60) or 3.8 percentiles higher (t-statistic of 7.44) using the normalized measure.

While the analysis matches on observed fund characteristics, there is always a concern that we are omitting a relevant variable. Thus in Panel B we additionally match on the fund's pre-period loadings on orthogonal projections of Vanguard benchmarks (see Section 4.2 for details of their estimation). To the extent that similar funds covary together on a wide variety of possible characteristics, this should control for the characteristics not explicitly included in the match. Results are similar after matching on these loadings. One globe funds experience outflows of -0.52% relative to the matched sample and five globe funds experience outflows of -0.19% per month. The results suggest that pre-period differences do not account for the results.

The internet appendix contains additional analysis ruling out further possible concerns. To examine whether the results are due to a general trend related to sustainability we construct pseudo globe ratings based on KLD scores in the years prior to the Morningstar rating publication. The pseudo ratings do not predict fund flows. While Morningstar ratings are sticky, they are recalculated every month and funds do change categories. The internet appendix shows that funds experience more extreme flows when they possess the extreme rank compared to months that they do not.

In order for our results to be capturing something other than the impact of the globe ratings,

the ratings would have to be correlated with some other variable which is accounting for flows. This variable would have to be related to the discrete globe ratings to account for the discontinuity analysis, but not the underlying sustainability score or more continuous percentile ranks. The alternate variable could not be capturing fixed fund attributes, as we find the effect is stronger when funds are ranked high or low in sustainability than in months when they are not. The variable must not be captured by our explicit controls, or correlations on factor loadings and must begin having its impact only when the ratings are published as the placebo analysis showed it was not present before. While these alternatives are not impossible, we feel that the results strongly support the parsimonious explanation that the globe ratings had a causal impact on fund flows.

3.5 Economic Impact

The inflows to five globe funds and outflows from one globe funds provide evidence that investors on average view sustainability as a positive attribute. While statistically strong, how economically meaningful was the impact of the globe ratings?

We conduct a back of the envelope analysis to estimate the overall impact. We take all funds with a five globe or a one globe rating and multiply their prior month TNA by the regression coefficient. This serves as an estimate for how much higher or lower the flows were because of a globe rating. Examining Table 3, for one globe funds the smallest regression coefficient is -0.352 while the largest is -0.457. Using these estimates, one globe funds lost between 12 and 15 billion dollars in outflows in the 11 months after publication. Using the range of estimates for five globe funds where the smallest coefficient is 0.281 and the largest coefficient is 0.379, five globe funds received inflows of between 24 and 32 billion dollars as a result of their globe ratings.

Another metric for evaluating the magnitude of the ratings is by comparison to the impact of the Morningstar star ratings, which Del Guercio and Tkac (2008) argue are the "undisputed market leader" for fund ratings which are "arguably the primary inputs to many investors' decisions." Reuter and Zitzewitz (2010) find that moving up one star rating results in 43 to 52 basis points higher flows per month. Thus the impact of the sustainability rating on flows is of a similar magnitude to that

of the Morningstar star rating system.

These magnitudes are estimates of the net-impact of the ratings publication and associated publicity and roll out campaign by Morningstar and should be viewed in this context. The initial sorting measured here will be greater than the long-run effect we expect to occur after investors have sorted into various funds based on their sustainability. Although we are unable to examine whether the flow effects are permanent or transitory given the data's short time-series, we would not expect these effects to continue at the same magnitude without further ratings changes.[20] On the other hand, these are estimates of net flows which means they underestimate the number of investors who flowed into these funds based on sustainability ratings. On net, investors flowed into high sustainability funds, but likely some investors flowed out as well. Thus the estimates represent a lower bar for the proportion of investors who value these sustainability ratings in the market as a whole.

Increasing size is clearly an important aspect of overall fund health and as such the impact of the flows should be apparent in other fund attributes. One such attribute is the probability of a fund closing down. Table 6 examines the probability a fund shuts down based on its globe rating. We define a fund as closing if the final month a fund is present in our data occurs before the last month of the sample and Morningstar lists the fund as liquidated for each share class in our sample. Column 1 shows that 13 one globe funds shut down, while only 6 five globe funds did. The one globe rate of closure of 0.41% is more than double that of any of the other globe categories.

To further examine the pattern, we estimate linear probability models where the left hand side variable is equal to one if the fund is liquidated and the right hand side is the fund's globe rating. The coefficient on a globe rating can be interpreted as the difference in probability of a fund's liquidation in a given month relative to the omitted category.[21] Column 2 shows that a one globe fund is 0.24 percentage points more likely to close (t-statistic of 2.50) than a three globe fund, and that the other ranked funds do not seem to close at a higher or lower rate. Column 3 shows

[20] This is why papers examining the impact of a rating system already in equilibrium are forced to rely on estimates such as regression discontinuity to estimate their impact (e.g. Reuter and Zitzewitz 2010).

[21] In the Internet Appendix we conduct a similar exercise using Cox proportional hazard models and find analogous results.

that one globe funds are 0.25% more likely to close than all the other funds (with a t-statistic of 2.78). Columns 4 and 5 add category by year by month fixed effects and the additional controls respectively and finds similar results. Combining them all together in Column 6 the point estimate decreases to an insignificant 0.12%. The results are suggestive that being rated one globe leads to a higher probability of closing down, but given the rarity of the event we lack the statistical power to say for certain after including the full battery of controls and fixed effects.

4 Why do investors value sustainability?

We now explore three separate hypotheses to examine why investors place a positive value on sustainability. The first hypothesis is that institutional investors value sustainability due to constraints imposed by their institution. The second hypothesis is that investors (rightly or wrongly) view sustainability as a signal of higher future returns. The third hypothesis is that investors have a preference for sustainability for non-pecuniary reasons, such as altruism. These hypotheses are not mutually exclusive and it is likely that each has a hand in our results to some degree. In this section, we explore the extent to which each is important, but we are not be able to offer definitive answers as to the driving force underlying the demand for high sustainability rated mutual funds.

One remaining possibility that we cannot directly examine is that investors react to the globe rating as an arbitrary ranking without regard to the sustainability it is attempting to measure. This could occur either due to the salience of the image or because people believe that any rating Morningstar creates is a positive signal due to its reputation. While this is likely true for some investors, we believe it is unlikely to be the main driver of flows for several reasons. First, Morningstar spent significant resources attempting to make it clear to investors that the rating was measuring sustainability. Further, investors – especially institutional investors – presumably spent significant amounts of time and effort on their decisions, and they should therefore be likely to understand that the globe ratings were constructed to capture a fund's sustainability. Finally, the Google search analysis shows that roughly as many people search directly for the phrase "Morningstar sustainabil-

ity rating" as "Morningstar star rating." This suggests there are a large number of individuals who are sufficiently knowledgeable to search directly for the sustainability rating and who are not simply responding to the globe image at the top of the Morningstar webpage. Thus, it seems reasonable to assume that the flows we observe are driven significantly by an aspect related to sustainability.

4.1 Institutional Constraints

We begin by examining the hypothesis based on institutional constraints. For example, a University endowment may impose implicit or explicit constraints on its managers to avoid or invest in certain types of funds irrespective of maximizing returns.[22] If the results are being driven by such constraints, then the reaction by institutions should be different from that of non-institutional investors who do not share the same constraints. The ideal analysis would be specifically examining institutions that we knew were subject to such constraints. While we do not have this exact data, we can isolate the flows into and out of institutional share classes based on sustainability ratings.[23]

The use of institutional share class warrants caution when interpreting the results. If institutional investors are present in the market, we assume they are taking advantage of their size and investing primarily in institutional share classes. However, flows in these share classes may also be capturing the behavior of participants in retirement plans with access to institutional share classes (e.g. Sialm et al. 2015). If the institutional share classes only represent retirement plan participants, this would indicate that institutional investors were absent from the US mutual fund market and are not driving the effects we document. If institutions are the main driver of the flow patterns we observe, as long as institutions are present in the institutional share classes to some extent, the effects should be concentrated in the institutional share classes, but not in the non-institutional share classes.

Table 7 repeats the analysis examining the differential impact of institutional funds based on globe ratings. Analysis is run at the share-class level and standard errors are clustered by fund and

[22] Evidence supporting this hypothesis would be consistent with prior literature showing that institutional investors drive firms' environmental and social investments (e.g., Dyck et al. 2017) and the general importance of institutional investors more broadly (e.g. Gillan and Starks 2000; Gillan and Starks 2003).

[23] We use Morningstar's classification of institutional shares which typically require an investment of greater than $100,000.

month. Column 1 includes only non-institutional share classes. If institutions were solely responsible for the results, we would not expect there to be a large effect of the sustainability rating in this specification. Examining the regression, the spread between one and five globe funds was about 60 basis points per month, with a p-value of 0.008. The second column examines only institutional share classes. While statistically slightly weaker, the 72 basis point estimate of the spread between one and five globe funds is slightly larger, with a p-value of 0.06.

We next test whether the institutional share classes behaved differently from the non-institutional share classes. Specifically, in Column 3 we use data from both institutional and non-institutional share classes, include globe rating dummy variables, and also include interactions between the globe rating dummies and a dummy variable equal to one if the share-class is institutional. Including the globe dummy variables and the interaction terms means that the coefficient on the institutional interaction represents how different the flows into the institutional share classes with a given globe rating compare to the non-institutional share classes with the same globe rating. The insignificant interaction terms, with t-statistics of -0.23 on the one globe interaction and 0.36 on the five globe interaction, suggest similar responses to the globe ratings for institutional and non-institutional share classes. Finally, we examine whether non-institutional and institutional share classes within the same fund received different flows. We include a fixed effect for each fund by month combination and the institutional dummies interacted with globe ratings. The insignificance of the coefficients, with t-statistics of 0.09 on the one globe interaction and 0.22 on the five globe interaction, suggests that institutional and non-institutional share classes within the same fund received similar flows based on their sustainability rating.

While the institutional share classes represent a portion of the effect that we observe, the effects are still present and significant in the non-institutional share classes, suggesting that institutional behavior cannot fully account for the results. One interpretation of these results is that institutions behave in a manner similar to non-institutional investors. This could be because institutions have similar preferences to the non-institutional investors, or it could be that they face constraints forcing them to behave as if their preferences were similar. Another interpretation is that this analysis does

not reflect the preferences of institutional investors at all as the behavior represents individual investors trading in their retirement accounts. It could also be that institutions behave differently from non-institutional investors, but they are combined together in the institutional share class in such a way that we lack the power to detect different behavior of institutional investors. Under any of these interpretations, including the likely combination of all of them, the results suggest institutions are not the sole driver of the results that we document.

4.2 Rational Performance Expectations

The pattern in fund flows could also have been due to investors rationally viewing sustainability as a positive predictor of future fund performance. If investors had a rational belief that high sustainability funds would deliver high performance, we would hope that such out-performance would manifest itself in the data. While it is difficult to make a definitive conclusion examining 11 months of return data, we find evidence more consistent with an inverse relation or no relation between globe ratings and returns rather than the positive relation that would be necessary to account for the flow results under an explanation based on rational expectations.

4.2.1 Observed Performance

We examine returns relative to a variety of benchmarks in Table 8. Column 1 examines returns in excess of the risk free rate. Column 2 examines returns minus the value weighted return of funds in that Morningstar category (e.g. Pástor et al. 2015; Pástor et al. 2017). Column 3 examines returns in excess of a fund-specific benchmark based on Vanguard loadings. To do so, we follow Berk and Van Binsbergen (2015) to construct an orthogonal basis set of Vanguard index funds using data from 2014 to January 2017.[24] Fund specific betas on these projections are estimated prior to the globe rating publication and these betas are used to construct a fund's Vanguard benchmark return in the post-publication period. A similar methodology is used to construct a fund's 4-factor benchmark

[24]We utilize the same list of funds, though add the total bond market, short-term bond, intermediate-term bond and long-term bond. Our complete list (in order of inception date is thus): VFIAX, VBTLX, VEXAX, VSMAX, VEUSX, VPADX, VVIAX, VBIAX, VBIRX, VBILX, VBLLX, VEMAX, VIMAX, VSGAX and VSIAX.

using beta estimates on the factors of market, size, value and momentum. These measures of performance are regressed on globe ratings and are value weighted in Panel A and equal weighted in Panel B. For example, Column 1 Panel A shows the value weighted excess returns of one globe funds were 31 basis points higher than the excess returns of middle globe funds. Five globe funds received excess returns 25 basis points lower than middle globe funds received. The constant in this regression is 159 basis points which indicates that over this period one globe funds earned returns in excess of the risk-free rate of 1.9% per month, while five globe funds earned excess returns of 1.34% per month. Below the regression, we display the 56 basis point difference between the five globe and the one globe coefficient with the p-value that this difference is zero of 0.02.

Examining the 16 point estimates, each one globe estimate is positive and each five globe estimate is negative. Five of the eight five globe coefficients are significantly negative at the 10% level and two of the one globe coefficients are significantly positive at the 10% level. The point estimate of the spread between one and five globe funds is negative in each instance, ranging from 16 to 56 basis points per month with p-values on the difference ranging from 0.02 to 0.24.

In Panel C we form portfolios that are long firms that are rated five globes and short firms that are rated one globe. We regress this portfolio on just the market factor in columns 1 and 3 and on the market, size, value and momentum factors in columns 2 and 4. We report the alpha from these regressions in basis points. Value weighted, the four factor alpha returns -48 basis points (with a t-statistic of -2.14) and equal weighted the alpha is -18 basis points (with a t-statistic of -1.33). The portfolio sorts thus yield a similar estimate to the panel regressions in Panel A and B.

The short time series and volatility of returns makes it difficult to make definitive statements on the relation between returns and globe ratings in this natural experiment. The evidence does not support higher performance of five-globe funds relative to one globe funds which is what is necessary to explain the observed fund flows with a rational performance-based explanation, though it remains possible that such a belief was ex-ante justified and did not manifest itself in the relatively short 11 months of data. The evidence is consistent with both the hypothesis that one and five globe funds performed similarly as well as the hypothesis that one globe funds outperformed five

globe funds. The point estimate on five globes is lower then that for one globe in every specification suggesting the low sustainability funds outperformed the high sustainability, though the weak statistical significance in some specifications is also consistent with a lack of relation between globe ratings and performance.

4.2.2 Potential Explanations of Return Predictability

A variety of arguments have been made consistent with sustainability either positively predicting performance, negatively predicting performance or having no relation with performance. The focus of this paper is on how investors responded to the sustainability ratings, while what accounts for the return patterns is more closely related to the question of how funds responded to them. Although fully answering this question is beyond the scope of this paper, we discuss various explanations of fund performance as a function of sustainability ratings.

We group potential explanations of return predictability into three distinct categories. The first relates to the scale of funds with decreasing returns to scale. Berk and Green (2004) assume that funds have decreasing returns to scale which is empirically supported by the findings of Grinblatt and Titman (1989), Chen et al. (2004), Pástor et al. (2015) (though Reuter and Zitzewitz (2010) do not find such an effect). If an investor believed that the sustainability rating would cause flows to funds already at their optimal scale in a competitive equilibrium, the investor would expect high sustainability funds to underperform after their inflows and low sustainability funds to overperform after their out flows.[25]

The second class of theories relates to funds buying assets with high Sustainalytics ratings in order to achieve better fund ratings. Such an effect could be specific to funds competing on the Morningstar rating, or indicate general marketwide shifts in demand for sustainable investments. If funds were aware that ratings induce flows, they may actively trade to receive a higher sustainability

[25] A fund at its optimal scale is expected to earn zero abnormal returns, so inflows to high sustainability funds already at this scale would induce negative abnormal returns. In the context of Berk and Green (2004), if most investors cared only about returns they would undo aggregate flow effect induced by the sustainability ratings. The subset of investors who valued sustainability would shift into high sustainability funds and out of low sustainability funds while the profit maximizing investors would do the opposite. If such a pattern occurred we would see no aggregate flow response in our data.

rating, potentially at the expense of future returns. If many funds engaged in such a strategy, this could increase the price of assets with high Sustainalytics ratings. This price increase would yield a period of high returns for funds holding such assets, but would lead to subsequent underperformance as the price pressure reversed.[26]

The third class of explanations relates to the characteristics of the underlying assets, not fund behavior. For example, Hong and Kacperczyk (2009) argue that many investors are hesitant to hold "sin stocks," which leads these stocks to earn higher returns. Applying this intuition to our setting, if investors believed that there was a hesitance to hold low sustainability stocks, they might expect an inverse relation between returns and globe ratings. On the other hand, Edmans (2011) finds that employee satisfaction predicts positive returns, consistent with the idea that socially responsible screens can positively predict performance if the market is incorrectly pricing such signals.[27] If an investor believed that the market was not correctly pricing attributes correlated with high sustainability, they would expect higher returns for more sustainable funds.

To identify the relative importance of each to fund performance necessitates examining fund trading both prior to the ratings publication and also in reaction to the publication itself. We cannot perform this analysis since our data lacks holdings and is anonymized so it cannot be merged to publicly available holdings data. The internet appendix examines returns before and after the rating publications and finds aspects consistent and inconsistent with each of the three explanations as the noise inherent in fund level returns over a short period makes drawing definitive conclusions difficult. Holdings level data would allow a direct test of whether funds were systematically buying positions with high sustainability ratings and also whether this impacted the valuation of the underlying stocks. We leave it to future researchers with data that can be merged to holdings to further examine this issue.

[26] Similar to the return patterns observed for index inclusions (Harris and Gurel 1986; Shleifer 1986; Kaul et al. 2000), and dividend issuance (Hartzmark and Solomon 2013).

[27] Existing literature supports the possibility that sustainability could help a firm since it is well positioned to deliver warm-glow feelings to consumers (Becker 1974; Andreoni 1989; Cahan et al. 2015), or because corporate goodness could be used as a method for deterring harmful regulation or enforcement (Baron 2001; Hong and Liskovich 2015; Werner 2015) or broadly signal good governance (Deng et al. 2013; Dimson et al. 2015; Ferrell et al. 2016). Other papers have found evidence of sustainable investments being negative for a firm (e.g. Di Giuli and Kostovetsky 2014; Dharmapala and Khanna 2016; Fernando et al. 2017).

4.3 Naive Performance Expectations and Non-Pecuniary Motives

The remaining explanations are that investors either naively assumed that a high sustainability rating was predictive of high future fund returns or had a non-pecuniary preference for holding more sustainable mutual funds.[28] Unfortunately, the natural experiment from Morningstar does not allow for testable predictions that distinguish between naive beliefs about future returns versus preferences for sustainable funds because under either hypothesis the prediction is that more money would be allocated to high sustainability funds without observing higher subsequent performance. The difference between these two behaviors comes from the underlying motivation. Under the performance expectations hypothesis, the decision to invest more in high sustainability funds is driven by these performance expectations, while under the non-pecuniary motives hypothesis, the decision is driven by altruism, warm glow, or social motives. Thus differentiating between these two hypotheses requires a measure of expectations of future performance.

4.3.1 Experiment Overview

To obtain such a measure, we ran an experiment based on the Morningstar ratings to elicit the impact of the globe rating on expected future performance.[29] We gave participants information about three hypothetical mutual funds, derived from Morningstar's website. We picked three similar funds rated one globe, three globes and five globes, all with five star ratings on Morningstar's site. We randomized the sustainability ratings across these three funds in the experiment, and we gave participants Morningstar sustainability information along with fund information related to past performance and other fund characteristics. The display containing the globe ratings was taken directly from Morningstar's website to most closely simulate the information an investor would be seeing. However, it is possible that participants in the experiment did not understand the globe rating scale in the same way as a typical Morningstar investor. This would lead to a different motivation driving the responses of our experimental subjects than the Morningstar investors they

[28] For example, investors in funds with a socially responsible mandate derive utility from the social responsible aspect of the investment and are less sensitive to negative returns (Bollen, 2007; Renneboog et al., 2011).

[29] Additional details and survey materials are available in the online appendix.

are meant to represent. Thus, we replaced the text at the bottom of the Morningstar sustainability rating with a description of the globe ratings.[30]

Each participant was asked to (a) report how well she thought the fund would perform over the next year on a seven point Likert scale (b) report how risky she considered an investment in the fund to be on a seven point Likert scale and (c) allocate $1,000 between the fund and a savings account.[31] We examined MBA students at the University of Chicago Booth School of Business (269 students participated) to draw conclusions that would be more likely to be representative of market participants. In addition, we ran the experiment on 576 participants on Amazon Mechanical Turk (MTurk) to see how decisions were made in a likely less financially sophisticated subject pool.[32]

4.3.2 Performance Expectations

If flows to high sustainability funds are driven by increased performance expectations, then more globes will be positively correlated with these expectations. We first analyze whether people associate globe ratings with higher performance and find that they do. In Figure 6 Panel A, we graph the average performance rating for each of the three globe ratings, after removing an individual fixed effect. To the left, we examine the MBA students and see that moving from one globe to five globes is associated with an increase in expected performance of about 0.4, which is a statistically significant difference with a t-stat of 3.23 clustered by participant. To the right we see a similar, slightly stronger pattern for MTurk participants with a difference between extreme globe ratings of about 0.8 which is statistically significant with a t-statistic of 7.69. Thus the globes seem to have a slightly higher impact on MTurk participants than MBA students, but both groups strongly believe

[30] This text was taken from the Morningstar site and read, "This score provides a reliable, objective way to evaluate how investments are meeting environmental, social, and governance challenges." To avoid drawing additional attention to the globe ratings, this detail was designed to closely mimic text that appears in the globe display on the Morningstar site. Among the MTurk participants, half of participants saw the original text stating that the "Sustainability Mandate information is derived from the fund prospectus", and half saw the more informative message. We did not see meaningful differences in responses as a function of these messages and combine results for subsequent analysis.

[31] Participants responded to questions about performance for all three funds in one block, questions about risk for all three funds in one block, and questions about allocations for all three funds in one block. The order of these question blocks was counterbalanced across participants.

[32] Research examining this platform finds that participants recruited through MTurk tend to perform similarly on tasks (Casler et al., 2013) and better in attention checks (Hauser and Schwarz, 2016) than traditional participant pools recruited through labs, while representing a more diverse set of participants (Paolacci and Chandler, 2014).

that higher globe ratings lead to higher future performance.

One possibility is that these participants expected a fund with a higher globe rating to have higher performance because they thought five globe funds were riskier. We plot the expectations of risk in Figure 6 Panel B and find a strong inverse correlation between perceptions of risk and globe ratings, the opposite of what would be necessary to explain the performance expectations with risk. MBA students rated 5 globe funds as about 0.6 points less risky than one globe funds, with a t-statistic on the difference of -4.67. MTurk participants exhibit similar, slightly stronger behavior with a difference of roughly 0.8, with a t-statistic of -6.86. Thus it is unlikely that the positive correlation between globe ratings and performance is due to compensation for perceived risk. Participants reported that higher globe ratings would result in higher performance at lower risk.

Although the finding that investors believe both that performance will be superior and that risk will be lower for funds rated high in sustainability may appear surprising, it is consistent with existing research in psychology. The affect heuristic (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2004, 2005, 2007) and research examining the role of affect in decision making (Loewenstein et al. 2001; Nisbett and Wilson 1977; Klauer and Stern 1992) have been used to explain a range of contexts where risks and benefits are positively correlated, but people believe them to be negatively correlated (Fischoff and Lichtenstein, 1978; Slovic et al., 1991; McDaniels et al., 1997). This research posits that people rely on affect and emotion - rather than reasoned analysis - to assess attributes of a given stimulus and make subsequent decisions.[33] To the extent that the high sustainability rating causes positive affect towards a mutual fund, the affect heuristic would suggest that people are likely to judge it to be both higher in returns and lower in risk.

4.3.3 Non-Pecuniary Motives

While higher expected performance alone could account for the patterns we observe in Morningstar data, this does not rule out the possibility that non-pecuniary motives also play a role. In other

[33]For example, Finucane et al. (2000) experimentally manipulate participants' affective evaluations of items such as nuclear power and find that perceptions of both risks and benefits shift to be congruent with the overall evaluation.

words, people may be investing in highly sustainable funds only because they believe they will outperform, or also because they value sustainability and are willing to pay for it. This preference could derive from a number of non-economic motivations, and would be consistent with evidence and theorizing that people are concerned with increasing social welfare (Charness and Rabin 2002; Fehr and Schmidt 1999). For example, investors may experience altruism or warm glow (Andreoni 1989, 1990), in which case they would want to invest in sustainability because they derive value from the fact that others benefit, or feel good because they are responsible for benefiting others. Alternatively, it could stem from social motives and pressures such as the desire to impress others or to avoid contempt or social backlash (Becker 1974; DellaVigna et al. 2012; Olson 2009).

In the context of our experiment, one potential measure of non-pecuniary motives is the extent to which an investor allocates funds towards five globe funds or away from one globe funds that is not explained by their expectation of future performance or risk. If participants cared about the globe ratings solely as indicators of fund performance, we would expect the globes to impact expectations of future performance and risk. Under such an explanation, after controlling for these expectations, the globe ratings would have no further explanatory power. In Table 9, we examine how dollars allocated to portfolios vary with expectations of risk, performance and globe ratings. Regressions include a participant fixed effect and a fund fixed effect. If there is a significant difference between the one and five globe dummy variables, this indicates that a participant is more or less likely to invest in the given globe level than can be accounted for by performance and risk expectations alone. Thus, a positive difference between the five globe and one globe dummy variables in this analysis is consistent with altruism. We do caution that interpreting the results in such a manner requires the assumption that the portfolio weights for an investor who only cares about performance and risk increase linearly in the measures based on a Likert scale. While not definitive, we believe that it offers insight into a question with little information currently available.

The first column of Table 9 shows that dollars allocated to a fund are strongly positively correlated with expected performance and strongly negatively correlated with expected risk. Column 2 shows that without controlling for either risk or performance, investors allocate more money to

five globe funds and less to one. MBA students allocate $108 more to five globe funds than to one globe funds (with a p-value of roughly 0 on the difference) and MTurk participants allocate about $130 more (again with a p-value of roughly 0).

Column 3 includes risk, performance and the globe ratings to identify whether this difference in allocations is explained by performance expectations alone or whether non-pecuniary motives also play a role. After including the controls for risk and performance, the difference between funds allocated by MBA students towards one versus five globe funds drops, but remains meaningful at $48 (with a 95% confidence interval of $1 to $95), with a p-value of 0.04. For MTurk participants this difference is $71 (with a 95% confidence interval of $42 to $100), with a p-value of roughly 0. The point estimates suggest that slightly less than half of the difference in money allocated to one and five globe funds can be attributed to non-pecuniary motives for the MBA students, while non-pecuniary motives account for slightly more than half of the difference for MTurk participants.

If a portion of the difference in allocation is driven by non-pecuniary motives related to sustainability, this effect of globe ratings should be concentrated among participants who considered sustainability when making their decisions. After making their choices, we asked participants the extent to which they considered ESG factors when making their investment decisions. Participants who said they did not consider ESG factors have no reason to exhibit non-pecuniary motives, so we would expect the globe dummy variables capturing such motives to lose their explanatory power for these investors. This is what we find when we restrict the sample to these investors in Column 4. MBA students in this group exhibit only a $5 difference in allocation between 1 and 5 globe funds while MTurk subjects exhibit a marginally significant $41 difference. Column 5 shows strong evidence consistent with non-pecuniary motives when examining participants who considered ESG factors. MBA students allocated a significant $79 more dollars towards five globe funds (with a 95% confidence interval ranging from $16 to $142) and MTurk participants allocated a significant $86 towards five globe funds (with a 95% confidence interval from $51 to $122). Thus we see evidence that dollar allocations are driven by expected performance and risk, but also by altruism (or other

non-pecuniary motives) above and beyond these factors.[34]

5 Conclusion

We present causal evidence that investors collectively value sustainability and rule out the possibility that investors are indifferent to this information or that they penalize a fund for maintaining a portfolio of sustainable investments. We find that funds with the highest globe ratings receive a more than $24 billion increase in fund flows while those with the lowest globe ratings face a more than $12 billion reduction in fund flows. This suggests that a large portion of the market views sustainability as a positive company attribute.

Although investors are presented with detailed information about the percentile rank of sustainability within Morningstar categories, they largely ignore this information and instead respond to the simpler and more salient globe ratings, consistent with the psychological literature on categorization. They further respond mainly to the extreme ranked categories, largely ignoring the others, consistent with literature on the salience of extreme ranks. The results suggest that how categories are constructed, especially extreme categories, can have a significant impact on how decisions are made in a financial setting and impact marketwide variables such as fund flows.

Our natural experiment in which a large portion of the market experiences a quasi-exogenous shock that does not impact fundamentals is rare in financial markets. This allows us to cleanly identify the causal effect of the sustainability ratings on mutual fund flows. We propose and find support for several explanations of the response to the publication of the ratings. The flow pattern is present among institutional share classes, especially for high sustainability funds, consistent with social constraints placed upon institutions being partially responsible for the effect. However, the pattern persists among non-institutional investors as well. We do not find evidence supporting a

[34] The results also suggest that the experiment is not capturing a pure attention effect induced by the ratings. Under such an explanation, any salient ranking we presented would induce the observed empirical pattern in allocations due to the picture itself, but not the underlying context of the rating. If this were the case, the amount that an investor considers environmental factors would be unlikely to influence investment decisions. These differential responses suggest that the patterns we observe in the experimental setting were largely due to considerations related to sustainability, and not simply an attention effect unrelated to sustainability.

rational belief that more sustainable funds perform better, instead the evidence is more consistent with the opposite. In spite of this, our experimental evidence suggests that investors have a strong belief that better globe ratings positively predict future returns. We also find suggestive evidence of non-pecuniary motives, consistent with altruism or warm glow.

Taken together, our experimental findings support the importance of affect in investment decisions. Specifically, the finding that participants expect that funds rated high in sustainability will both perform better and have lower risk is consistent with prior research on the affect heuristic (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2004, 2005, 2007). The patterns we observe may speak to a general phenomenon, whereby attributes that are not related to performance can alter an investor's feelings about an investment. A positive shift in affect increases investor expectations of future returns and lowers perceived risk while a negative shift results in the opposite. This is consistent with findings on halo effects, in which an impression formed in one area influences overall evaluations (Nisbett and Wilson, 1977; Klauer and Stern, 1992). In our setting, an investor who values sustainability for non-pecuniary reasons, such as moral ideals, may have the positive affect impact their perception of performance.

An additional question that emerges is how investors in our dataset and participants in our experiment are interpreting the sustainability ratings. For example, although we found that people tend to associate sustainability with the environment, people may be considering the Morningstar sustainability rating to be specific to environmental factors, or more broadly indicative of a fund's corporate social responsibility. It is also possible that due to Morningstar's reputation, investors trust that Morningstar has measured sustainability in the most sensible way and respond to it without giving additional thought to what they are measuring. We have not attempted to define sustainability throughout this paper, instead simply using Morningstar's definition of the concept. What investors actually are responding to when they view the sustainability ratings, or any number of other socially responsible investment objectives, is an interesting and open question for further study.

References

Abadie, Alberto, and Guido W Imbens, 2006, Large sample properties of matching estimators for average treatment effects, *econometrica* 74, 235–267.

Abadie, Alberto, and Guido W Imbens, 2011, Bias-corrected matching estimators for average treatment effects, *Journal of Business & Economic Statistics* 29, 1–11.

Alhakami, Ali Siddiq, and Paul Slovic, 1994, A psychological study of the inverse relationship between perceived risk and perceived benefit, *Risk Analysis* 14, 1085–1096.

Andreoni, James, 1989, Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of political Economy* 97, 1447–1458.

Andreoni, James, 1990, Impure altruism and donations to public goods: A theory of warm-glow giving, *The economic journal* 100, 464–477.

Barber, Brad M, Adair Morse, and Ayako Yasuda, 2017, Impact investing.

Baron, David P, 2001, Private politics, corporate social responsibility, and integrated strategy, *Journal of Economics & Management Strategy* 10, 7–45.

Becker, Gary S, 1974, A theory of social interactions, *Journal of political economy* 82, 1063–1093.

Bénabou, Roland, and Jean Tirole, 2010, Individual and corporate social responsibility, *Economica* 77, 1–19.

Benson, Karen L, and Jacquelyn E Humphrey, 2008, Socially responsible investment funds: Investor reaction to current and past returns, *Journal of Banking & Finance* 32, 1850–1859.

Berk, Jonathan B, and Richard C Green, 2004, Mutual fund flows and performance in rational markets, *Journal of political economy* 112, 1269–1295.

Berk, Jonathan B, and Jules H Van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.

Bialkowski, Jedrzej, and Laura T Starks, 2016, Sri funds: Investor demand, exogenous shocks and esg profiles, Technical report.

Birru, Justin, 2017, Day of the week and the cross-section of returns .

Bollen, Nicolas PB, 2007, Mutual fund attributes and investor behavior, *Journal of Financial and Quantitative Analysis* 42, 683–708.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2012, Salience theory of choice under risk, *The Quarterly journal of economics* qjs018.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013a, Salience and asset prices, *The American Economic Review* 103, 623–628.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013b, Salience and consumer choice, *Journal of Political Economy* 121, 803–843.

Cahan, Steven F, Chen Chen, Li Chen, and Nhut H Nguyen, 2015, Corporate social responsibility and media coverage, *Journal of Banking & Finance* 59, 409–422.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik, 2018, Regression discontinuity designs using covariates, *Review of Economics and Statistics* .

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, 2014, Robust nonparametric confidence intervals for regression-discontinuity designs, *Econometrica* 82, 2295–2326.

Casler, Krista, Lydia Bickel, and Elizabeth Hackett, 2013, Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing, *Computers in Human Behavior* 29, 2156–2160.

Charness, Gary, and Matthew Rabin, 2002, Understanding social preferences with simple tests, *The Quarterly Journal of Economics* 117, 817–869.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D Kubik, 2004, Does fund size erode mutual fund performance? the role of liquidity and organization, *The American Economic Review* 94, 1276–1302.

Cheng, Ing-Haw, Harrison Hong, and Kelly Shue, 2013, Do managers do good with other people's money?, Technical report, National Bureau of Economic Research.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Chowdhry, Bhagwan, Shaun Davies, and Brian Waters, 2017, Investing for impact .

Christensen, Hans B, Eric Floyd, Lisa Yao Liu, and Mark Maffett, 2017, The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records, *Journal of Accounting and Economics* 64, 284–304.

Del Guercio, Diane, and Paula A Tkac, 2008, Star power: The effect of monrningstar ratings on mutual fund flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, 2012, Testing for altruism and social pressure in charitable giving, *The quarterly journal of economics* 127, 1–56.

Deng, Xin, Jun-koo Kang, and Buen Sin Low, 2013, Corporate social responsibility and stakeholder value maximization: Evidence from mergers, *Journal of Financial Economics* 110, 87–109.

Dharmapala, Dhammika, and Vikramaditya S Khanna, 2016, The impact of mandated corporate social responsibility: Evidence from india's companies act of 2013 .

Di Giuli, Alberta, and Leonard Kostovetsky, 2014, Are red or blue companies more likely to go green? politics and corporate social responsibility, *Journal of Financial Economics* 111, 158–180.

Diecidue, Enrico, and Peter P Wakker, 2001, On the intuition of rank-dependent utility, *Journal of Risk and Uncertainty* 23, 281–298.

Dimson, Elroy, Oğuzhan Karakaş, and Xi Li, 2015, Active ownership, *The Review of Financial Studies* 28, 3225–3268.

DiNardo, John, and David S Lee, 2011, Program evaluation and research designs, *Handbook of labor economics* 4, 463–536.

Dyck, IJ, Karl Lins, Lukas Roth, and Hannes Wagner, 2017, Do institutional investors drive corporate social responsibility? international evidence, Technical report.

Edmans, Alex, 2011, Does the stock market fully value intangibles? employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621–640.

Feenberg, Daniel, Ina Ganguli, Patrick Gaule, and Jonathan Gruber, 2017, It's good to be first: Order bias in reading and citing nber working papers, *Review of Economics and Statistics* 99, 32–39.

Fehr, Ernst, and Klaus M Schmidt, 1999, A theory of fairness, competition, and cooperation, *The quarterly journal of economics* 114, 817–868.

Fernando, Chitru S, Mark P Sharfman, and Vahap B Uysal, 2017, Corporate environmental policy and shareholder value: Following the smart money, *Journal of Financial and Quantitative Analysis* 52, 2023–2051.

Ferrell, Allen, Hao Liang, and Luc Renneboog, 2016, Socially responsible firms, *Journal of Financial Economics* 122, 585–606.

Finucane, Melissa L, Ali Alhakami, Paul Slovic, and Stephen M Johnson, 2000, The affect heuristic in judgments of risks and benefits, *Journal of behavioral decision making* 13, 1.

Fischoff, Baruch, and Sarah Lichtenstein, 1978, Don't attribute this to reverend bayes., *Psychological Bulletin* 85, 239.

Geczy, Christopher, Robert F Stambaugh, and David Levin, 2005, Investing in socially responsible mutual funds, Technical report.

Gillan, Stuart, and Laura T Starks, 2003, Corporate governance, corporate ownership, and the role of institutional investors: A global perspective .

Gillan, Stuart L, and Laura T Starks, 2000, Corporate governance proposals and shareholder activism: The role of institutional investors, *Journal of financial Economics* 57, 275–305.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of business* 393–416.

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures, *the Journal of Finance* 41, 815–829.

Harris, Lawrence E, Samuel M Hartzmark, and David H Solomon, 2015, Juicing the dividend yield: Mutual funds and the demand for dividends, *Journal of Financial Economics* 116, 433–451.

Hart, Oliver, and Luigi Zingales, 2017, Companies should maximize shareholder welfare not market value .

Hartzmark, Samuel M., 2015, The worst, the best, ignoring all the rest: The rank effect and trading behavior, *Review of Financial Studies* 28, 1024–1059.

Hartzmark, Samuel M, and David H Solomon, 2013, The dividend month premium, *Journal of Financial Economics* 109, 640–660.

Hartzmark, Samuel M, and David H Solomon, 2019, The dividend disconnect, *Journal of Finance* Forthcoming.

Hauser, David J, and Norbert Schwarz, 2016, Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants, *Behavior research methods* 48, 400–407.

Heal, Geoffrey, 2005, Corporate social responsibility: An economic and financial framework, *The Geneva papers on risk and insurance Issues and practice* 30, 387–409.

Hirshleifer, David, and Tyler Shumway, 2003, Good day sunshine: Stock returns and the weather, *The Journal of Finance* 58, 1009–1032.

Hong, Harrison, and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hong, Harrison, and Inessa Liskovich, 2015, Crime, punishment and the halo effect of corporate social responsibility, Technical report, National Bureau of Economic Research.

Huberman, Gur, 2001, Familiarity breeds investment, *The Review of Financial Studies* 14, 659–680.

Imbens, Guido W, and Thomas Lemieux, 2008, Regression discontinuity designs: A guide to practice, *Journal of econometrics* 142, 615–635.

Kaul, Aditya, Vikas Mehrotra, and Randall Morck, 2000, Demand curves for stocks do slope down: New evidence from an index weights adjustment, *The Journal of Finance* 55, 893–912.

Kitzmueller, Markus, and Jay Shimshack, 2012, Economic perspectives on corporate social responsibility, *Journal of Economic Literature* 50, 51–84.

Klauer, Karl Christoph, and Elsbeth Stern, 1992, How attitudes guide memory-based judgments: A two-process model, *Journal of Experimental Social Psychology* 28, 186–206.

Lee, David S, and David Card, 2008, Regression discontinuity inference with specification error, *Journal of Econometrics* 142, 655–674.

Loewenstein, George F, Elke U Weber, Christopher K Hsee, and Ned Welch, 2001, Risk as feelings., *Psychological Bulletin* 127, 267.

Margolis, Joshua D, Hillary Anger Elfenbein, and James P Walsh, 2009, Does it pay to be good... and does it matter? a meta-analysis of the relationship between corporate social and financial performance .

McDaniels, Timothy L, Lawrence J Axelrod, Nigel S Cavanagh, and Paul Slovic, 1997, Perception of ecological risk to water environments, *Risk analysis* 17, 341–352.

Nisbett, Richard E, and Timothy D Wilson, 1977, The halo effect: Evidence for unconscious alteration of judgments., *Journal of Personality and Social Psychology* 35, 250.

Olson, Mancur, 2009, *The logic of collective action*, volume 124 (Harvard University Press).

Paolacci, Gabriele, and Jesse Chandler, 2014, Inside the turk: Understanding mechanical turk as a participant pool, *Current Directions in Psychological Science* 23, 184–188.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2017, Do funds make more when they trade more?, *The Journal of Finance* 72, 1483–1528.

Quiggin, John, 1982, A theory of anticipated utility, *Journal of Economic Behavior & Organization* 3, 323–343.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2008, Socially responsible investments: Institutional aspects, performance, and investor behavior, *Journal of Banking & Finance* 32, 1723–1742.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2011, Is ethical money financially smart? nonfinancial attributes and money flows of socially responsible investment funds, *Journal of Financial Intermediation* 20, 562–588.

Reuter, Jonathan, and Eric Zitzewitz, 2010, How much does size erode mutual fund performance? a regression discontinuity approach, Technical report, National Bureau of Economic Research.

Riedl, Arno, and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *The Journal of Finance* .

Schmeidler, David, 1989, Subjective probability and expected utility without additivity, *Econometrica: Journal of the Econometric Society* 571–587.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *The Journal of Finance* 41, 579–590.

Sialm, Clemens, Laura T Starks, and Hanjiang Zhang, 2015, Defined contribution pension plans: Sticky or discerning money?, *The Journal of Finance* 70, 805–838.

Simonson, Itamar, and Amos Tversky, 1992, Choice in context: Tradeoff contrast and extremeness aversion, *Journal of marketing research* 29, 281.

Skowronski, John J, and Donal E Carlston, 1989, Negativity and extremity biases in impression formation: A review of explanations., *Psychological bulletin* 105, 131.

Slovic, Paul, Melissa L Finucane, Ellen Peters, and Donald G MacGregor, 2004, Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality, *Risk Analysis* 24, 311–322.

Slovic, Paul, Melissa L Finucane, Ellen Peters, and Donald G MacGregor, 2007, The affect heuristic, *European journal of operational research* 177, 1333–1352.

Slovic, Paul, Nancy Kraus, Henner Lappe, and Marilyn Major, 1991, Risk perception of prescription drugs: report on a survey in canada, *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique* 82, S15–S20.

Slovic, Paul, Ellen Peters, Melissa L Finucane, and Donald G MacGregor, 2005, Affect, risk, and decision making., *Health Psychology* 24, S35.

Sussman, Abigail B, 2017, Valence in context: Asymmetric reactions to realized gains and losses., *Journal of Experimental Psychology: General* 146, 376.

Sussman, Abigail B, and Eldar Shafir, 2012, On assets and debt in the psychology of perceived wealth, *Psychological Science* 23, 101–108.

Thistlethwaite, Donald L, and Donald T Campbell, 1960, Regression-discontinuity analysis: An alternative to the ex post facto experiment., *Journal of Educational psychology* 51, 309.

Tversky, Amos, and Daniel Kahneman, 1986, Rational choice and the framing of decisions, *Journal of business* S251–S278.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and uncertainty* 5, 297–323.

Tversky, Amos, and Itamar Simonson, 1993, Context-dependent preferences, *Management science* 39, 1179–1189.

Weber, Elke, and Britt Kirsner, 1997, Reasons for rank-dependent utility evaluation, *Journal of Risk and Uncertainty* 14, 41–61.

Werner, Timothy, 2015, Gaining access by doing good: The effect of sociopolitical reputation on firm participation in public policy making, *Management Science* 61, 1989–2011.

Figure 2
Example of Globe Rating on Morningstar Website

This picture is an example from Morningstar's website of how sustainability information is displayed on a fund's webpage.

Figure 3
Google Search for Sustainability and Star Rating

This graph shows monthly google search volume based on sustainability rating and Morningstar star rating. The maroon line is based on searches for "Morningstar globe rating" while the navy line represents searches for "Morningstar star rating." Data is from a search for both terms jointly so the magnitude of the two lines can be directly compared. The monthly measure is the average of the weekly measure where months are defined based on month ending period. Data cover January 2015 through January 2017.

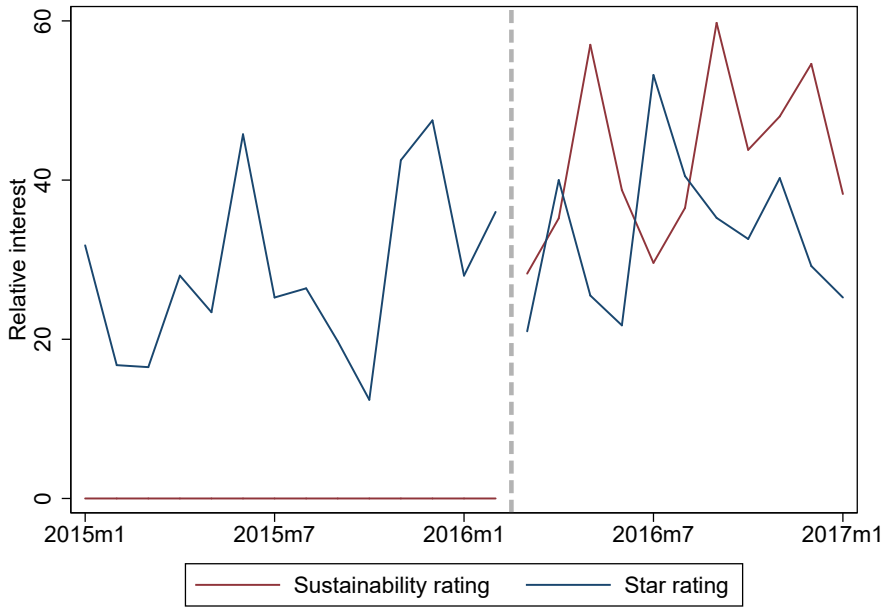
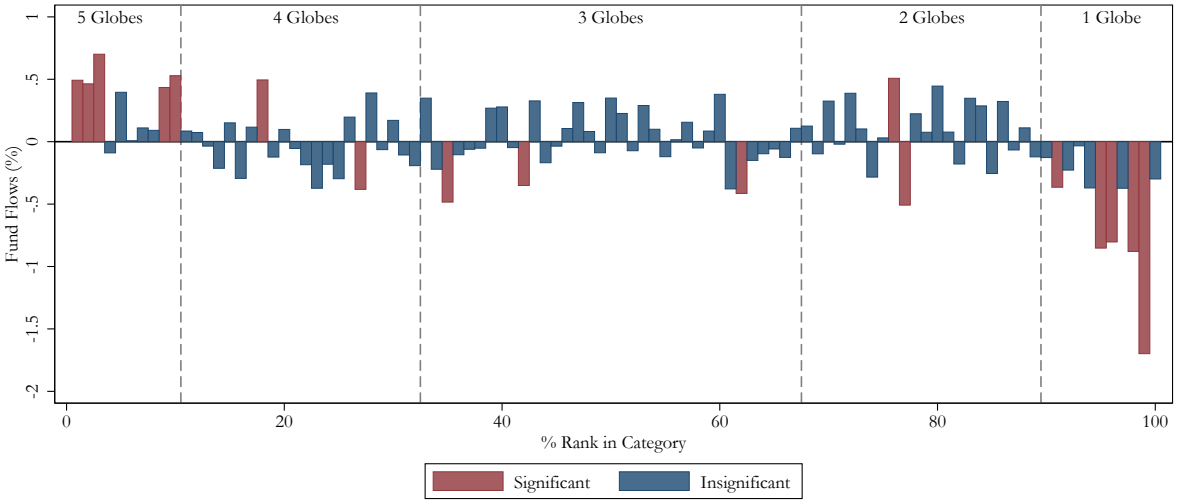


Figure 4
Flows by Percentile Rank of Sustainability

This graph shows average percentage flows for each sustainability percentile rank after controlling for year by month fixed effects. Panel A shows the averages of this variable. Panel B examines the normalized flow measure, normalized to be mean 0. Significant indicates the average flow is significant at the 90% level.

Panel A: Flows by Percentile Rank



Panel B: Normalized Flows by Percentile Rank

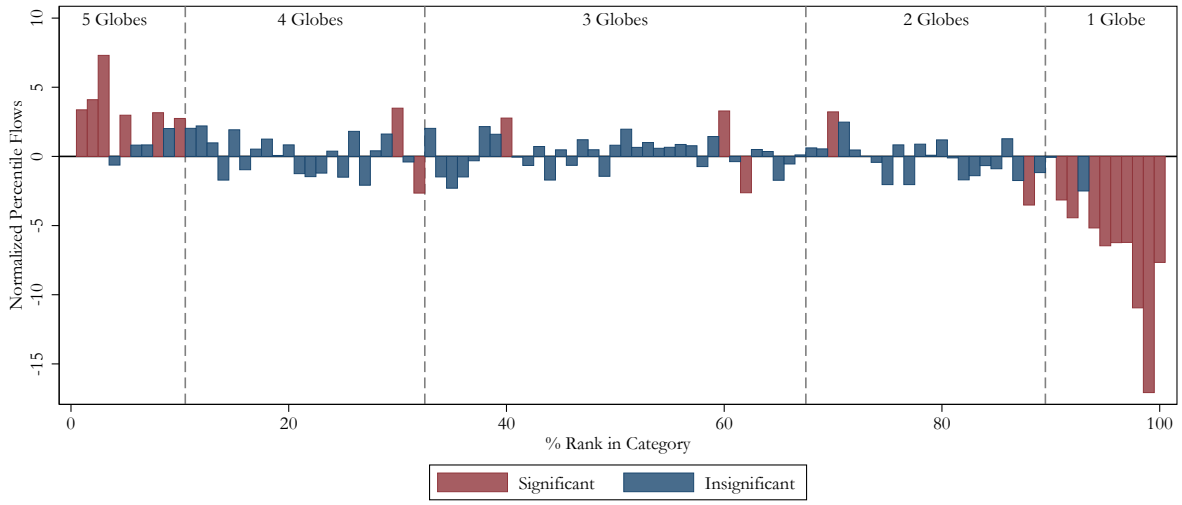
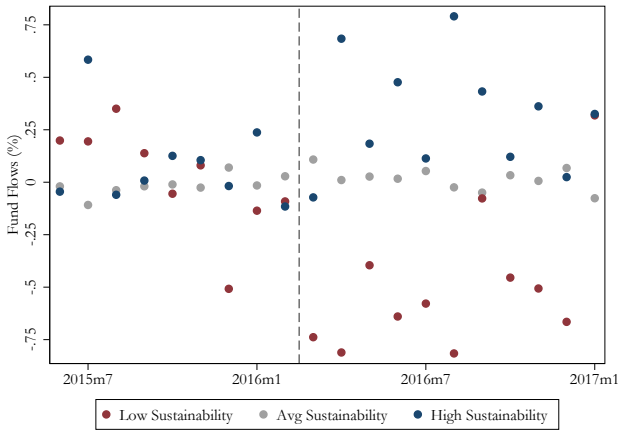


Figure 5
Flows by Month

This graph shows average percentage flows by month controlling for a year by month fixed effect. Panel A shows the average of the variable for each month and Panel B shows a local linear plot with 90% confidence intervals.

Panel A: Average Flows



Panel B: Local Linear Figure

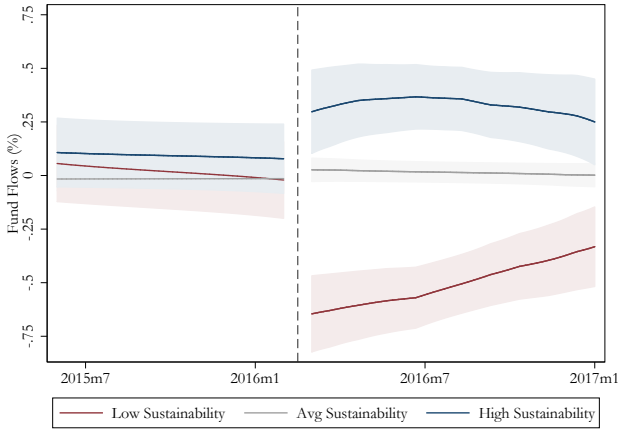
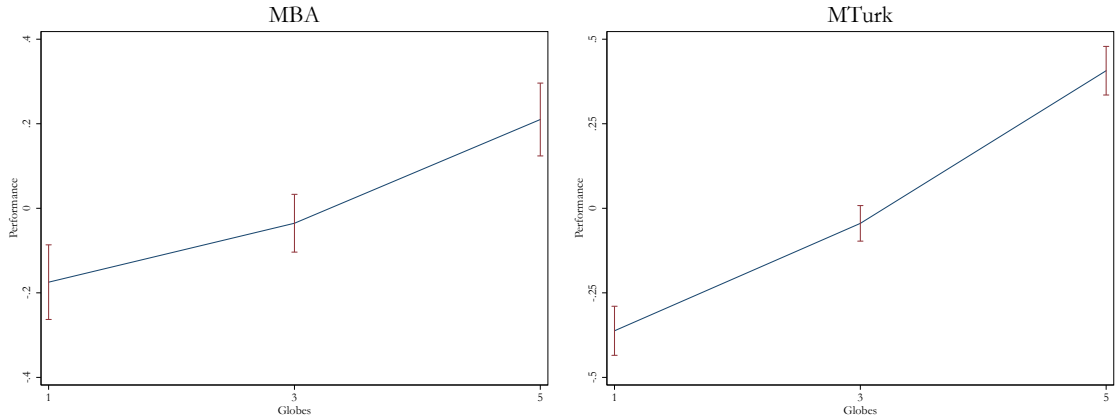


Figure 6

Experimental Expectations of Future Performance and Risk by Sustainability Rating

This graph shows the average performance rating in Panel A and risk rating in Panel B after taking out an individual fixed effect by globe rating. The left graphs are MBA students while the right graphs examine MTurk subjects. Maroon bars indicate the 90% confidence interval.

Panel A: Expectation of Performance



Panel B: Expectation of Risk

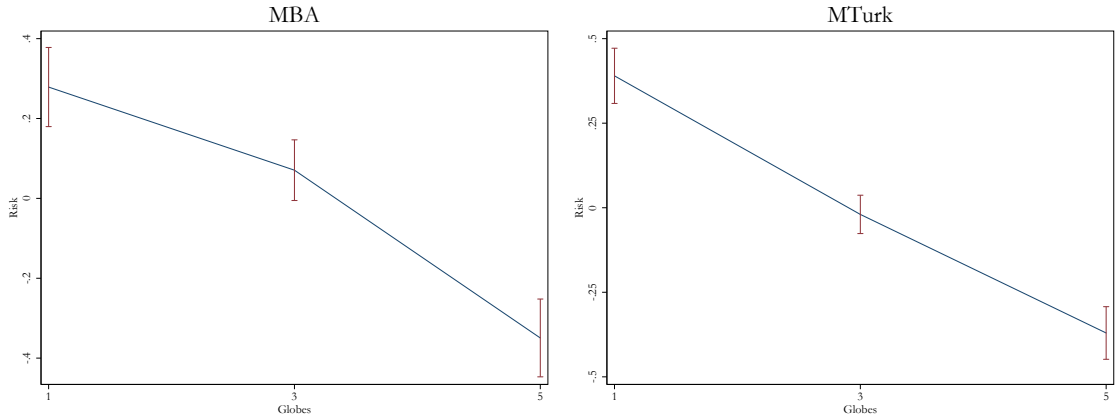


Table 1
Summary Statistics

This table shows summary statistics of the data. Panel A examines all funds post-publication, from March 2016 through January 2017. Statistics are at the fund level. Panel B examines the data by Globe for the 9 months prior to publication where Globes are defined as the rating the fund receives in March 2016. Panel C examines the data by globe after publication. Panel D shows the transition matrix from month to month for each globe rating after publication.

Panel A: Post-Publication Summary Stats

	Mean	SD	p10	p25	p50	p75	p90
Flow	-0.41	4.68	-3.43	-1.60	-0.60	0.37	2.46
Visits	209.45	474.48	1	14	44	159	521
Size	2184.33	8617.62	19.98	76.36	350.29	1370.35	4105.12
Rating	3.01	1.00	2	2	3	4	4

Panel B: Pre-Publication Summary Stats By Globe

	Obs	Size	Flows	Normalized Flows	Visits	Return	Age	Rating
All	28713	2112.38	0.10	50.33	229.80	-1.27	183.94	3.03
1 Globe	2982	1392.05	0.12	48.12	235.64	-1.40	178.57	2.92
2 Globes	6215	2370.89	0.28	52.52	223.62	-1.23	184.36	3.10
3 Globes	9891	2353.41	0.01	50.10	229.37	-1.29	183.85	3.10
4 Globes	6422	1937.82	0.03	49.54	218.84	-1.24	194.52	3.02
5 Globes	3174	1885.88	0.19	50.43	259.84	-1.18	167.08	2.76

Panel C: Post-Publication Summary Stats By Globe

	Obs	Size	Flows	Normalized Flows	Visits	Return	Age	Rating
All	34105	2184.32	-0.41	50.33	209.45	1.62	195.35	3.01
1 Globe	3170	1039.96	-0.90	44.69	164.34	1.72	180.58	2.74
2 Globes	7207	2438.41	-0.32	50.36	205.35	1.71	195.89	3.05
3 Globes	12183	2298.06	-0.41	50.70	201.52	1.62	198.11	3.10
4 Globes	7816	2197.10	-0.45	50.69	207.97	1.56	201.51	3.00
5 Globes	3730	2267.72	-0.10	53.13	284.43	1.53	184.94	2.83

Panel D: Transition Probability

		Next Month Rating				
		1 Globe	2 Globes	3 Globes	4 Globes	5 Globes
Current Month Rating	1 Globe	2297	539	37	8	0
		(79.73%)	(18.71%)	(1.28%)	(0.28%)	(0.00%)
	2 Globes	436	4869	1170	29	6
		(6.70%)	(74.79%)	(17.97%)	(0.45%)	(0.09%)
	3 Globes	64	983	8753	1185	28
		(0.58%)	(8.93%)	(79.48%)	(10.76%)	(0.25%)
	4 Globes	18	93	1032	5415	512
		(0.25%)	(1.32%)	(14.60%)	(76.59%)	(7.24%)
	5 Globes	4	14	61	467	2837
		(0.12%)	(0.41%)	(1.80%)	(13.80%)	(83.86%)

Table 2
Survey on the Meaning of Sustainability

This table shows summary statistics from a survey on the perceived meaning of sustainability. 482 participants on Amazon Mechanical Turk responded to the question "Recently, many companies have been trying to becoming more sustainable. Which of the following elements of a company's business practices do you think "sustainability" refers to?" Participants were given a list of categories with examples based on KLD definitions and were asked to select all categories that applied.

	% Selecting
Environment (e.g., pollution prevention, recycling)	79%
Products (e.g., product quality and safety, provision of products for the economically disadvantaged)	48%
Human Rights (e.g., labor rights in outsourcing, no operations in Myanmar)	34%
Community (e.g., generous giving, support for housing)	32%
Diversity (e.g., promotion of women and minorities, outstanding family benefits)	26%
Employee Relations (e.g., strong union relations, cash profit sharing)	23%
Corporate Governance (e.g., limited compensation to executives, lack of tax disputes)	22%
I do not Know	2%
Other	1%

Table 3
Fund Flows in Response to Sustainability Rating

This table shows how mutual fund flows vary with measures of sustainability. The dependent variable is fund flows which are regressed on three proxies of sustainability, namely the raw sustainability score, the percentile rank within category and dummy variables for globe rankings. Columns 4 through 6 of Panel A and the even numbered of Columns of Panel B include additional controls of return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating in the prior month and the log of fund age. Panel A does not weight regressions, while Panel B weights by log of TNA the month prior in Columns three through six. All Columns include year by Morningstar category by month fixed effects. Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at the fund level. Standard errors are clustered by month and fund, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Baseline Flow Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Sustainability Score	0.0744					0.0612
	(0.94)					(0.65)
Category Percent Rank	0.000983					0.00398
	(0.23)					(0.89)
1 Globe		-0.441**	-0.457***	-0.352**	-0.402***	-0.408**
		(-2.80)	(-3.21)	(-2.81)	(-3.38)	(-2.55)
2 Globes		0.0964		0.134		
		(0.92)		(1.33)		
4 Globes		-0.0353		0.0440		
		(-0.40)		(0.51)		
5 Globes		0.297	0.281*	0.379**	0.331**	0.319*
		(1.81)	(1.87)	(2.39)	(2.35)	(1.85)
Diff: 5 Globe-1 Globe		0.737	0.738	0.731	0.733	0.727
P-value: 5 Globe=1 Globe		0.00384	0.00382	0.00377	0.00376	0.0296
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	Yes	Yes	Yes
R^2	0.0503	0.0511	0.0510	0.0909	0.0907	0.0909
Observations	34046	34046	34046	32421	32421	32421

Panel B: Normalized Flow and Size Weighted Regressions

	Normalized Flows		Size Weighted Flows		Size Weighted Normalized Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
1 Globe	-5.743***	-4.427***	-0.444***	-0.353**	-5.802***	-4.389***
	(-5.53)	(-4.62)	(-3.20)	(-3.07)	(-5.47)	(-4.59)
5 Globes	2.474**	3.253***	0.302*	0.358**	2.686**	3.465***
	(2.25)	(3.19)	(2.07)	(2.64)	(2.35)	(3.29)
Diff: 5 Globe-1 Globe	8.217	7.680	0.746	0.711	8.487	7.855
P-value: 5 Globe=1 Globe	0.000402	0.000397	0.00299	0.00330	0.000380	0.000347
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes	No	Yes
R^2	0.0711	0.157	0.0498	0.0883	0.0725	0.160
Observations	34046	32421	34046	32421	34046	32421

Table 4

Regression Discontinuity Tests of Fund Flows Around Sustainability Rating Breakpoints

This table conducts regression discontinuity tests of mutual fund flows around Globe breakpoints. The first row shows the conventional RD estimate while the second corrects for the bias described in Calonico et al. (2014). Panel A examines flows at the one globe breakpoint while Panel B examines flows at the five globe breakpoint. Columns include monthly fixed effects in estimation. Bandwidths are selected using the procedure described in Calonico et al. (2014) and Calonico et al. (2018) and are reported below. Odd columns use the same bandwidth on either side of the breakpoint while even columns estimate separate bandwidths. Columns 1 through 4 calculate z-stats using a heteroskedasticity robust nearest neighbor estimator. Other columns calculate z-stats using a cluster robust plug-in residuals variance estimator as described in Calonico et al. (2018) with the clustering variable indicated below the estimates. Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at fund level. z-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Flows at One Globe Breakpoint

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	-0.463**	-0.319**	-0.466***	-0.326**	-0.412*	-0.305	-0.556***	-0.385**
	(-2.46)	(-2.02)	(-2.60)	(-2.15)	(-1.66)	(-1.42)	(-2.58)	(-2.00)
Bias-corrected	-0.536***	-0.388**	-0.536***	-0.391***	-0.490**	-0.376*	-0.624***	-0.446**
	(-2.85)	(-2.46)	(-2.99)	(-2.58)	(-1.98)	(-1.75)	(-2.90)	(-2.32)
Cluster	No	No	Rank	Rank	Fund	Fund	Month	Month
Bandwidth Left	42.43	98.66	42.21	98.60	47.34	102.7	34.27	69.42
Bandwidth Right	42.43	42.40	42.21	41.38	47.34	43.81	34.27	35.85
Observations	31668	31668	31668	31668	31668	31668	31668	31668
Effective Obs.	10597	16938	10597	16911	11399	17331	9166	13868

Panel B: Flows at Five Globe Breakpoint

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	-0.747***	-0.503**	-0.777***	-0.528***	-0.712**	-0.477*	-0.846***	-0.585***
	(-2.98)	(-2.56)	(-3.19)	(-4.45)	(-2.28)	(-1.81)	(-4.59)	(-3.09)
Bias-corrected	-0.821***	-0.578***	-0.852***	-0.603***	-0.807***	-0.564**	-0.883***	-0.632***
	(-3.27)	(-2.94)	(-3.50)	(-5.09)	(-2.59)	(-2.14)	(-4.79)	(-3.34)
Cluster	No	No	Rank	Rank	Fund	Fund	Month	Month
Bandwidth Left	19.31	22.15	17.55	16.34	21.88	25.03	13.18	15.54
Bandwidth Right	19.31	77.68	17.55	82.38	21.88	83.75	13.18	65.97
Observations	32241	32241	32241	32241	32241	32241	32241	32241
Effective Obs.	7041	15694	6765	15758	7855	16440	5240	14071

Table 5
Fund Flows in Response to Sustainability Rating Matching on Pre-period Variables

This table reports the average treatment effect from nearest neighbor matching of Globe ratings on mutual fund flows. In Panel A funds are matched within Morningstar star rating based on flows, size, return over the prior 12 months and fund age, each based on the value before the publication of the rating in February 2016. In Panel B funds are matched on these characteristics as well as their loadings on Vanguard benchmark portfolios using the methodology of Berk and Van Binsbergen (2015) described in Section 4.2.1. An extreme rated fund is matched to another fund, based on the initial rating in March of 2016. Matching is adjusted for the bias discussed in Abadie and Imbens (2006; 2011). Abadie-Imbens standard errors are used and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Matching on Characteristics

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
1 Globe	-0.715***		-6.700***	
	(-9.07)		(-11.60)	
5 Globes		0.206***		3.799***
		(2.60)		(7.44)
Observations	33262	33262	33262	33262

Panel B: Matching on Characteristics and Loadings

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
1 Globe	-0.523***		-4.391***	
	(-5.91)		(-6.36)	
5 Globes		0.185***		5.589***
		(2.68)		(9.75)
Observations	33232	33232	33232	33232

Table 6
Fund Liquidation Based on Globe Rating

This Table examines how the probability of mutual fund liquidation varies with Globe ratings. A dummy variable equal to one if a fund is liquidated is examined. In Column 1 we show the raw counts and proportion of funds liquidating, while in the other columns it is regressed on dummy variables for globe rankings. Columns 4 and 6 include category by year by month fixed effects. Columns 5 and 6 include the additional controls return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating in the prior month and the log of fund age. Data is restricted to March 2016 and after excluding the final month, the period when the Globe ratings were published and analysis is at the fund level. Standard errors are clustered by month and fund, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Summary Stats	Regressions				
	(1)	(2)	(3)	(4)	(5)	(6)
1 Globe	13	0.00239**	0.00250***	0.00222**	0.00238***	0.00121
	[0.41%]	(2.50)	(2.78)	(2.34)	(2.58)	(1.13)
2 Globes	8	-0.000613				
	[0.11%]	(-0.82)				
3 Globes	21					
	[0.17%]					
4 Globes	15	0.000194				
	[0.19%]	(0.45)				
5 Globes	6	-0.000114				
	[0.16%]	(-0.09)				
Cat by YM FE		No	No	Yes	No	Yes
Other Controls		No	No	No	Yes	Yes
R^2		0.000327	0.000285	0.0191	0.00548	0.144
Observations	34046	34046	34046	34046	32422	32421

Table 7
Institutional Share Class Flows in Response to Sustainability Rating

This Table shows how mutual fund flows vary with Globe ratings comparing institutional to non-institutional share classes. Fund flows are regressed on dummy variables for globe rankings, a dummy variable equal to one if the share class is institutional and interactions of globe ratings and the institutional dummy variable. Columns labeled "Non-Inst" include only non-institutional share classes, columns labeled "Inst" include only institutional share classes and columns labeled "All" include all share classes. Columns one through three and five through seven include category by year by month fixed effects while columns four and eight include fund by month fixed effects. All columns include the additional controls return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating in the prior month and the log of fund age. Analysis is at the share class level. Standard errors are clustered by month and fund, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Flow				Normalized Flow			
	(1) Non-Inst	(2) Inst	(3) All	(4) All	(5) Non-Inst	(6) Inst	(7) All	(8) All
1 Globe	-0.271**	-0.176	-0.219*		-3.168***	-1.926	-2.940***	
	(-2.31)	(-0.71)	(-1.85)		(-3.85)	(-1.48)	(-3.52)	
2 Globes	0.0224	0.0255	0.0279		-0.126	-0.186	-0.130	
	(0.27)	(0.19)	(0.34)		(-0.22)	(-0.26)	(-0.23)	
4 Globes	0.0527	-0.00696	0.0594		0.121	0.822	0.132	
	(0.54)	(-0.05)	(0.63)		(0.25)	(1.22)	(0.27)	
5 Globes	0.327*	0.551*	0.347**		2.372**	3.854***	2.363**	
	(2.14)	(2.19)	(2.28)		(2.39)	(3.33)	(2.43)	
1 Globe*Institutional			-0.0579	0.0257			0.409	0.765
			(-0.23)	(0.09)			(0.31)	(0.45)
2 Globes*Institutional			0.0111	-0.0403			-0.115	-0.197
			(0.07)	(-0.23)			(-0.16)	(-0.22)
4 Globes*Institutional			-0.0924	-0.180			0.512	0.114
			(-0.57)	(-1.04)			(0.71)	(0.13)
5 Globes*Institutional			0.0970	0.0596			1.190	1.271
			(0.36)	(0.22)			(1.08)	(1.19)
Diff: 5 Globe-1 Globe	0.598	0.727	0.565		5.539	5.780	5.303	
P-value: 5 Globe=1 Globe	0.00781	0.0581	0.0104		0.00119	0.00831	0.00155	
Cat by YM FE	Yes	Yes	Yes	No		Yes	Yes	No
Fund by YM FE	No	No	No	Yes	No	No	No	Yes
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.0823	0.0859	0.0756	0.398	0.170	0.139	0.161	0.492
Observations	71771	23351	95136	84739	71771	23351	95136	84739

Table 8
Returns Based on Globe Rating

This Table shows how mutual fund performance varies with Globe ratings. In Panels A and B performance is regressed on a one globe and a five globe dummy variable. In Panel A regressions are value weighted based on the prior month's TNA and in Panel B regressions are equal weighted. Column 1 shows returns in excess of the risk free rate. Columns 2, 3 and 4 show returns in excess of a benchmark. The Benchmark in Column 2 is the value weighted return in a Morningstar category. Column 3 measures returns in excess of Vanguard benchmarks using the methodology from Berk and Van Binsbergen (2015). Column 4 estimates returns in excess of a fund benchmark based on the market, SMB, HML and momentum. The benchmark's in Columns 3 and 4 use fund-specific beta estimates from the two years prior to the globe rating. Below the regression, the difference between five and one globe funds is reported along with the p-value for the test that they are equal. Regressions are at the fund level and returns are in percentages. Standard errors are clustered by month and fund, and t-statistics are in parentheses. In Panel C, portfolios are formed based on globe ratings. The difference portfolio long five globe stocks and short one globe stocks is regressed on the market in the "CAPM" Columns and on the market, size, value and momentum in the "4-factor" Columns. Returns are before fees and data is restricted to March 2016 and after, the period when the Globe ratings were published. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value Weighted

	Excess Return	Morningstar Benchmark	Vanguard Benchmark	4-Factor Benchmark
	(1)	(2)	(3)	(4)
1 Globe	0.311**	0.0514	0.209**	0.159
	(2.01)	(0.44)	(2.43)	(1.24)
5 Globes	-0.252**	-0.158*	-0.0995	-0.193
	(-2.57)	(-1.87)	(-0.69)	(-1.35)
Diff: 5 Globe-1 Globe	-0.563	-0.209	-0.309	-0.351
P-value: 5 Globe=1 Globe	0.0247	0.236	0.131	0.165
R^2	0.00144	0.00224	0.000798	0.00149
Observations	34083	34083	33307	33307

Panel B: Equal Weighted

	Excess Return	Morningstar Benchmark	Vanguard Benchmark	4-Factor Benchmark
	(1)	(2)	(3)	(4)
1 Globe	0.0924	0.0776	0.113	0.0782
	(0.88)	(0.94)	(1.32)	(0.87)
5 Globes	-0.0961***	-0.102**	-0.0494	-0.150***
	(-2.65)	(-2.00)	(-0.45)	(-2.59)
Diff: 5 Globe-1 Globe	-0.189	-0.179	-0.163	-0.228
P-value: 5 Globe=1 Globe	0.0906	0.122	0.238	0.103
R^2	0.000152	0.000620	0.000129	0.000466
Observations	34095	34095	33319	33319

Panel C: Portfolios

	Value Weighted		Equal Weighted	
	(1) CAPM	(2) 4-Factor	(3) CAPM	(4) 4-Factor
Long 5 Globe - Short 1 Globes	-0.460*	-0.479*	-0.138	-0.173
	(-2.03)	(-2.14)	(-1.01)	(-1.33)
Observations	11	11	11	11

Table 9
Experimental Results

This Table shows how Globe ratings impact expectations of returns, risk and portfolio allocations in an experimental setting. Panel A examines MBA students while Panel B examines MTurk subjects. Dollar allocation amounts are regressed on performance expectations and globe rating dummy variables. Below the regression, the difference between five and one globe funds is reported along with the p-value for the test that they are equal. Column 4 includes subjects indicating they did not consider environmental, social or governance (ESG) factors when making decisions while Column 5 includes subjects that indicated that they did consider ESG factors. All regressions include subject fixed effects. Standard errors are clustered by subjects, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: MBA Students

	All			No ESG Consideration	ESG Consideration
	(1)	(2)	(3)	(4)	(5)
Performance	75.14***		71.32***	92.04***	53.92***
	(5.44)		(5.22)	(3.81)	(3.44)
Risk	-54.83***		-49.73***	-32.67	-59.70***
	(-4.60)		(-3.99)	(-1.52)	(-4.20)
1 Globe		-50.56**	-27.99	-13.89	-30.82
		(-2.24)	(-1.32)	(-0.43)	(-1.13)
5 Globes		57.36***	20.11	-8.080	48.51*
		(2.78)	(1.00)	(-0.27)	(1.75)
Diff: 5 Globe-1 Globe		107.9	48.10	5.809	79.33
P-value: 5 Globe=1 Globe		0.0000329	0.0485	0.876	0.0140
Acct FE	Yes	Yes	Yes	Yes	Yes
R^2	0.767	0.718	0.770	0.770	0.773
Observations	807	807	807	354	450

Panel B: MTurk Subjects

	All			No ESG Consideration	ESG Consideration
	(1)	(2)	(3)	(4)	(5)
Performance	58.29***		51.43***	51.43***	50.54***
	(9.38)		(8.07)	(3.96)	(7.06)
Risk	-30.69***		-25.58***	-31.42***	-23.18***
	(-5.13)		(-4.31)	(-3.25)	(-3.06)
1 Globe		-65.69***	-39.28***	-30.29	-43.66***
		(-5.02)	(-3.15)	(-1.49)	(-2.73)
5 Globes		64.43***	31.74**	11.44	42.75***
		(4.89)	(2.48)	(0.53)	(2.68)
Diff: 5 Globe-1 Globe		130.1	71.03	41.73	86.42
P-value: 5 Globe=1 Globe		5.26e-16	0.00000210	0.103	0.00000283
Acct FE	Yes	Yes	Yes	Yes	Yes
R^2	0.755	0.719	0.763	0.812	0.725
Observations	1728	1728	1728	624	1101

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

ECGI Working Paper Series in Finance

www.ecgi.global/content/working-papers

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html