DO MUTUAL FUNDS REPRESENT INDIVIDUAL INVESTORS?

Jonathon Zytnick*

Georgetown University Law Center

October 7, 2022

* Jonathon Zytnick is an Associate Professor at Georgetown University Law Center. He can be reached by phone: (202) 662-9817, email: jz859@georgetown.edu. For valuable comments and discussions I am grateful to Patrick Bolton, Alon Brav, Ryan Bubb, Chuck Callan, Harrison Hong, Edwin Hu, Robert Jackson, Enrichetta Ravina, José Scheinkman, Arielle Zimran, and participants in Columbia University's Financial Economics Seminar, the 2022 American Law and Economics Associate Annual Meeting, and the Northwestern Law School Law and Economics Seminar.

Abstract

In recent decades, investors have wielded their voting power to influence the direction of corporate policy. Although mutual funds have widely varying voting patterns and predictable ideological disagreements, little is known about whether their underlying investors have similar preferences. I provide the first systematic documentation comparing the voting preferences of individual investors in the United States to those of the mutual funds they invest in. I begin by presenting aggregate voting statistics, showing that, in the aggregate, individual investors vote similarly to mutual funds on SRI and management proposals. Since these numbers come from the subset of individual investors who vote, I adjust for sample selection bias on unobservable variables using access to additional voting methods, which provides variation in turnout that is unrelated to voting choices. I show (preliminarily) that sample selection bias does not appear to have a large aggregate impact. I next turn to the relationship between how individuals vote and how the funds they own vote. I find that although individual investors are highly ideological in their voting, there is generally no relationship between how a fund votes and the preferences of its individual investors. I explore the sources of this divergence, providing evidence for limited attention of individual investors.

JEL Classification: G11, G18, G23, G34, G38, D72.
Keywords: Individual Investor Voting, Shareholder Proposal, Corporate Governance.

I. Introduction

"There is another government… consisting of a handful of gigantic institutional asset managers… who own (on behalf of their customers) most of the stocks of most of the public companies, and can, in some loose sense, tell those companies how to behave. They are not chosen democratically, exactly, but they are *representative*; millions of people give their money to those institutions and trust them to make decisions for them." Matt Levine, Bloomberg Opinion, June 25, 2020 (emphasis in original).

The majority of investment by individuals is conducted via institutional intermediaries—especially mutual funds—who vote on their behalf. Institutional investors wield vast voting powers and take strong and consistent ideological positions, forcing shifts in corporate policies (Bubb and Catan (2020), Bolton et al. (2020)). But despite extensive research on mutual fund voting, there has been little research on the preferences and voting behaviors of individual investors. As observed in Bolton et al. (2020), "[w]hether these ideological differences [across institutional investors] reflect the ideology of the institutions' client bases, we cannot say. It is not even clear that clients are aware that the funds they invest in have systematic ideological biases."

In this paper, I use novel microdata on individual investor ownership and voting to examine whether variation in mutual fund voting choices, which has previously been shown to be strong and consistent, is reflected in the preferences of the underlying investors. I first present aggregate summary statistics on voting choices and correct for sample selection bias on unobservable variables by exploiting a variable that shifts turnout without connecting to ideology. I then describe patterns in individual investor voting, showing they have their own ideologies which are strongly held and highly consistent. Finally, I show most funds' voting is entirely uncorrelated with the preferences of the investors who own the fund.

These findings present a puzzle: why can institutions wield power with no fidelity to the preferences of their underlying investors? I show evidence that individual investors are constrained by a lack of information about the ideologies of their mutual funds. This finding calls into question the legitimacy of interventions by large institutional investors, particularly those related not to specific details of firm management but rather broad questions of corporate

purpose. Although individual investors' demand function has a strong ideological component, mutual funds' decisions to exercise power do not derive from an informed marketplace for ideology.

A. Overview

To date, researchers have been unable to explore questions on the ideological preferences of individual investors due to lack of data on individual preferences and shareholdings. In this paper, I introduce novel microdata on individual ownership and voting to compare mutual fund votes to the votes of individuals who own those mutual funds.

I begin with investor preferences, presenting some aggregate summary statistics. Contrary to conventional wisdom, individuals are not, in the aggregate, more pro-social responsibility than funds; in fact, they vote quite similarly.

Next, I evaluate the extent to which the preferences of individual investors who vote can be used as a proxy for the preferences of individual investors who do not. To adjust for sample selection bias on unobservable variables, I use an "excluded variable" which increases individual investor turnout but is not directly connected to ideology—whether they receive a mail ballot permitting them to vote by mail. I also use inverse probability weighting to adjust for selection on observable variables. I find, preliminarily, that sample selection bias does not appear to play much of a role in individual investor preferences.

To better understand whether mutual funds represent individual investor preferences, I focus on social responsibility proposals, which relate more to differing corporate objectives and feature fewer firm-specific information asymmetries than other proposal types, and compare individual and mutual fund voting pattern. However, variation in individual voting differs from variation in fund voting. Individuals vote highly consistently across their holdings and over time; they do so to a much greater extent than do funds, whose votes tend to shift based on the proposal at issue. Individual investors are best characterized as polarized and ideological.

I next turn to the relationship between individual preferences and those of the specific funds they own. I find that individuals—the same individuals who vote in a highly ideological fashion on their direct holdings—do not sort into funds along ideological lines. Given that funds and individuals both have strong and consistent ideological preferences, we might expect a close

association between individual and fund ideologies. But instead, on average, two funds with identical objects but vote oppositely have ideologically identical customers.

An exception is environmental, social, and governance (ESG) funds, which explicitly define themselves by their ideological mission. The individual investors in ESG funds have distinctly different ideologies from the investors of other funds. As a result, ESG funds appear to serve a unique role in the mutual fund ecosystem, as a channel for individual ideology.[1]

Why don't shareholders choose like-minded funds? I find evidence that investors have limited attention: the costs of acquiring more granular detail about funds, as compared to readily available information, exceed the benefits. In their choice of stocks, funds, and votes, I show that individuals consistently choose ideologically based on high-salience information but not low-salience information. Consistent with this possibility, investors who make larger investments feature a significantly stronger relationship between their own ideology and that of the funds in which they invest.

This is a macro-finance paper. My contribution is not to identify a new factor that individuals incorporate in their portfolio purchasing decisions; in fact, to the extent that individuals sort into like-minded mutual funds, such sorting may be based on the mutual fund's overall sales pitch, not its votes. Rather, it is to document, at an economy-wide level, the extent to which intermediated ownership channels the preferences of underlying beneficiaries. I show that individuals have ideological preferences that connect their voting, equity purchases, and purchases of ESG funds. Mutual funds also vote in ideologically distinct blocs, but they don't advertise it. The result is an ideological party structure without party members: mutual fund voting is correlated with individual investor ideology only to the extent that pro-SRI shareholders sort into a small bloc of highly ideological ESG funds.

B. Literature Review

This paper contributes to a new literature studying the relationship between individual ideological preferences and fund choices. While much is known about the voting behavior of institutional intermediaries, little is known about the preferences of their underlying investors. Aggarwal, Erel, and Starks (2014) document a positive relationship between fund voting and

[1] I find similar results using the political voting results of the zip codes of mutual fund holders, which do not suffer from sample selection bias.

public opinion over time. Brav, Cain, and Zytnick (2021) describe the voting of individual investors who directly hold equities. Gordon and Gilson (2013) and Bebchuk, Cohen, and Hirst (2017), describe the potential for agency conflicts by institutional investors in their stewardship of firms.

A large body of literature describes mutual fund voting preferences. Matvos and Ostrovsky (2010) first established that mutual funds vote ideologically, and Bubb and Catan (2020) and Bolton et al. (2020) both map out the dimensions of institutional investor ideology.

Several papers focus on the causes of the rise of ESG funds. Riedl and Smeets (2017) combine administrative data from a Dutch mutual fund provider with survey and experimental evidence to establish that individuals who invest in socially responsible funds are willing to forego returns and have pro-social preferences. Hartzmark and Sussman (2019) show that fund flows are connected to Morningstar's sustainability ratings. Relatedly, a growing literature discusses the social responsibility of firms and how firm social responsibility relates to the firm's investors (Hart and Zingales (2017), Dyck, Lins, Roth, Wagern (2019), Broccardo, Hart, and Zingales (2020)).

The concept of representation is the subject of a large body of writing in political science (see Urbinati and Warren (2008) for a review of theoretical developments on the topic). I incorporate concepts from the representation literature into the field of shareholder voting.

The rapid growth of ESG funds over the past decade and the growing importance of SRI proposals suggest that some shareholders may have objectives other than maximizing risk-adjusted returns, or, in the alternative, different beliefs about how to maximize risk-adjusted returns. Hart and Zingales (2017) argue that firms ought to maximize the utility of shareholders rather than firm value, and they propose shareholder voting as a mechanism to achieve such optimization.[2] However, since most shareholders are institutional intermediaries, the interests of the entities casting ballots may not align with the interests of their underlying beneficiaries (see, e.g., Gordon and Gilson (2013) and Bebchuk, Cohen, and Hirst (2017)).

Given that the mutual fund industry is competitive (Wahal and Wang (2013)) and mutual fund voting is publicly disclosed, we may expect the U.S. mutual fund industry to produce

[2] I note also the popularity, both in academia and the popular press, of the notion of "shareholder democracy", referring to the primacy of shareholders to make decisions about the firm through voting (Fairfax (2009), Fos and Tsoutsoura (2014), Levit, Malenko, and Maug (2019), Gantchev and Giannetti (2018), Sorkin (2019)).

ideological representation similar to U.S. political institutions. The U.S. Securities and Exchange Commission has created a regulatory disclosure system explicitly justified by the notion that individuals ought to be able to choose funds based on the funds' voting patterns, and it has recently emphasized the importance of funds voting in line with investor expectations.[3] My results show that, to the extent that the industry is representative, it is only because ESG funds constitute an overtly ideological "party." Increasing representation may require policies that convey less granular disclosure to individuals. Otherwise, as the twin trends of delegated investment to intermediaries and social responsibility voting continue to grow in size and importance, there may be a growing gulf between the preferences of the individual investor and how the investor's money is voted.

This paper is organized as follows. In Section II I describe the data. In Section III I provide background information on shareholder voting and ownership, including new empirical facts. In Section IV, I compare aggregate individual votes to aggregate fund votes and attempt to adjust for sample selection bias. In Section V, I compare the patterns of variation across votes and voters for individuals and funds. In Section VI, I show the extent to which variation in fund voting mirrors variation in the voting by their underlying shareholders. In Section VII I deepen the analysis by examining possible explanations for the results in the previous sections. Section IX concludes.

II. Data

In this paper, I introduce an individual-level dataset of U.S. investors. These data contain, for U.S. individual investors, the public equities that they beneficially own; the proposal-by-proposal votes those individual investors cast on those equities; and the mutual funds that they own (for roughly one third of U.S. mutual funds). The data come from Broadridge, the service provider that handled shareholder voting for virtually all brokers, for both individual and institutional investors, for the data period 2015 – 2017. The set of direct ownership of, and votes

[3] The SEC's main rationale for requiring mutual funds to disclose all votes was to enable "fund shareholders to monitor their funds' involvement in the governance activities of portfolio companies." (SEC 2003). In March 2021, the Acting Chair of the SEC pledged to "examin[e] proxy voting policies and practices to ensure voting aligns with investors' best interests and expectations." (SEC 2021), and, in a separate speech, stated that, "[u]ltimately, corporate accountability is only possible when the funds that manage American investors' savings diligently exercise their authority to vote, clearly disclose their votes to investors, and operate in a system that efficiently provides accurate information about vote execution" (Lee 2021).

on, public company equities was first introduced in Brav, Cain, and Zytnick (2020), and the set of mutual fund ownership (which includes both open-end and closed-end funds) is newly introduced in this paper.[4]

For ownership of public equities and of mutual funds, at the time of a shareholder meeting of the fund or firm, I observe, for each individual investor account, among other things: (i) its number of shares owned, (ii) whether it voted, (iii) how it voted, and (iv) its zip code. For any firm or fund with a shareholder meeting, these data contain nearly-complete snapshots of individual beneficial ownership of the fund or firm at the time of its shareholder meetings (annual or special) during the period 2015 – 2017. Each account has a unique identifying number, allowing an account's entire portfolio to be linked across securities and over time.

The data's coverage of a fund depends on the frequency of the fund's shareholder meetings. Regular public equities and closed-end mutual funds are required to hold annual shareholder meetings, so for such entities, I observe annual ownership snapshots for the three-year period. However, open-end mutual funds are not required to hold annual meetings; generally, an open-end mutual fund holds a meeting when it wishes to revise its operating agreement. Consequently, I have no snapshot of ownership for many open-end funds operating in the period 2015-2017. Those open-end funds which did hold a meeting in the period 2015-2017 often held only a single meeting, so, for such funds, I have one snapshot in the three-year period.

To summarize, if mutual fund f has a shareholder meeting in March 2017 with a record date of December 2016, I would observe all accounts that own the mutual fund as of December 2016 with their number of shares. Because account identification numbers are consistent across sets, I also observe the account owner's zip code, other funds and direct equity holdings. Finally, for a's direct equity holdings, I observe whether she voted and how she voted.

Brav, Cain, and Zytnick (2020) extensively describe the dataset of direct ownership of operating company equities, including methodology in merging to other datasets and summary statistics. In the Data Appendix, I discuss in detail the data and methodology for cleaning and merging the Broadridge data on retail ownership of mutual funds, including tables describing the coverage of the data.

[4] Throughout, I distinguish between two types of ownership: direct ownership of firms, which I also refer to as operating companies or public equities, and indirect ownership via mutual funds.

Using these datasets as a base, I conduct an extensive data collection process. Mutual funds are required to report their votes on Form N-PX; I collect these votes, along with proposal-level information, from ISS Voting Analytics, which I merge by meeting and proposal to the Broadridge datasets using a combination of algorithmic matching and hand-matching. I divide proposals into categories as described in Brav, Cain, and Zytnick (2020).

I obtain financial information on firms, including share prices, from the CRSP monthly securities file and Compustat fundamentals annual. Following Hong and Kostovetsky (2012), I obtain SRI ratings on firms from MSCI (formerly KLD). I use SIC codes to determine if firms are in fossil fuels industries. I designate renewable energy firms as those that appear on NASDAQ's Clean Edge Green Energy index or Wilderhill's New Energy Global Innovation Index, which I hand-match by name to the set of firms.

I categorize a mutual fund as an explicitly-labeled ESG fund as follows. I begin with a list of sustainable funds in Appendix A1 of Hale (2018).[5] I then add to it funds that have "Sustainable", "ESG", "Social", or "Clean Energy" in their name, confirming in their prospectuses that they incorporate ESG criteria. Finally, I include all funds that belong to any of five fund families that are explicitly centered on ESG themes in their family-wide branding: Calvert, Parnassus, Pax, Praxis, and Trillium. No other fund family with explicit ESG themes in their family-wide branding had funds with shareholder meetings in the 2015 – 2017 period.

The CRSP Open-End mutual fund dataset provides information on open-ended mutual funds, including fund management, returns, and whether it is an index fund. I merge by fund name to the ISS Voting Analytics N-PX dataset, beginning with an algorithmic merge, hand-checking each match, and then hand searching for all unmatched funds. The result is a dataset combining fund votes from ISS Voting Analytics with fund features from CRSP. I use the CRSP monthly securities file to add prices of closed-end funds, and hand-merge these closed-end funds by name to the ISS Voting Analytics N-PX dataset. I merge this combined dataset to the Broadridge dataset by fund CUSIP, and hand-match funds that do not match by name.

For zip code level political voting outcomes, I use 2016 Presidential returns from the MIT Election Data and Science Lab. As in Brav, Cain, and Zytnick (2020), I obtain zip code-

[5] Hale (2018), as part of a Morningstar research report, begins with all funds that, in their prospectus, "state that they incorporate ESG criteria into their investment processes, or indicate that they pursue a sustainability-related theme, or seek measurable sustainable impact alongside financial return." He then removes funds that invest based on values rather than sustainability.

level demographic information from the Census Bureau and zip code-level adjusted gross income from the IRS website. As an additional measure of investor age, I identify investors who hold a "Target Retirement Date" fund, which I identify as funds that have "Target" in their fund name but do not have the words "College" or "Maturity", and hand-verify that each of the funds represents a target date retirement fund. I then use the retirement year in the fund name to impute a proxy for the age of the individual investor.

Data Appendix Table 1 shows I have roughly 30% of the open-end funds in CRSP's mutual fund database and slightly more than 50% of the asset value, along with almost all closed-end funds that appear in CRSP's monthly securities dataset.

In total, I have 80,209,211 unique fund-owning accounts, 16.7% of which also own direct equity securities. (I have 46,686,015 unique equity-owning accounts, as described in Data Appendix Table 5.)

III. Background

Mutual funds own assets, which may include publicly traded equity securities, and the fund's investors hold shares in the fund; thus, the individual investor is the ultimate economic beneficiary of the fund's performance but has no right to cast ballots in the underlying securities. Instead, a mutual fund's voting choices are determined by the fund's investment advisor and may be determined at the level of the individual fund or together with other funds of an adviser. Mutual funds are commonly regarded as having a fiduciary duty to vote, and mutual fund voter turnout is virtually 100%.

A substantial portion of individual investors directly hold public equities, permitting them to vote directly in shareholder meetings. In Figure 1, I display the distribution, across mutual funds, of the number of individual investors in each mutual fund; the number of individual investors in each fund who also own operating company equities; and the number of individual investors in each fund who also own and vote on operating company equities. A relatively small group of superstar funds have a disproportionate number of individual investors—the right side of Figure 1 is convex even on a logarithmic scale. Averaged across equal-weighted funds, 30.4% of individuals who own mutual funds own at least one equity, and 8.2% of individual fund investors cast at least one ballot on an equity holding in the three-year period.

Shareholder votes at firms are often significant focal points of contention. Shareholders vote on management-sponsored proposals for which investor approval is required for the firm to take an action; shareholder-sponsored governance proposals, which affect the power structure of the firm; and shareholder-sponsored social responsibility proposals (SRI proposals), which focus on social and environmental goals.

Voting choices may center on differences in beliefs regarding how to accomplish the shared objective of maximizing firm value (Kakhbod, Loginova, Malenko, and Malenko (2020)), or differences in what the objective of the firm should be (Hart and Zingales (2019)).

Most shareholders, whether institutions or individuals, vote remotely by proxy rather than physically at the shareholder meeting. Kahan and Rock (2008) detail the mechanics by which shareholders of firms are identified and notified and how they cast their ballots. Individual investors receive either a mailed package of materials or a notice entitling them to log onto a website to view materials. On the actual ballot, the individual sees the text of the proposal and the management recommendation (for SRI proposals, nearly always "Against"). Funds generally use a proxy voting service—ISS, Glass Lewis, or Broadridge— to facilitate the voting mechanics. Many funds also use proxy advisors, most commonly ISS or Glass Lewis, who guide them on how to vote.

In this paper, I largely focus on a single category of corporate proposals: shareholder SRI proposals. SRI proposals, which include both environmental and social proposals, provide a setting in which mutual funds have a minimal information advantage as compared to individual investors. Disagreement on SRI proposals likely arises from disagreement as to the objective function of the corporation, in which case the parties have a genuine conflict of interest, not information asymmetry. In the alternative, to the extent that disagreement on environmental and social proposals reflects differences of opinion about how best to maximize corporate value, as opposed to a disagreement as to the objective function of the corporation, assessing the long-run relationship between firm profitability and environmental and social practices does not require granular firm-specific knowledge. By contrast, if a fund's votes on, say, director elections diverge from those of its underlying investors, such divergence may simply reflect the fund's superior knowledge of the directors and firms.

IV. Individual Preferences and Sample Selection Bias

How do individuals and funds vote, in the aggregate? Table 1 shows descriptive statistics comparing how individuals vote to how funds vote. Panel A show simple aggregates across three types of proposals: shareholder social responsibility proposals, shareholder governance proposals, and management proposals. The table shows that on SRI proposals and management proposals, the two groups are quite similar, contrary to conventional wisdom. On governance proposals, however, individual investors are far more pro-management.

Panel B goes into more detail on SRI proposals, with more divisions in the table. The bottom portion of the table shows statistics on proposals reaching 50% in favor.

In the aggregate, individuals vote similarly to funds on SRI proposals. The inclusion criteria—who gets counted—and weighting rules—how votes are aggregated—make a large difference to the level of support for SRI proposals, though individuals vote similarly to funds for each set of rules.

Some empirical facts are clear from the table: (1) Individuals and funds with more shares tend to vote against SRI proposals, evidenced by value-weighted support that is substantially lower than support measured using other aggregation schemes. (2) Individuals and funds favor SRI proposals more than management does and less than ISS does; Glass Lewis, the second largest proxy advisor, votes similarly to individuals and funds. (3) Although there is a great deal of voting support, by both individuals and funds, for SRI proposals, few SRI proposals garner 50% support from either group.

A. Overview of Sample Selection Bias

In the results so far, I have taken individual shareholders who vote to be representative of all individual shareholders. However, only a fraction of shareholders who directly own stocks vote them. To what extent can we learn about the preferences of non-voting investors from the votes of voting investors? In this section, I assess the extent to which sample selection bias affects my assessments and conclude that sample selection bias does not meaningfully affect the results.[6]

[6] Sample selection bias, as will be apparent later in the section, has much in common with omitted variable bias. In fact, the tools used to identify and correct for sample selection bias have a strong resemblance to the tools used for causal identification. However, few papers describe the process of curing sample selection bias as they would describe the process of causal inference. An exception is Zimran (2019), on which this section relies heavily, which describes sample selection bias in terms that would be familiar to students of the identification revolution.

Suppose that each individual investor has an ideological position, $R_i \in [0,1]$, that can be observed through this individual's votes. The individual's ideology is given by the equation:

$$R_i = \alpha + x_i\beta + \varepsilon_i \qquad (1)$$

Here, x_i represents a vector of (demeaned) observable covariates that predict ideology, perhaps the dollar value of her portfolio or the mean income of her zip code.[7] My aim is to measure α, the intercept, which is equal to the average ideology among individual investors.

The problem of sample selection bias arises because I can only observe R_i for those individual investors who choose to vote. For simplicity, assume each investor has the opportunity to cast a single ballot, on which her votes on various proposals would reveal her ideology.[8] Her decision to vote—which is to say, to be included in the sample—is given by a second equation.

$$v_i^* = \mu + x_i\gamma + M_i\delta + u_i \qquad (2)$$

v_i^* is individual i's unobserved net utility from voting. She votes ($v_i = 1$) if and only if $v_i^* \geq 0$. In equation (2), I permit the net utility from voting to be impacted by a vector of variables, z, that do not affect her ideology, which will be critical later for measuring and correcting for sample selection bias. We may conceptualize M_i as her access to additional voting methods, which increases turnout but is unconnected to her ideology. In this model, v_i, x_i, and M_i are observed for all observations and R_i is observed only when $v_i = 1$.

The selection equation is given by:

$$\Pr(v_i = 1|x_i, z_i) = G(\mu + x_i\gamma + M_i\delta) \qquad (3)$$

Since I only observe ideology, R_i, for those shareholders who vote, a regression analysis on voters would yield $\hat{\alpha} = E[R_i|x_i; v_i = 1]$, which may be different from $E[R_i|x_i]$, the average ideology of all investors. The difference between them is equal to $E[\varepsilon_i|x_i, M_i, v_i = 1]$. Under mild assumptions, we can write this difference as follows:[9]

---

[7] The error term, ε_i, contains all unobserved predictors of ideology and is, by construction, uncorrelated with x_i. I assume that an individual investor's ideology is revealed by her votes on management and shareholder proposals.

[8] As in Zimran (2019), I model the decision to be selected into the sample as a single once-in-a-lifetime decision, even though in practice whether to vote is a series of sub-decisions that the voter makes over time.

[9] The standard selection model also requires certain mild technical assumptions on G. See Zimran (2019), footnote 6. Different sample selection methods require different assumptions on G. The most commonly used, the Heckman

$$\Omega(\mu + x_i\gamma + M_i\delta) \equiv E[\varepsilon_i | x_i, M_i, v_i = 1]$$

$\Omega(\cdot)$ is a function that captures the difference in average ideology between voters and non-voters (conditional on observables). It is the error term for an estimate that does not adjust for unobserved determinants of ideology when those determinants of ideology differ between voters and non-voters. If the unobserved error in ideology, ε_i, is correlated with the unobserved error in voting, u_i, $\Omega(\cdot)$ will be non-zero.

We can now re-write equation (1) as follows:

$$R_i = \alpha + x_i\beta + \Omega(\mu + x_i\gamma + M_i\delta) + \xi_i \qquad (4)$$

Where the error term ξ_i is uncorrelated with u_i.By including the additive error term $\Omega(\cdot)$ in the equation, we have replaced the correlated error term ε_i with the uncorrelated error term ξ_i.

The empirical problem, then, amounts to estimating $\Omega(\cdot)$ for each observation. Note also that, after obtaining $\Omega(\cdot)$, I can make an adjustment not just for the aggregate average, α, but for each individual observation. In future drafts, I intend to use bias-corrected versions of each individual's ideologies throughout.

B. Empirical Strategy and Intuition

In adjusting for sample selection bias, we must distinguish between selection on observable variables and selection on unobservable variables. Selection on observable variables—x_i in the discussion above—can be corrected rather straightforwardly.[10] Conceptually, we can analogize correcting for selection on observables to controlling for covariates in a linear regression, though the actual approach differs substantially.

Correcting only for observable variables, however, raises the possibility that there are elements that determine R_i that differ between voters and non-voters—that is, u_i may be correlated with ε_i above. Selection on unobservable variables requires estimating the error function $\Omega(\cdot)$.

(1979) method, assumes u_i is normally distributed and therefore G is the normal CDF function. Das, Newey, and Vella (2003) propose a non-parametric approach which requires weaker assumptions on G, but which is not suitable to this setting because it identifies the model only to a constant.

[10] M_i, by definition, contains observable variables that do not appear in the equation for R_i and therefore selection on M_i requires no adjustment.

Correcting for selection on unobservable variables uses a strategy similar to causal identification with an instrumental variable. Consider the vector of variables M_i. M_i must satisfy the relevance condition—that is, it must appear in the equation for R_i, the outcome variable of interest. And M_i must satisfy the exclusion restriction—that is, it cannot appear in the equation for v_i^*, the selection equation. In this way, M_i is much like an instrumental variable, although we technically call it an "excluded variable" because it does not appear in the equation for R_i.

If the researcher has an excluded variable, M_i, that satisfies the relevance condition and the exclusion restriction, then variation in M_i shifts one's likelihood of being a voter without shifting ideology R_i. Due to the exclusion restriction, any change in the outcome variable R_i as M_i shifts must be due to selection.

Intuitively, consider that as Pr $(v_i = 1|x_i, M_i)$ approaches 100%, the selection error $\Omega(\cdot)$ approaches zero. The selection error grows as Pr $(v_i = 1|x_i, M_i)$ increases. With an excluded variable that affects voting but not ideology, I can identify the error term.

I have such a variable, based on prior research with additional confirmation provided here. Brav et al. (2022) details how the voting method options that firms make available to individual investors affect their turnout. I will use these voting methods as an exogeneous source of variation in turnout that does not relate to voting choices.

C. Relevance Condition

Because the findings from Brav et al. (2022) serve as the basis for the excluded variable, I briefly summarize them here and replicate them using my data. In summary, individual turnout rates are determined in large part by the voting methods they are allowed to use.

Individual investors can choose what package of voting materials to receive from the firms they own. For those shareholders who do *not* make a selection, the firm chooses what package of voting materials to send. The firm can send either a full package of materials or a one-page notice. By law, the firm's choice determines what voting methods are available to the shareholder. If the firm sends the shareholder a full package of voting materials, it includes a mailable paper ballot and a votable telephone number. If instead the firm sends the shareholder a notice, by law it does not include a mailable paper ballot or a votable telephone number.

Brav et al. (2022) show that receiving mail/phone voting options causes a large increase in individual investor turnout. They use a difference-in-difference specification to show that

when a shareholder is switched from full package (which includes the option to vote by mail) to notice (which does not), her turnout drops significantly. When a shareholder is switched from notice to full package, her turnout rises. Further, they use a second difference-in-difference setup to show that the turnout effect is entirely driven by the loss or gain of mail/phone voting options, as opposed to the different information content sent through the mail.

The first difference-in-difference captures the effects of the firm switching an individual's voting materials. I replicate it here.[11] The outcome variable is v_{ict}, whether individual i voted at firm c's meeting at time t. A simple difference-in-difference would compare turnout at switching firms, pre- and post-switch, to turnout at non-switching firms. An extra layer is added by comparing firms that sent full materials at time 0 (and potentially switched to notice) to firms that sent notice at time 0 (and potentially switched to full materials). A final layer exploits the fact that only individuals who made no selection are affected by the firm's choice. In sum, then, the right-hand-side contains the interaction of four variables. $switcher_c$ indicates whether firm c switches the materials it sends during the sample period, and $post_{ct}$ indicates whether the observation in question (at time t) is post-switch. If firms only switched in one direction, from notice to full materials, the interaction of $switcher_c$ and $post_{ct}$ with firm fixed effects would be sufficient to capture the firm-level turnout effect. Since firms switch materials in either direction, I further interact with $full_{c0}$ or $notice_{c0}$, which indicate firm c's materials at the start of the sample period. Finally, to take advantage of the fact that only those shareholders who did not select materials are affected by the firm's switch, I make it a difference-in-difference-in-difference by including $NoSelection_{ic0}$, whether shareholder i did not select materials for firm c at time 0.[12] I estimate:

[11] I limit to annual meetings and, for simplicity, I remove firms that send a mix of materials at any of their meetings in the sample period. I identify firms that switch from notice to full package or vice versa; I keep switcher and non-switcher firms but remove firms that switch more than once.

[12] I use $NoSelection$ at time 0 rather than time t to reduce any potential endogeneity from the (perhaps implausible) possibility that an investor would select or decline to select her materials based on features of the upcoming shareholder meeting. Because an investor's choice is sticky from year to year, $NoSelection_{ic0}$ is highly predictive of $NoSelection_{ict}$, though not identical. Thus, $NoSelection_{ic0}full_{c0}switcher_c post_{ct}$ is technically an instrument for $NoSelection_{ict}full_{c0}switcher_c post_{ct}$, the latter of which fully determines an individual's materials at time t. In Appendix Table A1, Columns 1 through 3, I display the full three-column instrumental variable results with first-stage, reduce form, and IV results, where the first stage contains the relationship between $NoSelection_{ic0}full_{c0}switcher_c post_{ct}$ and the individual's materials, the reduced form (Equation 1 above) contains the relationship between $NoSelection_{ic0}full_{c0}switcher_c post_{ct}$ and the individual's turnout, and the IV results

$$v_{ict} = \beta_1 NoSelection_{ic0} full_{c0} switcher_c post_{ct} +$$

$$\beta_2 NoSelection_{ic0} notice_{c0} switcher_c post_{ct} + \phi_{ic} + \psi_{it} + \theta_{ct} \quad (5)$$

The specification includes individual-firm, individual-year, and firm-year fixed effects, so that I compare an individual's turnout to the individual's turnout in the same year at other firms, to the individual's turnout in different years at the same firm, and to other individuals' turnout at the same firm meeting.

Table 2, Column 1, estimates Equation 5. Switching an individual from notice to paper increases turnout by 2.736 percentage points, and switching her from paper to notice reduces turnout by 2.255 percentage points. Appendix Table A1 effectively combines those two figures from Column 2 and arrives at an adjusted overall turnout effect of 2.373 percentage points. Figure 2, Panel A shows these results graphically, broken out by year and limited to firms that switch in 2017 (or do not switch at all) to demonstrate parallel pre-trends. Each coefficient represents, for selecters as compared to non-selecters and as compared to time $t = 0$, the difference in turnout between individuals at switching firms and non-switching firms. The solid (treatment group) lines match the dashed (control group) lines for the pre-trend, then significantly diverge when the firm switches materials.

The results of Table 2 and Figure 2 demonstrate that receiving a full package of materials rather than notice causes a substantial increase in individual turnout. In Brav et al. (2022), we use a second difference-in-difference to show that this effect is entirely attributable to the ability to vote by mail, which is only permitted when one receives the full package of materials.

D. Exclusion Restriction

Next, I turn to assessing whether the shareholder's receipt of mail ballots, m_{ict}, is related to her voting choices, which would violate the exclusion restriction.

Exclusion restrictions must be justified using institutional knowledge. Here, it is intuitive that receipt of the ability to vote by mail would increase turnout but would have no obvious relationship to one's voting choices.

adjust the reduced form results by the first stage results. Because the first-stage coefficients are nearly 1, as shown in Appendix Table A1, in the main text we present the reduced form results only.

Exclusion restrictions cannot generally be empirically verified. However, in this setting, there is a control group of individuals who are unaffected by the firm's choice m_{ct}—those who select their voting materials—allowing me to empirically substantiate the exclusion restriction. I discuss two possibilities that could create a relation between an individual's likelihood of receiving mail/phone voting, m_{ict}, and her voting choices.

Possibility 1: Direct Violations of the Exclusion Restriction

The first possibility is that receiving a full package as opposed to a notice could directly cause individuals to vote differently than they otherwise would.[13] In this section, I assess the question empirically.

To avoid selection effects, consider shareholders at firms that switch materials who are inframarginal on turnout—that is, they turn out before and after the firm's switch. And consider those shareholders whose materials are switched (because they did not make a selection of materials) as compared to the control group, those whose materials are not switched (because they did make a selection). If the switch of materials affects *how* shareholders vote, then their votes should change as compared to the control group.

To test this, I mimic Equation 5 above, but instead of the shareholder turnout at the meeting as the outcome variable, I substitute the shareholder's agreement with management at the meeting:

$$R_{ict} = \beta_1 NoSelection_{ic0} full_{c0} switcher_c post_{ct} +$$

$$\beta_2 NoSelection_{ic0} notice_{c0} switcher_c post_{ct} + \phi_{ic} + \psi_{it} + \theta_{ct} \qquad (5A)$$

I calculate R_{ict}, by aggregating individual i's votes at firm c's meeting at time t to calculate the fraction of her votes that agree with management at the meeting, then regress that using a triple difference specification on the interaction of switching firms, post-switch, and whether the firm

[13] I am skeptical that mailing a package of paper materials as opposed to a notice would affect voting choice. I note, for example, that Brav et al. (2022) show that the entire effect on turnout of the firm's choice is from losing or gaining mail/phone ballots, not from receiving a different form of reading materials in the mail. Individuals who previously received a notice and voted by internet, and then received mailed reading materials and the ability to vote by mail, experience no turnout change. Given that the form of reading materials does not matter to turnout choice, it would be unlikely to matter to voting decisions.

started as a mail firm or a notice firm. As before, I include individual-year, individual-firm, and firm-year fixed effects.

In Table 2, column 2, I show the results of this test, with full excluded variable results in Appendix Table A1, columns 4 through 6. There is no significant effect of the firm switching materials on voting choices. In fact, a switch from notice to paper and from paper to notice have (insignificant) coefficients with the same sign, when we would expect opposite coefficients if there were an effect. Column 6 of Appendix Table A1 estimates the overall effect at 0.143%, with a standard error of 0.569%. When firms switch shareholders' voting packages from full package to notice or vice versa, there is no evidence that affected and unaffected shareholders at the firm vote differently.

Possibility 2: Indirect Violations of the Exclusion Restriction

Consider a second, perhaps more plausible possible violation of the exclusion restriction. Firms know that the voting methods they provide affect shareholder turnout, so they may choose what materials to send strategically. If firms are more likely to send mail/phone ballots before a meeting at which management expects weak voting results, that will create a correlation between the proxy for turnout and the voting results that is not driven by selection into voting.

Indeed, I have in other work shown that firms make their decision of what materials to send strategically. They send paper ballots when the structure of the voting rules rewards high turnout—namely, proxy contests, where both sides are attempting to boost turnout of their supporters, and special meetings, where votes on mergers typically require 2/3 of all shares outstanding. (My data does not include proxy contests and I exclude special meetings.) More importantly, firms are more likely to send mail ballots when their support from institutional investors is lower. However, individual and institutional votes are nearly uncorrelated, so it is possible that their decisions are uncorrelated with individual voting choices.

Again, I can empirically substantiate the exclusion restriction. If firms were more likely to include mail/phone ballots when they expected strong or weak voting results, then there should be a correlation between their choice and the voting results among unaffected investors—the ones who selected their materials.

Limiting to votes by such shareholders, I regress, at the individual-proposal vote level:

$$r_{ictp} = \beta_0 + \beta_1 m_{ct} + \varepsilon_{ictp}$$

In which r_{ictp} is an indicator for whether shareholder i voted with management on proposal p at firm c at time t.

With this specification, I test whether there is a relationship between the firm's choice of what voting methods to send shareholders, m_{ct}, and the voting choices made by its investors who are unaffected by its choice (those who select a voting method). By limiting to those who make a selection, who therefore receive the same materials regardless of the firm's choice, I isolate any relationship between the firm's decision and retail shareholder favorability towards management at the meeting.

Table 3, Panel A contains this result. There is no evidence of a relationship between voting choices and the firm's mail rate. An increase in the mail rate from 0 to 100% is associated with 0.215 percentage points less voting with management, with a standard error of 0.468.

Table 3, Panel A uses a parsimonious regression, with no covariates or fixed effects. I do this because any sample correlation between the firm's choice of voting methods and individual voting choices would suggest a violation of the exclusion restriction, regardless of whether it is caused by firm strategic behavior or by, say, composition effects from entry / exit of individuals or firms. Still, as an added check, in Table 3, Panel B, I re-estimate the above equation with various fixed effects, to rule out strategic behavior in which funds send mail voting options at meetings where individual investors vote differently, and again find null results.[14]

In summary, I have calculated, for all individual investors, how likely their firms are to send mail/paper ballots. I showed that this statistic, m_{ct}, aggregated to the individual level, is strongly correlated with an individual's likelihood of being a voter. I have shown, using a difference-in-difference, that the individual's receipt of a full package or notice does not directly affect her voting choices. And I have shown, using individual investors who made a selection and are therefore unaffected by the firm's choice, that the firm's choice of materials is uncorrelated with individual investor voting preferences for the meeting.

[14] Is Table 3 evidence that firms do not behave strategically? Hardly. Firms might shift their mail/phone voting options when *institutions* may turn against them, not when *individual investors* will. As I will discuss in Section V, individual voting choices are only weakly correlated with those of funds—instead, the retail shareholder vote tends to be consistent across all firms and years.

E. Excluded Variable Construction

Subsections IV.C and IV.D above appear to demonstrate that the voting methods an investor receives on a given ballot is related to her likelihood of turnout out and not her voting choices. Thus far, I have assumed, for simplicity, that each individual has the opportunity to vote on a single ballot, but in reality individuals may have many ballots to vote on. The choice of voting method affects her probability of voting on a given ballot, but to adjust for sample selection bias on unobservables, I require a variable that drives an individual's choice to *ever* vote without affecting *how* she votes. To obtain a variable that's at the level of the individual investor and not a single voting choice, I calculate an aggregated number for each individual investor based on the voting materials (full package or notice) chosen by all the firms she owns.[15]

I construct this variable in two steps. First, I calculate each firm's mail/phone ballot rate for a specific meeting as (i) the number of retail shareholders who do not select their voting materials *and* who receive mail/telephone ballots, divided by (ii) the number of retail shareholders who do not select their voting materials:

$$m_{ct} = \frac{1}{|\Im_{ct}|} \sum_{j \in \Im_{ct}} m_{jct}$$

Where j indexes individuals, $\Im_{ct}$ is the set of individuals who own firm c at time t and select their materials, and m_{jct} denotes whether individual j at firm c at time t received a mail/phone ballot.[16]

Second, because I am interested in whether individual i *ever* voted, I then aggregate this measure by individual across the firms they own. Specifically, I calculate the average mailing rate by firms at the meetings in which an individual is eligible to vote:

$$M_i = \frac{1}{|\mathbb{C}_i|} \sum_{ct \in \mathbb{C}_i} m_{ct}$$

[15] In future drafts I may explore different ways of constructing this variable to better capture the relationship between the voting methods one receives and one's likelihood of being a voter.

[16] m_{ct} is most commonly 0 or 1, but some firms send some materials to some shareholders and others to other shareholders.

Where $\mathbb{C}_i$ is the set of firm-years ct at which individual i is eligible to vote.[17] I next turn to showing that our newly constructed variable M_i satisfies the relevance condition and the exclusion restriction.

Using the notation above, we can say that Brav et al. (2022) and Table 2, column 1, show that whether shareholder i receives mail/phone voting options, m_{ict}, is causally related to whether she votes in meeting ct, v_{ict}. However, my excluded variable is not the mail/phone ballot for a single meeting m_{ict} but rather the constructed variable M_i, which is m_{ict} aggregated over the individuals in a firm and then over the firms in an individual's account. Similarly, my interest is not in the turnout for a single meeting v_{ict} but rather V_i, individual-level voting. I provide direct evidence that M_i and V_i are connected in Appendix Table A2. In all specifications, M_i is strongly correlated with whether shareholder i ever votes, V_i, with a t-statistic over 150, satisfying the relevance condition.

F. Estimation of Sample Selection Model

I am now ready to assess the extent to which our basic results on individual investor preferences are affected by sample selection bias.

I use a simple Heckman (1979) model, as is most common. Heckman (1979) proposes a maximum likelihood estimation with all variables estimated simultaneously.[18] The results of the first stage estimation, the selection equation, are shown in Appendix Table A2. In Appendix Figure B1, I show a binned scatterplot of the actual likelihood of V on the predicted likelihood of V, confirming that estimating $\Pr(v = 1)$ using a probit model generates a good (unbiased) approximation of the true likelihood of voting.[19]

In Table 4, I present the main results on selection. In each column, I show our predicted average R_i at the bottom of the table in bold, using three different methods to generate it.

[17] Note that, to avoid M_i having a direct relationship with an individual's choices, I calculate M_i for each individual regardless of whether that individual selects her voting materials. For those individuals who select their voting materials, m_{ict} and therefore M_i have no effect on them whatsoever—a fact that we exploit below to test the exclusion restriction.

[18] Heckman (1979) also proposes, as an alternative, a two-step method—first estimating selection with a probit model, using the results to calculate the Mills Ratio for each observation, then including the Mills Ratio as a covariate in a regression estimating equation 3. I do the full maximum likelihood estimation rather than the Heckman two-step method, as is preferred when the computational power is available (Puhani (2002)).

[19] I use the Heckman (1979) sample selection correction rather than the semi-parametric approach proposed by Newey (2009) because that approach does not separately identify the constant—fine when one is interested in identifying regression coefficients, but not when one is seeking to estimate the constant itself.

In Column 1, I run the benchmark regression, a simple OLS regression of R_i with no right-hand side variables, so the regression intercept represents the predicted average R_i.

In Columns 2 through 4, I adjust for selection on observables using Inverse Probability Weighting. I conduct this analysis in two parts. First, I run a probit regression of selection V_i to estimate Equation 3 on the full sample, adding more right-hand-side variables as we move from Column 2 to Column 4. Results of that probit regression are in Appendix Table A2. Second, I calculate regression weights as $1/\widehat{\Pr(V_i)}$ and run a weighted OLS regression of voting R_i to estimate Equation 1 in the subset of shareholders who voted. The goal is to estimate the average R; we have little interest in the coefficients. To produce a simple average estimate, I emulate Column 1 by running a weighted regression of R_i with no right-hand-side variables. This method produces our estimate of the predicted average voting choices adjusted for selection on observables, which we include in the bottom row of the table.[20]

In Columns 5 through 8, I use Heckman maximum likelihood estimation to adjust for selection on observables and unobservables. Each column repeats the right-hand-side variable selection from Columns 2 through 4, respectively. I jointly estimate the selection equation, Equation 3, and the outcome equation, Equation 1, with no weights. I include the excluded variable M_i, the account's mail rate, as an additional variable on the right-hand side of the selection equation and include the results of the selection equation in Appendix Table A2. Following Zimran (2019), I use the regression results to adjust R_i for selection for each observation in the voting subset, by calculating $\widehat{R_i} = R_i - \Omega(\mu + x_i\gamma + M_i\delta)$. As in Columns 2 through 4, I then calculate the average $\widehat{R_i}$ by regressing it with no right-hand-side variables, weighting the observations by $1/\widehat{\Pr(V_i)}$, and putting the result in the bottom row of the table.

Across columns, support for management ranges from roughly 86% to 88%. We can see that neither selection on observables nor selection on unobservables make a large difference in the totals. Voters and non-voters appear to have substantially similar preferences.

[20] All covariates are demeaned over the entire sample of individual investors, not just voters. I note that the constant term differs slightly from the predicted average R_i. This difference is expected and is driven by the fact that the regressions demean over the full sample, not the subset that voted, and therefore the intercept term does not precisely reflect the predicted average R_i.

A final point on using this method for correcting sample selection bias. In a causal inference framework, an instrumental variable yields the impact of treatment on the subset of respondents who comply with the instrument. Similarly, in this setting, the estimates of R_i are best interpreted as the ideology adjusting for sample selection on unobservables on the subset of individuals who comply with the excluded variable (that is, are shifted from not voting to voting).

I use this adjustment in conjunction with other methods to ascertain sample selection bias. Specifically, later in the paper, I show my results on representation using individual investors' zip code ideology, which is not subject to selection on voting, rather than their actual votes.

In future drafts, I intend to expand this analysis in several ways. First, this method allows me to adjust for sample selection bias for each individual observation. Thus, I can conduct bias-adjusted regression analyses and create bias-adjusted figures and tables. Second, the rest of this paper uses SRI voting, whereas this section uses voting on all proposals. I will revise this section so it uses SRI voting instead. For now, I use these results as evidence for the proposition that sample selection bias is not driving large effects in aggregate preferences.

V. Individual and Fund Ideology on SRI

In this section, I explore individual investor and fund voting patterns. Throughout the rest of the paper, I limit to SRI proposals, as the category of proposals in which there are the most substantial arguments for fund representativeness. The results for governance and management proposals are largely identical, and I intend to add them in future drafts.

A. Individual and Fund Ideological Variation

As a preliminary question, we may ask what is more important in investor voting: variation across proposals or variation across shareholders? Given that individuals and funds demonstrate similar aggregate support for SRI proposals, their voting choices have similar overall variance, but such variance could be driven by variation across proposals or variation across shareholders.

The question has implications for how we think about the representativeness of mutual funds. Consider two hypothetical versions of a polity and its views on environmental proposals. In the first, some environmental proposals are extremely popular and others are extremely unpopular; individuals tend to agree with each other on each proposal, sometimes voting in favor and sometimes against based on the quality of the proposal, and there is little polarization. In this polity, a representative legislature would support the popular proposals and oppose the unpopular ones. In the second hypothetical polity, most proposals are middlingly popular, due to some consistent strong supporters and consistent strong opposers, and there is strong polarization. In this polity, a representative legislature would be divided, reflected the divided population.

To understand the sources of variation in voting, I regress voting choices on proposal and investor fixed effects:

$$Y_{ipct} = \delta_0 + (\delta_1 X_{pct}) + (\phi_p) + (\psi_i) + \phi_t + \varepsilon_{ipct}$$

i indexes investors, p indexes proposals, c indexes firms, t indexes years, Y_{ipct} is i's vote on proposal p, X_{pct} is a vector of firm-month and proposal-level covariates, and ϕ_p, ψ_i, and ϕ_t are proposal, investor, and year fixed effects, respectively.

Table 5 contains results estimating the equation above separately for individuals and funds. Columns 1 and 2 include covariates rather than proposal or investor fixed effects. We see in column 2 that fund voting on SRI proposals is highly correlated with ISS and Glass Lewis recommendations and uncorrelated with firm performance. Individual investor voting in column 1 shows a much smaller (though still substantial) correlation with ISS recommendations and none with other covariates.

In columns 3 and 4, I include proposal fixed effects; in columns 5 and 6, I include investor fixed effects. Returning to the question posed above, variation across proposals explains a large portion of variation in fund voting: an R^2 of 22%, in column 4. However, it explains only a small portion of variation in individual voting: an R^2 of only 3% for individual investors in column 3. Note also that the 19% R^2 for funds in column 2, which is very close to the 22% R^2 in column 4, further implies that variation across proposals in fund voting is mostly determined by proxy advisor recommendations.

Investor identity determines a much greater portion of variation in voting than does proposal identity, and this difference is even more pronounced for individuals than for funds: in columns 5 and 6, with account fixed effects, the R^2 is 65% for individuals and 42% for funds.

Table 5 establishes an important empirical fact: in shareholder voting, especially individual voting, firm and proposal characteristics are of second order importance as compared to voter tendencies.

B. Individual and Fund Ideological Consistency

As implied by Table 5, individual voting on SRI proposals is marked by consistency: the identity of the investor is strongly predictive of how she will vote.

In Figure 3, Panel A, I plot the distribution of individual and fund ideologies by calculating, for shareholders with at least 10 votes, their fraction of SRI proposals voted favor, Y_i. A large portion of both individuals and funds almost always vote against SRI proposals, but individual investors have more mass near the tails: they are more consistent, or, put another way, more extreme.[21] An individual shareholder is more likely than a fund to vote the same way on all or almost all SRI proposals.[22]

Individuals exhibit consistency in voting that extends across firms and years. Following Matvos and Ostrovsky (2010), I measure the extent to which an individual's votes in a category are predicted by its vote at *other* firms the previous year:

$$Y_{ipct} = \beta_0 + \beta_1 Y_{i(-c)(t-1)} + \phi_t + \varepsilon_{ipct}$$

i indexes investors, c indexes firms (and $-c$ denotes firms other than c), and t indexes time. The left hand side is the investor's vote on a particular proposal, and the right hand side is her vote on a randomly drawn proposal from a different firm in the previous year.[23] ϕ_t represents time fixed effects. I do the same for funds.

Table 6 shows that an individual vote on one proposal in year $t-1$ is highly predictive of her votes on different firms in year t. An individual's vote in favor of a single randomly drawn SRI proposal increases her likelihood of voting in favor of an SRI proposal at a different company the next year by 47.7 percentage points, more than 10 percentage points more than funds.

[21] I note that by limiting to those with ten or more SRI votes, I subset to a group of individuals who vote more often and/or own more securities, both of which are associated with voting against SRI. I discuss sample selection bias and population measures of support in Section VI.

[22] This result is not driven by funds casting more ballots than individuals. I note again that only individuals with ten or more SRI votes are included. When comparing the histograms, individuals have excess mass at the edges and funds have excess mass in the 40% – 60% range.

[23] I modify Matvos and Ostrovsky to use a single (randomly selected) vote from last year on the right hand side, rather than the average of all the shareholder's votes, to avoid having greater precision in the right-hand side variable for those voters who cast more ballots last year. Using the average of all the shareholder's votes last year yields virtually identical results.

Individual consistency also extends across proposals. I aggregate the voting outcome on each proposal separately for individuals and funds and display the distribution of results in Figure 3, Panel B. SRI proposal voting features substantial heterogeneity in results among mutual funds, with many proposals strongly disfavored and many others strongly favored. In the individual investor vote, by contrast, but exhibits strikingly little heterogeneity among individuals.

We see, then, that with respect to SRI proposals, individual investors have distinctive ideologies on a uni-dimensional yes-or-no spectrum. This is in contrast to mutual funds, who are far more likely to vote for one SRI proposal and against another.

Individual sorting into funds compresses individuals' aggregate votes. Figure 3, Panel C displays a histogram of the average individual vote within funds, overlaid over the histogram of fund voting. Fund aggregates are far closer to each other than the funds themselves are, suggesting that individuals do not strongly sort across funds based on ideology.

VI. Representation in Voting

A. Hypothesis Development

How should we think about the role of funds as intermediaries in voting? I borrow from political science the notion of "representativeness" of elected representatives. An investor's choice of mutual fund shares many elements with electing a representative: the intermediary stands in for a large number of constituents; it casts ballots on various issues, some of which are more ideological in nature while others require expertise; the constituents have opportunities to change intermediaries based on past and expected performance; an intermediary may make its voting choices to maximize its appeal or based on other factors such as social welfare or private benefits; even an intermediary solely focused on its constituents' welfare may not vote exactly as its constituents would vote—for example, the intermediary may have superior information.

Of course, most investors likely view the primary role of the intermediary as choosing a portfolio rather than voting. But this is a difference of degree, not kind—voters choose representatives based on factors other than the representative's voting choices as well.

A vast literature in political science is devoted to empirically estimating the representativeness of political institutions. Representation captures the relationship between constituent opinion and variation in voting by intermediaries. A well-established finding in the

political science literature is the high degree of representation in U.S. politics (see Wlezien (2004) for a list of such findings and Shapiro (2011) for a survey).

I represent an investor's ideology as her fraction of SRI proposals voted in favor:

$$\bar{R}_i \equiv \frac{1}{|\mathcal{H}_i|} \sum_{p \in \mathcal{H}_i} R_{ip}$$

$R_{ip} \in \{0,1\}$ is i's vote on proposal p, $\mathcal{H}_{\mathrm{I}}$ is the set of SRI proposals that i votes on, and $|\mathcal{H}_i|$ is the number of proposals that i votes on.

Consider the following empirical model of the relationship between individual and fund voting:

$$\bar{R}_f = \beta_0 + \beta_1 \sum_{a \in \Theta_f} \bar{R}_a + \nu_f \qquad (6)$$

In which f indexes funds, a indexes individual accounts, Θ_f is the set of accounts at fund f, and $\bar{R}_a$ and $\bar{R}_f$ represent the individual's and fund's average vote on SRI proposals, respectively. As described in Section V, the probability in favor serves as a simple metric of the investor's orientation towards social value maximization versus shareholder value maximization, as in Bolton et al. (2020).

Equation 6 is the standard representativeness equation.[24] If there is no sorting into funds on preferences, then $\beta_1 = 0$. $\beta_1 \in (0,1)$ would imply that funds reflect—but dampen—individual ideology; $\beta_1 > 1$ would imply that funds amplify aggregate individual ideology. All values are *ex ante* plausible. For example, $\beta_1 > 1$ could occur if a fraction of individuals choose funds based on the fund's (exogeneous) voting choices. Low or negative β_1 would imply that funds do not serve as representative intermediaries for their customers.

B. Main Results

I begin in Figure 4 by plotting a scatterplot, at the fund level, of fund ideology (on the y-axis) and the mean ideology of the fund's underlying investors (on the x-axis). I weight funds and individuals equally.

Figure 4, Panel A, which separately demarcates ESG funds and non-ESG funds, reveals the basic picture of representativeness in shareholder voting. Although funds have a wide range of positions towards ESG proposals, with a thick band ranging from 0% to 70% in favor, their

[24] An alternative version of Equation 6, tailored for finance, might weight accounts by their ownership of the fund, rather than equal-weighting all accounts in the fund.

mean individual owner tends to vote similarly, with a narrow band 30% in favor. Even those funds whose underlying investors do not fit into the narrow band—those which appear on the left or right side of the figure—show no apparent correlation in how the funds themselves vote.

An exception is ESG funds, which vote substantially more in favor of SRI proposals and whose shareholders also vote substantially more in favor of SRI proposals. The gap between ESG funds and non-ESG funds provides clear visual evidence of representativeness (that is to say, sorting on ideology), though, if we exclude ESG funds, there is no obvious graphic evidence of representativeness.

Figure 4, Panel A visually makes clear the main result of the paper: variation in fund ideology has no relationship to individual investor ideology, with the exception of ESG funds.

In Figure 4, Panel B, I separately mark index funds and non-index funds. Index funds are substantially more opposed to SRI proposals than non-index funds.

Table 7, column 1 presents results of a regression estimating Equation 6. As in Figure 4, I aggregate each fund's investors and equal-weight both funds and individuals. This follows the standard political science regression, which is at the representative or constituency level, not the constituent level. I cluster at the institution-meeting level.[25]

The significantly positive coefficient in Table 7, column 1 suggests that, on the whole, there is substantial representation in voting by mutual funds. For every 10 percentage point increase in the SRI vote of the fund's underlying shareholders, the fund itself votes 8.95 percentage points more in favor of SRI. However, as I will show, this result relies on ESG funds and is sensitive to the specification.

C. Role of ESG Funds

Figure 4 implies that ESG funds are a factor driving representation. In this subsection, I explore more deeply the factors driving or undermining representation.

Basic representation does not necessarily incorporate causality. For example, we may care whether U.S. representatives' positions on foreign policy correlate with those of their constituents, even if the constituents do not know or care about their representatives' position

[25] Multiple funds at an institution—specifically, those managed by the same sub-advisor—tend to have their meetings together. Iliev and Lowry (2015), Liu, Low, Masulis, and Zhang (2020), and Butler and Gurun (2012) cluster at the fund or firm level, whereas Morgan et al. (2011) and Bubb and Catan (2020) cluster at the fund sub-adviser level. Following Abadie, Athey, Imbens, and Wooldridge (2017), since a fund is only included in my sample when it has a shareholder meeting, and since groups of funds have shareholder meetings together, these groups of funds are the appropriate level for clustering. The fund-meeting level is similar to the sub-adviser level.

and the representative does not know or care about the positions of her constituents. Any correlation in this dimension may be driven by other factors—whether policy-related, such as the political parties of the relevant actors, or not, such as the tone that the politician strikes in public communications. Controlling for political party or politician tone in a regression of representative ideology on constituent ideology would reduce interpretational clarity.

Rather than *controlling* for fund characteristics, I instead seek to assess how much of representativeness is *explained by* fund characteristics. A natural question is how much of the representativeness shown in Table 7, column 1, is driven by fund features such as ESG or index funds. Merely adding such explanatory covariates to the right-hand side of the equation would reduce the interpretational clarity; instead, we are interested in how these variables affect β_1, the coefficient on individual investor ideology.

Gelbach (2016) explores this class of regressions, using as an example the black-white pay gap from a regression of wages on a race indicator. The baseline regression of wages on race without covariates is substantively interesting; explanatory covariates (in Gelbach (2016)'s case, test scores and educational achievement) are interesting in how they change the coefficient on the race indicator. Gelbach (2016) develops a tool to measure the impact of additional covariates on the coefficient of interest (as well as the statistical significance of the impact), independent of the order the coefficients are added.

Consider:

$$\bar{R}_f = \gamma_0 + \gamma_1 \sum_{a \in \Theta_f} \bar{R}_a + \gamma_2 X_f + \eta_f \qquad (7)$$

In which X_f is a vector of categorical fund characteristics, such as fund objective or whether it is an index or ESG fund. γ_1 represents the within-category representativeness. A high γ_1 but low β_1 suggests that the overall system is not representative but that, once investors choose funds within a category, there is within-category representativeness. A significant gap between γ_1 and β_1 indicates that representativeness is being driven (or undermined) by selection into the category.

In Table 7, Columns 2 and 3, I estimate Equation 7, including ESG and index fund dummies as covariates. Below each column, following Gelbach (2016), I include the impact of the covariates on the coefficient on mean owner ideology (from the base regression in Panel A), as well as the standard error of the impact.

As anticipated by Figure 4, Panel A, ESG funds are the major driver of correlation between funds and investors. Including an ESG indicator decreases the coefficient by 0.888 percentage points, which is roughly the coefficient in column 1. That is, controlling for ESG funds eliminates any relationship between fund and individual ideology.[26]

ESG funds are small. Table 7, columns 1 through 4 weight each fund equally. In columns 5–8, I weight each fund by its number of account-holders so each individual investor is weighted equally. The coefficient on β_1 becomes insignificant in column 5, due to the greater weight placed on index funds (which are large) and the reduced weight placed on ESG funds (which are small). The result in column 5 implies that, as between two fund holdings by individuals, one of whom is more pro-SRI than the other, there is no statistically significant prediction as to whose fund is more pro-SRI. This is in sharp contrast to the political science literature, for which a correlation between constituent and representative ideology is a core established empirical fact.

In columns 6–8, I mirror columns 2–4 by adding in ESG funds and index funds, but this time weight observations by account. Column 7 shows that, with marginal significance but a high coefficient, index funds negatively drive overall representation. Index funds are substantially less pro-SRI than other funds, whereas their shareholder bases are significantly more pro-SRI than those of other funds (see Online Appendix Table A10). Due to the sheer size of index funds, the effect is negligible in column 3 and large in column 7.

D. Account Characteristics

In the previous subsection, I observe that index funds reduce the degree of overall representation with marginal statistical significance. Here, I show how this result may derive from the characteristics of pro-SRI individuals and individuals at different types of funds. Specifically, in Online Appendix Tables A3 and A4, I use account information and zip code characteristics to yield insights on demographic features of funds and individuals.

In Online Appendix Table A3, I regress fund characteristics on account and individual characteristics, first aggregating to the fund level, then with observations at the account-fund level (which places relatively greater weight on funds with more accounts). I demean all right-hand-side variables so that intercepts can be interpreted as the mean of the dependent variable.

[26] In column 4, I add fixed effects for fund objective, and find no change. For computational tractability, in column 4 and column 8, when using high-dimensional fixed effects, I do not calculate the impact on β_1 of adding additional variables (or the standard error of the impact).

Notably, whereas 1.5% of funds are ESG funds (Panel A), only 0.6% of account-fund observations are ESG funds (Panel B), and whereas 16.1% of funds are index funds, 24.5% of account-fund observations are index funds, indicating that ESG funds tend to have few accounts and index funds tend to have many. Online Appendix Table A3 makes clear that account zip-code level demographic characteristics are strongly correlated with fund choices. Variation in fund-level aggregate zip code demographic characteristics and account portfolio characteristics can explain a substantial fraction of variation in fund characteristics, including 40% of cross-fund variation in a fund's closed-end status.

In Online Appendix Table A4, I regress account SRI percent on zip code and account characteristics, again demeaning all right-hand-side variables so that intercepts can be interpreted as the mean of the dependent variable. Pro-SRI investors are poorer and live in denser, younger, more educated, more diverse, lower-income zip codes. Index fund owners are wealthier and hold smaller stakes in each fund, and also live in younger, denser, more diverse, more educated zip codes. These results give some indication why index funds reduce representation—index funds, which vote strongly against SRI, draw from pools of individuals who tend to be more pro-SRI.

To what extent can the above results be interpreted as describing individual investors, as opposed to merely where they live? Caution is warranted when interpreting group characteristics as proxies for individual micro-characteristics, since mutual fund holders are not randomly selected from each zip code but are rather a self-selecting group.[27] To the extent that the residual with respect to the (unobserved) individual characteristic is correlated with the zip-code level characteristic, the estimated coefficient would be biased. (For example, if older individuals in older zip codes and older individuals in younger zip codes have different propensities to buy index funds, then we could not accurately interpret the coefficients in Online Appendix Table A3 as the relationship between index fund ownership and investor age.) As an alternative, for investors who own "Target Date Retirement" mutual funds, I use the target date of their retirement as a micro-level proxy for age. Online Appendix Table A5, Panel A contains a validation check—the imputed age is strongly correlated with the zip code age. Online Appendix Table A5, Panel B shows that imputed age of the individual investor is closely related to fund

[27] Using a group characteristic also tends to downwardly bias the coefficient estimate even if mutual fund holders were a randomly selected group from each zip code. This should not affect the qualitative interpretation of these results, though it may cause Online Appendix Tables A3 and A4 to understate magnitudes.

choice and SRI preferences: a fund held by an investor ten years older is 2.5 percentage points less likely to be an index fund. An investor ten years older is 4.2 percentage points less supportive of SRI proposals on average (though estimated on the small group of accounts that vote on SRI proposals on equity holdings and own target retirement funds). These numbers are highly significant and fully consistent with those obtained using group characteristics above.

E. Individual Choice Model and Estimation

Equation 7 describes a potential empirical relationship but is not intended to describe the decision-making of any particular actor. As an alternative to Equation (1), I could model account a's decision whether to own fund f as follows:

$$Q_{af} = \kappa_0 + \kappa_1\left(\bar{R}_f - \bar{R}_a\right)^2 + \kappa_2 X_{af} + \phi_f + \theta_a + \eta_{af} \qquad (8)$$

Q_{af} equals 1 (multiplied by 100) if account a owns fund f, $\bar{R}_f$ and $\bar{R}_a$ are fund and account ideology, X_{af} is a vector of additional variables concerning the match between a and f, ϕ_f is fund fixed effects and θ_a is account fixed effects. Fund fixed effects ensure that Equation (3) does not capture the factors that drive total fund flow from individual investors but rather the factors that drive a match between a particular investor and fund. I include in X_{af} the distance between individual and fund headquarters to add to the literature on home bias whether individuals invest in mutual funds that are headquartered closer to home. Equation 8 uses a standard quadratic loss function, as in Ansolabehere (2008).

Such a model takes fund ideology as exogeneous but allows individual investors to choose funds based on the fund's votes. This assumption is plausible. For one, mutual funds cannot observe their beneficial owners, though they may use market research to ascertain preferences. Even if funds vote to attract customers, they may set their votes based on the ideologies of potential customers or institutional customers, not existing individual customers.

Estimating Equation 8 with an observation for every possible account-fund combination would be computationally intractable, generating a vast sparse matrix that is too large to estimate (Bertsimas and Cory-Wright 2020). Instead, in Table 8, I estimate Equation 8 by selecting at random, for each account, twenty control group funds that the account does not own and adjusting the estimation for the sampling weights. The estimation results appear in Table 8. Observations are at the account-fund level. As in Table 7, column 5, there is no significant overall correlation between individual and fund ideology when not weighting funds equally.

Table 8, Column 2 shows that controlling for ESG funds reduces the representation correlation, implying that ESG funds increase representation. It also shows that an account that votes 100% in favor of SRI proposals is 0.07 percentage points more likely to own a *particular* ESG fund than one that votes 0% in favor—quite large, given that the account has a 0.05 percentage point likelihood of owning the ESG fund unconditionally.

Table 8 also provides clearer evidence of an index fund effect. Column 3 shows that pro-SRI accounts are substantially more likely to own index funds, and index funds, as we have seen elsewhere, are substantially less pro-SRI. Table 8, column 4 shows that just as there is local home bias by fund managers (Coval and Moskowitz (2002)), individuals are also more likely to select mutual funds that are located closer to home.

F. Role of Proxy Advisors

Bubb and Catan (2020) find that fund voting is effectively divided into three clusters: those that vote with management, those that vote with ISS, and a smaller cluster that vote with Glass Lewis. ESG funds, which vote against management more than the three clusters, are a small collection of funds outside of the clusters.

In Appendix Figure B2, I graph the relationship between a fund's distance to its underlying investors and its agreement rate with management, ISS, and Glass Lewis recommendations. High rates of agreement with Glass Lewis are associated with greater similarity to a fund's investors, and funds that closely follow Glass Lewis recommendations tend towards significantly greater ideological similarity to their investors than those that closely follow management or ISS recommendations. By contrast, similarity to a fund's investors is maximized when agreeing with management recommendations on 70 to 80% of SRI proposals.

G. Robustness Checks and Extensions

Sample Selection Bias. In Section IV, I concluded that individual votes on proposals are not, in the aggregate, strongly affected by sample selection bias, though I have yet to apply it to SRI votes or apply the selection bias adjustment to the results on representativeness.

As a different check on whether my results hold for individual investors as a whole rather than just voters, I use a proxy for one's ideology that is available for all mutual fund holders. The political voting results for the investor's zip code—specifically, the two-party share of the vote received by Hillary Clinton in the 2016 Presidential Election—are available for all accounts

except for a small fraction that are missing a zip code or that live in a zip code for which political results are unavailable. As a validation test, in Online Appendix Table A6, I test zip code ideology as a proxy for SRI votes and show that, for the subset of mutual fund holders who vote on equities, there is a large, highly significant correlation between zip code ideology and support for SRI. A 10 percentage point increase in the account's zip code's Democratic two-party share is associated with a 1.6 percentage point increase in the account's percentage in favor of SRI proposals. Online Appendix Figure B3 shows a binned scatterplot of the relationship between zip code political ideology and fraction in favor of SRI proposals.

In Figure 5, I repeat Figure 4, Panel A, using, rather than the fund's mean account SRI vote, the fund's mean account zip code ideology, excluding all accounts with 10 SRI votes (i.e. excluding accounts that appear in the main tests). The results show that the basic picture of representation is the same using this alternative metric: there is ideological sorting into ESG funds, but otherwise no evidence of sorting by ideology, though the latter point must be tempered by the high degree of noise added by this metric.

In Section IV, I correct for selection on both observables and unobservables. Although I have not yet carried forward the correction on unobservables, I have carried forward the correction on observables. Specifically, in the Online Appendix, I use Inverse Probability Weighting to adjust for sample selection bias on observables. The critical assumption in inverse probability weighting is that, for selection variable v, Pr $(v = 1|x_i, z_i)$ is known when $v = 1$, where Θ is a set of observable variables.

I use to predict Pr $(v = 1|x_i, z_i)$ for all observations. Then, I simply weight each observation by $\frac{1}{\Pr\,(\widehat{v=1}|x_i,z_i)}$. In Online Appendix Table A7, I repeat my main specifications but weight each observation by its inverse probability weight. The results are similar to the main results.

Voting on the Intensive Margin. Above and in Section IV, I attempt to adjust for sample selection bias by, effectively, comparing voters and non-voters. An alternative approach that could yield insights would be to compare those who vote frequently to those who vote infrequently. If infrequent voters more closely resemble non-voters, any observed ideological differences between infrequent and frequent voters could suggest sample selection bias.

Here, the results show that infrequent voters are, in fact, more pro-SRI than frequent voters. In Online Appendix Table A8, Panel A, I regress SRI percentage in favor on vote

frequency, conditional on being a voter on SRI proposals at all, and find a statistically significant result. In Online Appendix Table A8, Panel B, as an alternative metric that does not require conditioning on being a voter, I regress the zip code political vote on vote frequency, with similar results—those in Republican zip codes vote more frequently. In future drafts, I intend to explore this question further by using my excluded instrument to explore selection into voting on the intensive margin.

Management and Governance Proposals. It is possible that individuals sort into funds on ideology other than SRI ideology. In Online Appendix Table A9, I provide results for governance and management votes, as well as the account's composite voting rate in line with management recommendations. Unsurprisingly, governance proposals appear to feature weaker (marginally significant) representation than SRI proposals, owing to a smaller (marginally significant) effect from ESG funds, and management proposals have no significant representation and no significant effect of ESG funds on representation. The impact of index funds on representation in governance proposals is at least as strong as for SRI proposals.

Using Points in the Distribution of Fund Owners Other Than The Mean. In Figure 4 and Table 7, my measure of aggregation of individual accounts is the fund's mean account ideology. However, the mean is just one way to represent the aggregate preferences of a fund's individual investors. In Online Appendix Figure B4, I replicate Figure 4, Panel A, replacing the fund's mean account ideology with a series of points in each fund's distribution of account ideologies. The basic picture is identical: strong evidence of sorting into ESG funds, and little evidence of sorting apart from that. Funds with relatively few anti-SRI voters (non-zero 10^{th} and 25^{th} percentiles) are very likely to be ESG funds, and ESG funds have far more pro-SRI median voters. These results show that all funds tend to have a complete range of account ideologies, but non-ESG funds have a large mass of investors who almost always vote against SRI proposals and a small mass of pro-SRI investors, whereas ESG funds have a small mass of anti-SRI investors and a large mass of pro-SRI investors.

Switching Left-Hand and Right-Hand Side Variables. The coefficients in Table 7 must be interpreted in light of Figure 4. Namely, Table 7 features fairly large standard errors, permitting the possibility that the true population coefficient could be large. Fund aggregate account ideologies are tightly bunched on the x-axis (other than ESG funds); therefore, with fund aggregate account ideology on the right-hand-side (as in Table 7, following the standard political

science representation equation), the line of best fit tends towards vertical with large standard errors. In Appendix Table A10, I place the fund account ideology on the left-hand-side with fund ideology on the right-hand-side and show that standard errors and coefficients drop substantially, since fund ideology has extensive variation and fund account ideology has little variation. The interpretation is identical, of course: funds have a wide range of voting ideologies, and there is little sorting by account ideology except to ESG funds.

VII. Limited Attention

What explains the lack of sorting into like-minded mutual funds other than ESG funds? In this section I explore potential explanations. I look at suggestive evidence related to the granularity of information related to an investor's potential acquisition. I also show that the relationship between individual and fund ideology increases significantly with the individual's investment size. This section is necessarily speculative in nature.

One possibility is that individual investors seek fund types for which there is a lack of variation in mutual fund ideology. This does not appear to be the case. Appendix Figure B5 shows a series of histograms of fund ideologies for common fund objectives separately for index and non-index funds and excluding ESG funds. These fund types provide ideological options for an individual seeking a fund specific ideology. In Online Appendix Table A11, I show that there is no relationship between the number of funds with a certain type and the strength of the match between fund and individual ideology. These results suggests that lack of availability is not driving lack of sorting.

Another explanation could be that individuals wish for their funds to vote differently than they do. Such a desire could be due to an information channel: funds receive private signals as to the impact on firm value of a given proposal, and so individuals might expect the informed funds to pursue the individuals' interests by voting differently than they do. Alternatively, individuals may receive a consumption benefit from voting a certain way but do not want their voting choices effectuated; for example, individuals may enjoy voting for SRI proposals but not want to see SRI policies actually imposed on firms. A related possibility is that pro-SRI individuals, even if they prefer pro-SRI policies effected, prefer to own mutual funds that vote against SRI policies as a signal of the fund's commitment to value maximization.

A. Granularity of Information

Another candidate is limited attention, which may cause an ideological shareholder to invest little in acquiring information about the social impact of the firms or funds she invests in. Hirshleifer and Teoh (2003), Peng and Xiong (2006), and Dellavigna and Pollet (2009) discuss the role of limited attention in investor purchasing decisions. Rational inattention may have an especially pronounced impact in a voting context: a shareholder might not invest in information acquisition on a proposal due to the small likelihood of her vote being pivotal. But it also applies to her choice of delegated voting intermediaries: even if the shareholder has complete information on all proposals, if she lacks information about how the intermediary would vote on those proposals, she might not invest in information acquisition on the intermediary's choices due to the small likelihood of her investment causing the intermediary's vote to be pivotal.

Information about fund voting may be costly for individuals to acquire. Individuals may not know that funds disclose their votes and voting procedures, and, to the extent they are aware, may not be willing to expend the costs required to read and understand these materials.

Funds disclose votes on two forms. First, funds disclose all their past votes on Form N-PX. A fund may own hundreds of securities, sorted by firm, each of which may have ten or more proposals each year. Online Appendix Figure B6, Panel A, contains a representative excerpt from the Form N-PX for a mid-sized institution for a single year. The full form totals 423 pages. Funds also are required to disclose their voting policies and procedures on a forward-looking basis. Online Appendix Figure B6, Panel B, contains a representative excerpt of one such page; the policy totals three pages. These policies and procedures tend to be vague, aspirational, and give little insight into how the fund would vote in any particular situation; and, in any event, they may still be costly for an individual to acquire and understand.

I present evidence consistent with the notion that individuals use broader information, not granular information, in making portfolio choices. First, I explore the connection between pro-SRI individual investors and ESG funds. In Table 9, Panel A, I estimate the following equation:

$$\sum_{a \in \Theta_f} \bar{R}_a = \delta_0 + \delta_1 X_f + \varepsilon_f$$

The left hand side contains fund f's mean account ideology and the right hand side contains a vector of fund features related to environmental goals. Columns 1 and 2 show the Morningstar Sustainability Rating, which Hartzmark and Sussman (2019) show drive fund flows.

There is no relationship between the rating and the mean account ideology. I column 3, I include the nature of the firms in the fund portfolios—the value-weighted fraction of holdings that are fossil fuel or renewable energy firms. There is a strong significant relationship between account ideology and ownership of renewable energy firms, but the relationship goes away once an ESG indicator is added in column 4. ESG funds serve to match individuals with portfolios that fit their voting ideology, but, aside from ESG funds, individuals show no matching by such portfolio characteristics.

In column 5, I add an indicator for whether the fund is explicitly a fossil fuels or commodities fund—whether it has "Oil", "Commodities", "Energy", "Natural Resources", "Global Resources", or "Pipeline" in its name (other than ESG funds). Funds with these in the name have marginally significantly lower mean account SRI voting, consistent with individuals making ideological choices based on salient information.

B. Individual Selection of Equities

Next, I perform a similar test for equities.

One could envision that individual investors view ownership and voting as substitutes, and that individuals who vote against SRI proposals may own greener firms. I regress an account's voting percentage on ownership of fossil fuel firms and renewable energy firms:

$$\bar{R}_i = \zeta_0 + \zeta_1 K_c + \varepsilon_{ic}$$

Table 9, Panel B, contains results estimating this equation. I find that individuals who vote in favor of SRI proposals are more likely to own renewable energy firms and less likely to own fossil fuel producers. I conclude that individual ideology with respect to SRI proposals is positively associated with the shareholder's choice of equities. Column 1 provides clear evidence that individuals incorporate ideology into their equity portfolio decisions.[28]

In columns 2–4, I test KLD's SRI scores, as well as other KLD ratings that may be subtler than the general firm business line. Despite the evidence from column 1 that ideological voters seek ideologically aligned equities, I find no relationship—if anything, there is a slightly negative relationship.

Table 9 provides suggestive evidence that individual investors make choices based on the broad categorizations of firms and funds—ESG, renewable energy, fossil fuel—and not on more

[28] In Online Appendix Table A12, I repeat this specification where, rather than using an account's entire voting history, I exclude the account's votes on the firm in question. The results are qualitatively and quantitatively highly similar.

granular information on the fund's holdings or using third party information. Such an explanation would be consistent with evidence, such as Barber, Odean, and Zheng (2005), that individual investors react more to more salient information.[29]

C. Investment Size

To the extent rational inattention plays a role, individuals making larger investments may show a closer relationship between individual and fund attention due to a greater willingness to pay to acquire information (or due to greater financial sophistication). In Table 10, Panel A, I show that the degree of representation between individuals and funds is significantly larger for individuals with larger investments. I estimate Equation 8, modified as follows:

$$Q_{af} = \kappa_0 + \kappa_1\left(\bar{R}_f - \bar{R}_a\right)^2 + \kappa_2\bar{R}_f * LogAverageAmount_a + \kappa_3\left(\bar{R}_f - \bar{R}_a\right)^2 * LogAverageAmount_a + \kappa_4 X_{af} + \phi_f + \theta_a + \eta_{af} \qquad (8B)$$

Where $LogAverageAmount_a$ is the log of the average stake value in account a's securities, and where I exclude all ESG funds.[30] In column 1, I find a significantly positive coefficient κ_3: that is, the relationship between individual and account ideologies is greater for individuals with higher investment amounts, even excluding ESG funds. In columns 2 and 3, I add controls for the account's number of SRI votes cast, to ensure the results are not driven by greater precision in measuring account preferences.

The results of Table 10, Panel A, show that as one's investment size increases, a pro-SRI investor is more likely to pair with a pro-SRI fund, and an anti-SRI investor is more likely to pair with an anti-SRI fund, even excluding ESG funds.

As an alternative specification, for each actual individual fund-pairing, I regress:

$$\bar{R}_f = \beta_0 + \beta_1\bar{R}_a + \beta_2 LogAmount_{af} + \beta_3\bar{R}_a * LogAmount + \varepsilon_{af}$$

In which $LogAmount$ represents the log stake value of account a in fund f. The results are in Online Appendix Table A13; I find a significantly positive coefficient on β_3.

In Table 10, Panel B, I re-estimate Table 7 but weight accounts within a fund by their ownership portion of the fund, not equally, when calculating the fund's mean account ideology. The results are striking: a complete reversal of the null result in Table 7, column 2. Even

[29] It may also be consistent with reports in the popular press that individuals choose ESG funds even if those funds' portfolios are not environmentally friendly (Otani 2019).

[30] I use an account's average investment value across her portfolio holdings, rather than the value of any particular holding, because this regression includes both actual account-fund pairings and control group account-fund pairings in which the account has no investment in the fund.

controlling for ESG funds, funds have a strong significant correlation with their mean account ideology when weighting those accounts by the account's share of ownership of the fund.

Individuals with larger investments have a closer ideological fit between their own votes and the votes by their mutual funds. This relationship may be because such individuals are more financially sophisticated or more willing or able to expend effort to learn about their investment. Another explanation might be funds take into account the ideologies of their wealthier investors in determining how to vote. Regardless of the explanation, the picture of representativeness looks markedly different for large investors.

VIII. Conclusion

In this paper, I use novel microdata on individual ownership of mutual funds and ownership and voting of equities to evaluate the relationship between fund and individual voting. I begin by evaluating sample selection bias. Using an excluded variable—access to voting methods—that is related to voting turnout but not vote choice, I find (preliminarily) that, in the aggregate, sample selection bias does not appear to make a huge difference in individual investor preferences. Next, I show that individual investors are consistent in their preferences on SRI: they are more ideologically extreme than funds and they vote consistently across proposals. I show that individual voting and mutual fund voting have a high correlation. This correlation is driven entirely by ESG funds, which have an ideologically distinct shareholder base. Despite the variety in preferences among mutual funds, each mutual fund (excluding ESG funds) has ideologically similar owners, and variation in mutual fund voting (excluding ESG funds) has no relationship to variation in individual voting.

ESG funds serve a unique role in the mutual fund ecosystem, as they produce ideological sorting by ordinary investors. To the extent that non-ESG funds have diverse ideology, it appears unrelated to the preferences of ordinary individual investors.

These results may be driven by limited attention. Individuals select funds and firms ideologically by type, but do not appear to incorporate subtler information. Wealthier shareholders are likelier to hold funds that match their ideological preferences, even excluding ESG funds.

The canonical model of voting participation would suggest that, due to their negligible probability of being pivotal, individuals should not turn out to vote, and, to the extent they do

turn out, should not invest resources in information acquisition regarding their vote. And yet, in corporate shareholder meetings as in political elections, some individuals do vote, and when they vote, they vote in a coherent and ideological manner. That coherent ideology extends to mutual fund selection when individuals select ESG funds. However, most variation in voting across mutual funds has no reflection in the ideology of the owners of those funds.

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey M. Wooldridge, 2017, When Should You Adjust Standard Errors for Clustering?, NBER working paper.

Aggarwal, Reena, Isil Erel, and Laura T. Starks, 2015, Influence of Public Opinion on Investor Voting and Proxy Advisors, Fisher College of Business working paper.

Ansolabehere, Stephen, 2008, Voters, Candidates, and Parties, *The Oxford Handbook of Political Economy*.

Barber, Brad, Terrance Odean, and Lu Zheng, 2005, Out of Sight, Out of Mind: The Effect of Expenses on Mutual Fund Flows, *Journal of Business* 78(6): 2095-2120.

Bebchuk, Lucian, Alma Cohen, and Scott Hirst, 2017, The Agency Problems of Institutional Investors, *Journal of Economic Perspectives* 31(3), 89-102.

Bertsimas, Deimitris and Ryan Cory-Wright, 2020, A Scalable Algorithm for Sparse Portfolio Selection, Working Paper.

Bolton, Patrick, Tao Li, Enrichetta Ravina, and Howard Rosenthal, 2020, Investor Ideology, *Journal of Financial Economics* 137(2), 320–352.

Brav, Alon, Matthew Cain, and Jonathon Zytnick, 2020, Retail Shareholder Participation in the Proxy Process: Monitoring, Engagement, and Voting, ECGI Finance Working Paper No. 637/2019.

Broccardo, Eleanora, Oliver Hart, and Luigi Zingales (2020), Exit v. Voice, NBER Working Paper 27710.

Bubb, Ryan and Emiliano Catan, 2020, The Party Structure of Mutual Funds, working paper NYU Law School.

Butler, Alexander W. and Umit G. Gurun, 2012, Educational Networks, Mutual Fund Voting Patterns, and CEO Compensation, *The Review of Financial Studies* 25(8), 2533–2562.

Coval, Joshua, Tobias Moskowitz, 2002, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *Journal of Finance* 54(6), 2045-2072.

Dellavigna, Stefano and Joshua Pollet, 2009, Investor Inattention and Friday Earnings Announcements, *Journal of Finance* 64(2), 709-749

Downs, Anthony (1957), An Economic Theory of Democracy. New York: Harper Collins.

Dyck, Alexander, Karl Lins, Lukas Roth, and Hannes Wagern, 2019, Do Institutional Ivnestors Drive Corporate Social Responsibility? International Evidence, *Journal of Financial Economics* 131(3): 693-714.

Fos, Vyacheslav and Margarita Tsoutsoura, 2014, Shareholder democracy in play: Career consequences of proxy contests, *Journal of Financial Economics* 114(2), 316-340.

Gantchev, Nickolay and Mariassunta Giannetti, 2018, The Costs and Benefits of Shareholder Democracy, CEPR Discussion Paper No. DP13358.

Gelbach, Jonah B., 2016, When Do Covariates Matter? And Which Ones, and How Much?, *Journal of Labor Economics*, 34(2), 509 – 543.

Gilson, Ronald and Jeffrey Gordon, 2013, The Agency Costs of Agency Capitalism: Activist Investors and the Revaluation of Governance Rights, *Columbia Law Review* 113(4), 863-927.

Hale, Jon, 2018, Sustainable Funds U.S. Landscape Report, Morningstar Research.

Hart, Oliver and Luigi Zingales, 2017, Companies Should Maximize Shareholder Welfare Not Market Value, ECGI-Finance Working Paper, 2017.

Hartzmark, Samuel and Abigail Sussman, 2019, Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows, *Journal of Finance* 74(6), 2789-2837.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337– 386.

Hong, Harrison, and Leonard Kostovetsky, 2012, Red and Blue Investing: Values and Finance, *Journal of Financial Economics* 103(1): 1-19.

Iliev, Peter and Michelle Lowry, 2015, Are Mutual Funds Active Voters? *The Review of Financial Studies* 28(2), 446–485.

Kahan, Marcel, and Edward B. Rock, 2008, The Hanging Chads of Corporate Voting, Georgetown Law Journal, 96.

Kakhbod, Ali, Uliana Koginova, Andrey Malenjo, Nadya Malenko, 2020, Advising the Management, Working Paper.

Lee, Allison, 2021, Every Vote Counts: The Importance of Fund Voting and Disclosure, Speech dated March 17, 2021.

Levine, Matt, 2020, The Government Wants ESG Out of Pensions, *Bloomberg Opinion*.

Levit, Doron, Nadya Malenko, and Ernst Maug, Trading and Shareholder Democracy, ECGI Finance Working Paper No. 631/2019.

Liu, Claire, Angie Low, Ronald W. Masulis, and Le Zhang, 2020, Monitoring the Monitor: Distracted Institutional Investors and Board Governance, *The Review of Financial Studies*, 33(10) 4489–4531.

Malenko, Nadya and Yao Shen, 2016, The Role of Proxy Advisory Firms: Evidence from a Regression-Discontinuity Design, *Review of Financial Studies* 29(12): 3394-3427.

Matvos, Gregor and Michael Ostrovsky, 2010, Heterogeneity and Peer Effects in Mutual Fund Proxy Voting, *Journal of Financial Economics* 98(1), 90-112.

Morgan, Angela, Annette Poulsen, Jack Wolf, and Tina Yang, 2011, Mutual Funds as Monitors: Evidence from Mutual Fund Voting, *Journal of Corporate Finance* 17(4), 914-928.

Otani, Akane, 2020, ESG Funds Enjoy Record Inflows, Still Back Big Oil and Gas, Wall Street Journal, available at https://www.wsj.com/articles/top-esg-funds-are-all-still-invested-in-oil-and-gas-companies-11573468200

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, Journal of Financial Economics 80, 563– 602.

Poole, Keith T. and Howard Rosenthal, 1985, A Spatial Model for Legislative Roll Call Analysis, *American Journal of Political Science*, 29(2), 357-384.

Riedl, Arno and Paul Smeets, 2017, Who Do Investors Hold Socially Responsible Mutual Funds? *Journal of Finance* 72(6): 2505-2550.

Securities and Exchange Commission, 2003, Final Rule: Disclosure of Proxy Voting Policies and Proxy Voting Records by Registered Management Investment Companies.

Securities and Exchange Commission, 2021, SEC Division of Examinations Announces 2021 Examination Priorities.

Shapiro, Robert Y., 2011, Public Opinion and American Democracy, *The Public Opinion Quarterly* 75(5), 982-1017.

Sorkin, Andrew, 2019, How Shareholder Democracy Failed the People, *The New York Times*.

Urbinati, Nadia and Mark Warren, 2008, The Concept of Representation in Contemporary Democratic Theory, Annual Review of Political Science 11: 397-412.

Wahal, Sunil and Albert Wang, 2013, Competition Among Mutual Funds, *Journal of Financial Economics* 99(1), 40-59.

Wlezien, Christopher, 2004, Patterns of Representation: Dynamics of Public Preferences and Policy, *Journal of Politics* 66, 1-24.

Figures and Tables

Figure 1. Number of Fund Owners

This figure plots, for each of the 4,965 funds in the dataset, the number of individual investors who own the fund, on a logarithmic scale. The figure also displays the number of owners of the fund who also own at least one equity security and the number of owners of the fund who vote on at least one equity security. Funds are sorted by number of investors, then number of direct equity-holders, then by number of equity voters. The y-axis contains the number of investors; the x-axis represents the fund's ordinal ranking with respect to number of investors.

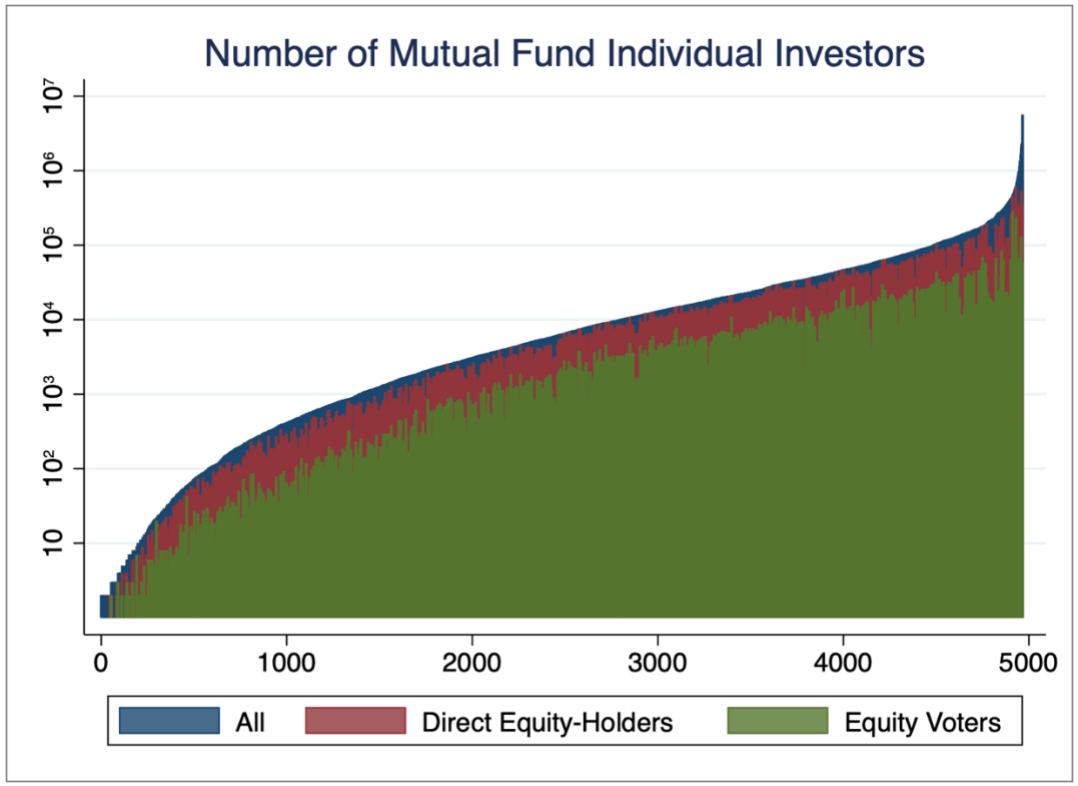


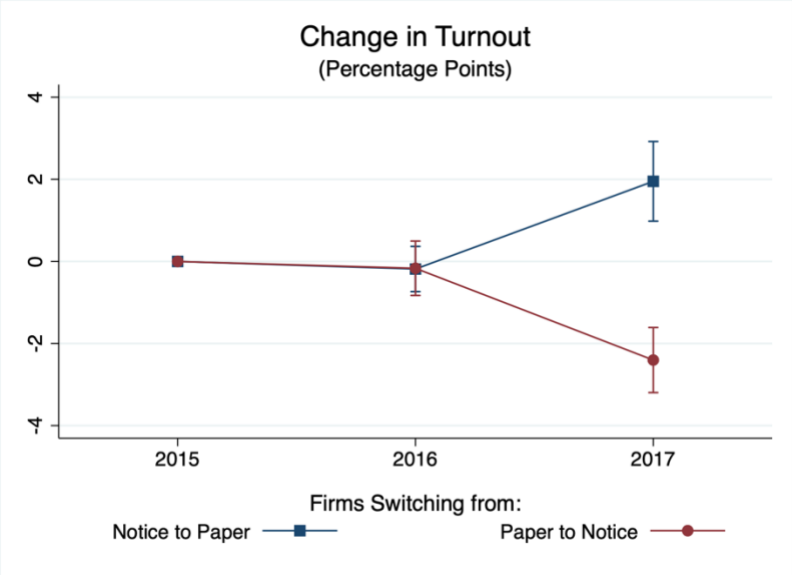
Figure 2. Effects of a Change in Materials Delivered

In this Figure, I measure the effects of the firm switching an individual's voting materials. In Panel A, the outcome variable is v_{ict}, whether individual i voted at firm c's meeting at time t. In Panel B, the outcome variable is r_{ict}, the fraction of proposals voted in favor of management. I limit to annual meetings and remove firms that send a mix of materials at any of their meetings in the sample period. I identify firms that switch from notice to full package or vice versa; I keep switchers and non-switchers but remove firms that switch more than once. For column 1, to make the sample size computationally tractable, I randomly sample 3,000 shareholders for each firm for non-switching firms that have more than 3,000, and weight observations by the inverse sampling probability. For Panel B, I limit to investors that vote in all of the firm's meetings during the sample period. I limit to firms that either do not switch or that switch in 2017, to allow for parallel trends. I estimate:

$$y_{ict} = \sum_{\tau=2015}^{2017} \beta_\tau \, 1_{t=\tau} NoSelection_{ic0} switcher_c \; + \phi_{ic} + \phi_{ct}$$

separately for firms that send full package and notice in 2015, making two regressions in total. $switcher_c$ is an indicator for whether firm c switches the materials it sends during the sample period; $post_{ct}$ is an indicator for whether the observation in question is post-switch; and $NoSelection_{ic0}$, indicates whether shareholder i did not select for firm c at time 0. The specification includes individual-firm and meeting fixed effects. Each coefficient represents, for non-selecters as compared to selecters, as compared to time $t = 0$, the difference between switching firms and non-switching firms. 95% confidence intervals are based on standard errors clustered at the account and meeting level.

Panel A



Panel B

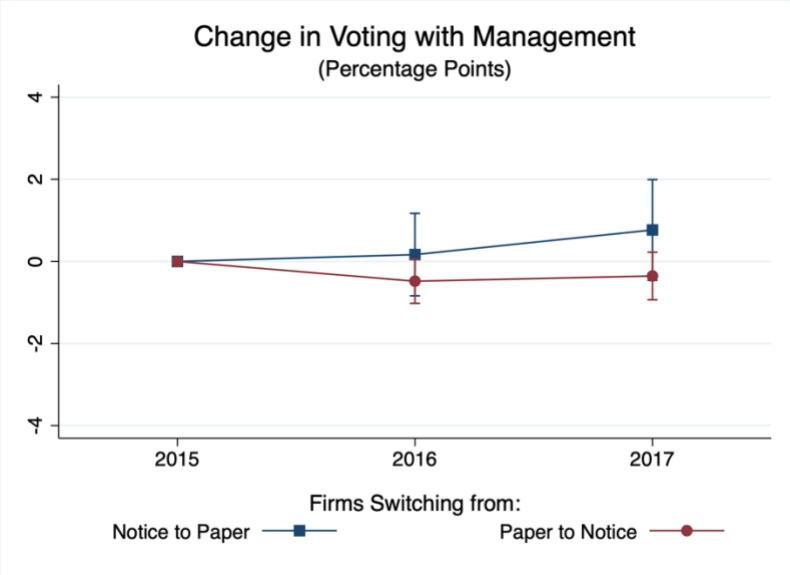
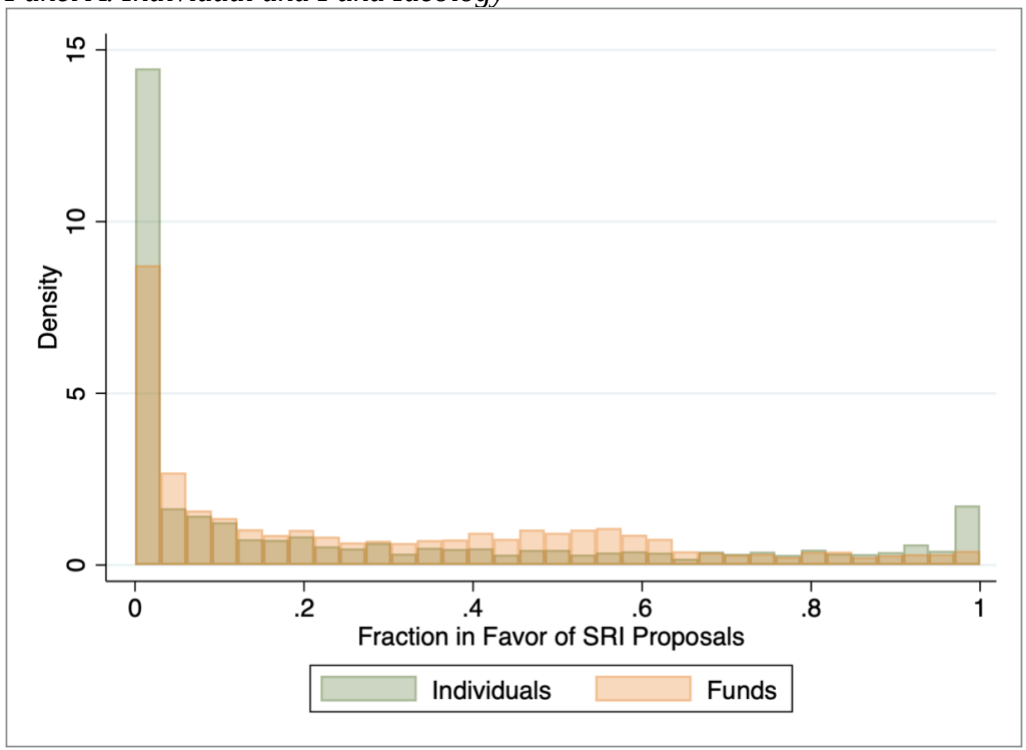

Change in Voting with Management
(Percentage Points)
4
2
0
-2
-4
2015
2016
2017
Firms Switching from:
Notice to Paper
Paper to Notice

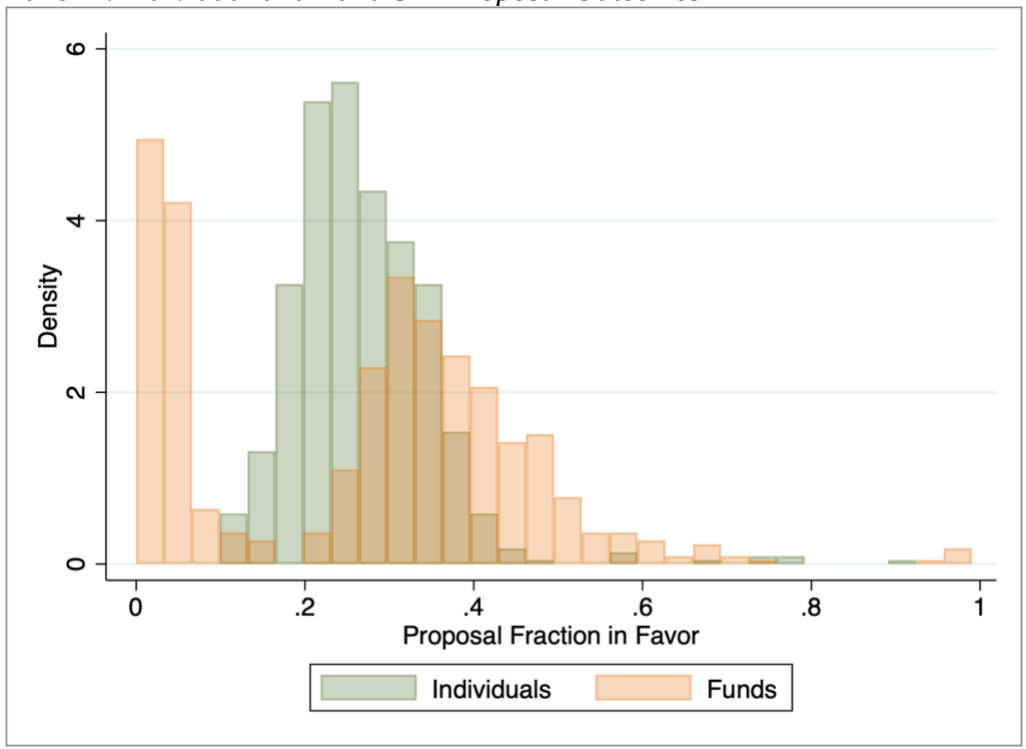
Figure 3. Ideology of Individuals and Funds

This figure displays distributions related to the ideology of individuals and funds with respect to SRI proposals. Panel A displays the distribution of individual and fund ideologies at the shareholder level, calculated as the fraction of SRIs proposals voted in favor. The figure is limited to investors with at least 10 votes on SRI proposals. Panel B displays the distribution of SRI proposal results among individuals and mutual funds, calculated as the fraction of votes in favor, votes weighted equally. Panel C displays the distribution of fund ideologies (with at least 10 SRI votes) and the distribution of the mean account at the fund (with at least 15 accounts).

Panel A. Individual and Fund Ideology



Panel B. Individual and Fund SRI Proposal Outcomes



Panel C. Fund Ideology and Fund Mean Account Ideology

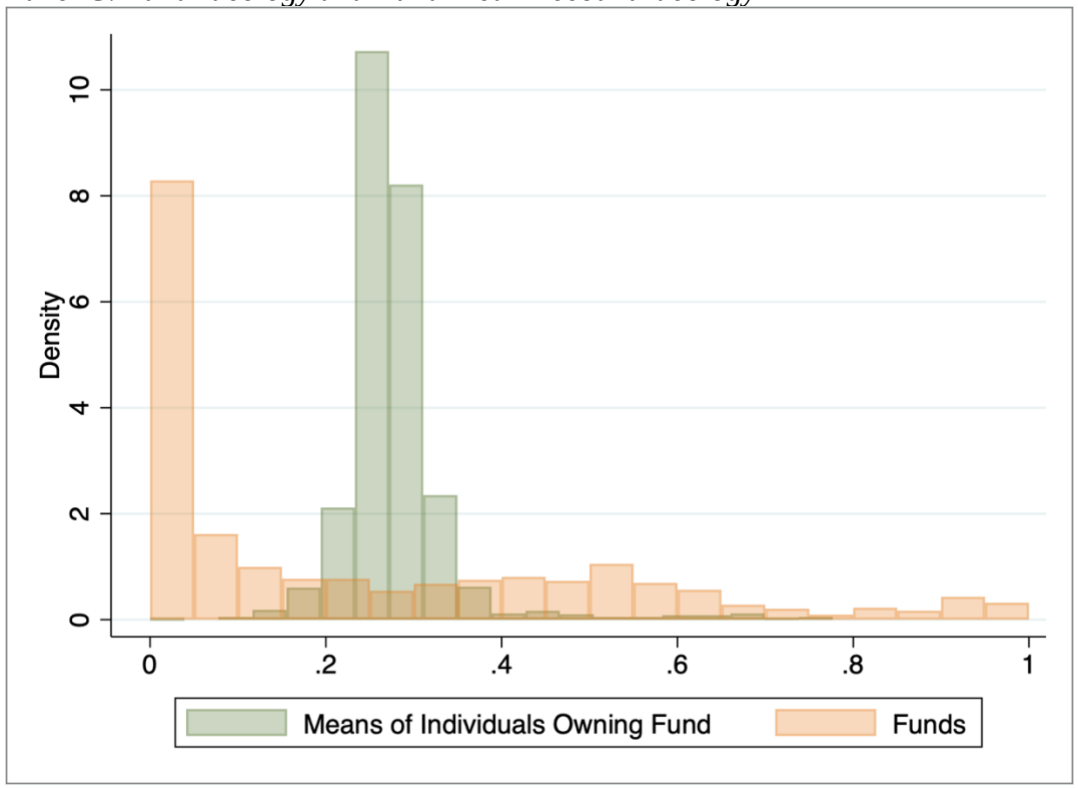


Figure 4. Scatterplots of Fund Ideology with Mean Account Ideology.

This figure plots fund votes and the aggregate votes of the individuals who own the fund. Observations are at the fund level. The sample is limited to individuals with at least 10 votes on SRI proposals and funds with at least 15 qualifying accounts and 10 votes on SRI proposals. Fund Votes on SRI proposals refers to the fund's fraction of SRI proposals voted in favor; Fund's Account Mean Votes on SRI Proposals refers to the average, across investors in the fund, of their percentage of SRI proposals voted in favor.

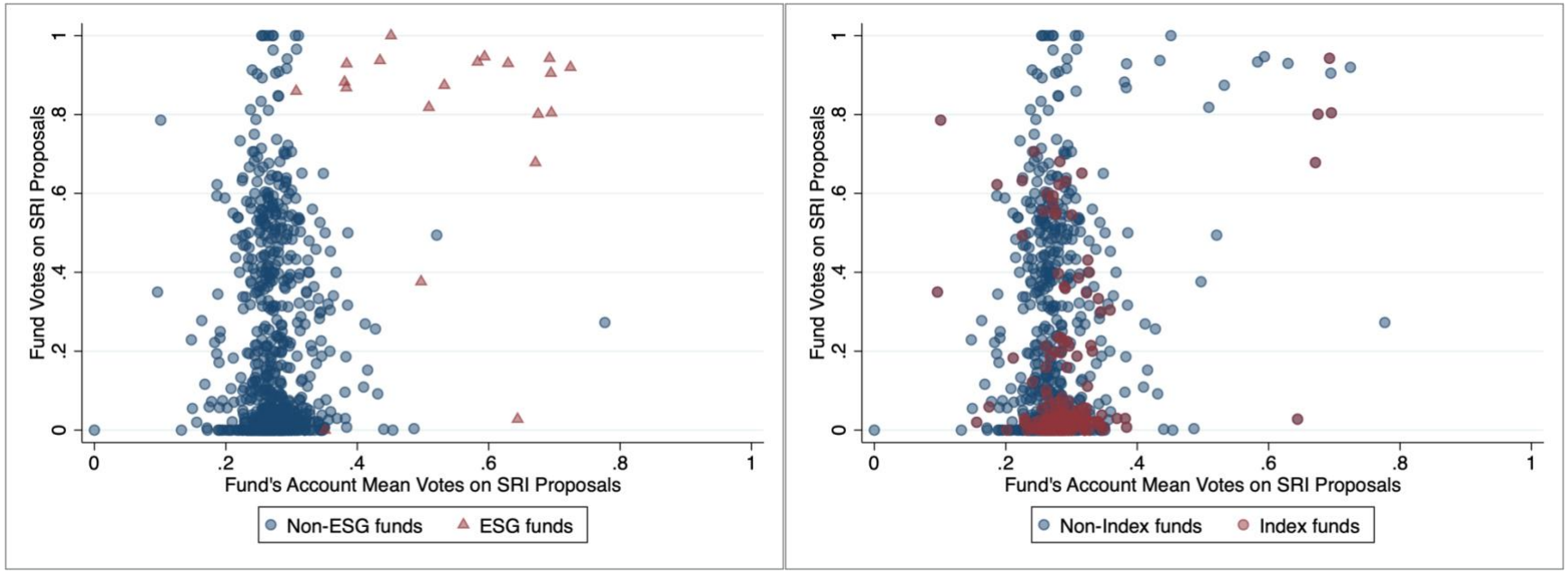


Figure 5. Replication of Fund Scatterplots with Zip Code Ideology

This figure plots fund votes and the aggregate zip code ideology of the individuals who own the fund. Observations are at the fund level. Zip code Democrat share is calculated as the two-party share earned by Hillary Clinton in the 2016 U.S. Presidential Election. The sample is limited to funds with at least 100 qualifying accounts and 15 votes on SRI proposals. Fund Votes on SRI proposals refers to the fund's percentage of the time voting in favor of SRI proposals; Fund's Account Mean Zip Code Democrat Shares refers to the average, across investors in the fund, of their zip code Democrat share.

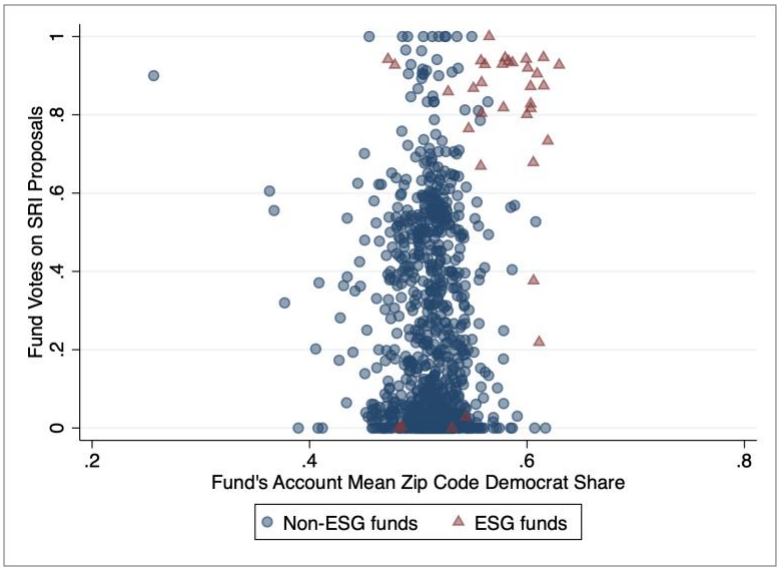


Table 1. Summary Statistics on Voting on SRI

This Table contains summary statistics on Environmental and Social shareholder proposals. Different rows reflect different methods of aggregation. The row marked "Votes weighted equally" totals up all votes across all proposals, each vote weighted equally, so if there were two proposals with 50 ballots cast and 150 ballots cast, respectively, then the percentages in the table would reflect the percentage of "For" votes out of the 200 total votes cast. The rows marked "Investors weighted equally" aggregate to the voter level (that is, each individual or fund's percentage in favor) before averaging across voters, each voter weighted equally. For Panel A, when weighting investors equally, I limit to investors who cast at least 10 votes on SRI, governance, or management proposals, respectively. For Panel B, when weighting investors equally, I limit to investors who cast at least 10 SRI votes in the first row, and make no limitation in the second. The row marked "Value-weighted" weights each vote by the value of the voter's stake in the firm, averaging across all votes on all proposals. The section titled "Percent of Proposals with 50% Support" treats proposals as the unit rather than votes, so if 20 out 100 proposals achieve 50% support the table would report 20%. The column titled Funds consists of all funds that appear in CRSP and ISS Voting Analytics that cast a ballot on an SRI proposal (regardless of whether they appear in my individual investor data).

Panel A. Overall support

	SRI		Gov.		Mgmt.	
	Funds	Individual Voters	Funds	Individual Voters	Funds	Individual Voters
Votes weighted equally	24.3	24.8	47.1	26.8	94.4	92.4
Investors weighted equally—minimum 10 votes	25.6	25.0	49.8	26.6	93.4	91.8

Panel B. SRI Votes Only

	Funds	Individual Voters Who Own Funds	All Individual Voters	Management	ISS	Glass Lewis
Percent of Votes in Favor						
Votes weighted equally	24.3	24.8	24.8	0.8	65.3	32.8
Investors weighted equally—minimum 10 votes	25.6	25.0	25.0	--	--	--
Investors weighted equally—all voters	29.3	29.8	29.8	--	--	--
Value-weighted	8.6	12.2	11.7	0.0	50.0	22.5
Num. votes	396,421	36,510,289	37,702,595	669	669	632
Num. investors	5,171	3,534,694	3,650,092	1	1	1
Percent of Proposals with 50% Support						
Votes weighted equally	8.0	1.4	1.4	--	--	--
Value-weighted	4.1	1.5	1.5	--	--	--

Table 2. Effects of a Change in Materials Delivered

In this Table, I measure the effects of the firm switching an individual's voting materials. In column 1, the outcome variable is v_{ict}, whether individual i voted at firm c's meeting at time t. In column 2, the outcome variable is r_{ict}, the fraction of proposals voted in favor of management. I limit to annual meetings and remove firms that send a mix of materials at any of their meetings in the sample period. I identify firms that switch from notice to full package or vice versa; I keep switchers and non-switchers but remove firms that switch more than once. For column 1, to make the sample size computationally tractable, I randomly sample 3,000 shareholders for each firm for non-switching firms that have more than 3,000, and weight observations by the inverse sampling probability. For column 2, I limit to investors that vote in all of the firm's meetings during the sample period. I estimate:

$$y_{ict} = \beta_1 NoSelection_{ic0} full_{c0} switcher_c post_{ct} +$$
$$\beta_2 NoSelection_{ic0} notice_{c0} switcher_c post_{ct} + \phi_{ic} + \psi_{it} + \theta_{ct}$$

Where $switcher_c$ is an indicator for whether firm c switches the materials it sends during the sample period; $post_{ct}$ is an indicator for whether the observation in question is post-switch; $full_{c0}$ and $notice_{c0}$, capture firm c's materials at the start of the sample period; and $NoSelection_{ic0}$, indicates whether shareholder i did not select for firm c at time 0. The specification includes individual-firm, individual-year, and firm-year fixed effects. Standard errors clustered at the account and meeting level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

	(1)	(2)
Outcome Variable:	Vote Cast (%)	With Management (%)
$NoSelection_{ic0} notice_{c0} switcher_c post_{ct}$	-2.255***	-0.348
	(0.154)	(0.641)
$NoSelection_{ic0} notice_{c0} switcher_c post_{ct}$	2.736***	-1.003
	(0.225)	(0.939)
Constant	15.002***	89.179***
	(0.013)	(0.042)
Account-Firm FE	Yes	Yes
Account-Year FE	Yes	Yes
Firm-Year FE	Yes	Yes
N	3,472,023	1,424,693
Number of clusters	7,839	7,236
R^2	0.9268	0.8892

Table 3. Relationship Between Firm's Choice and Individual Voting Decisions for Individuals Unaffected by Firm's Choice

In this table, I test whether there is a relationship between the firm's choice of what voting methods to send shareholders (its mail rate) and the voting choices made by its investors who are unaffected by its choice (those who select a voting method). The sample consists of 45 million randomly drawn individual votes at the proposal level, then limited to those shareholders who selected the materials they would receive for that meeting. I estimate:

$$r_{ictp} = \alpha_0 + m_{ct} + \varepsilon_{ictp}$$

In which r_{ictp} is an indicator for whether shareholder i voted with management on proposal p at firm c at time t, and m_{ct} is the fraction of the firms individual investors *who did not make a selection* that the firm sent full package materials to as opposed to notice. In Panel A, the first three columns each contain meetings from a single year; the fourth column contains all observations. Panel B adds fixed effexts. Standard errors clustered at the account and meeting level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

Panel A

	(1) 2015	(2) 2016	(3) 2017	(4) All
Meeting Mail Rate	-0.894	-0.472	0.312	-0.215
	(0.830)	(0.797)	(0.714)	(0.468)
Constant	90.738***	89.077***	89.333***	89.630***
	(0.288)	(0.325)	(0.399)	(0.214)
N	4,695,163	5,521,120	5,979,296	16,195,579
Number of clusters	4,126	4,096	3,970	12,192
R^2	0.0001	0.0000	0.0000	0.0000

Panel B

	(1)	(2)	(3)
Meeting Mail Rate	-0.360	-0.205	0.019
	(0.454)	(0.386)	(0.393)
Constant	89.651***	89.750***	90.006***
	(0.203)	(0.144)	(0.068)
Year Fixed Effects	Yes	Yes	Yes
Investor Fixed Effects	No	Yes	Yes
Investor-Firm Fixed Effects	No	No	Yes
N	16,195,579	15,711,599	13,006,818
Number of clusters	12,192	12,028	10,197
R^2	0.0005	0.4343	0.6288

Table 4. Sample-Selection Adjusted Voting Choices

In this table, I measure average individual investor voting choices adjusting for sample selection bias on observables and unobservables. Specifically, I estimate the equation:

$$R_i = \alpha + \kappa X_i + \varepsilon_i \qquad (2)$$

In which i indexes individuals and R_i is the individual's percentage voting with management. In the bottom (bolded) row of the table, I display the average R_i using different adjustments for sample selection bias. In Column 1 I estimate Equation 2 with no right-hand-side variables, producing a simple unweighted average R_i. In Columns 2 through 4, Inverse Probability Weighting, I first predict the likelihood of shareholder i being a voter using a probit regression (with probit results displayed in corresponding Appendix Table A2), calculate regression weights as the inverse probabilities, and use those regression weights in an OLS regression of Equation 2. The bottom row, our average R_i estimate, displays the results of a regression with those regression weights and no right-hand-side variables. In Columns 5 through 7, Heckman selection, I jointly estimate (by maximum likelihood estimation) the likelihood of being in the sample (with results displayed in corresponding Appendix Table A2) and Equation 2, with the former including an exclusion-restricted variable and the latter adjusting for sample selection bias by including the Mills Ratio as a covariate. Columns 4 and 7 include additional unshown covariates: log account value to the second, third, fourth, and fifth powers; log number of firms owned to the second and third power, the interaction of account value and firms owned, County-level 2016 Republican Presidential Share squared, zip code log average income and its square, county-level 2016 presidential political turnout, zip code-level percentage with bachelors' degree and its square, zip code-level percentage with post-bachelors degree, zip code median age, zip code percentage over age 65 squared, zip code-level fraction Black or Hispanic, zip code-level density, latitude, longitude, longitude squared, average log market cap of firms owned squared, average record date for firms owned (and its square), fraction of firms owned that are in oil or gas, average return on assets for firms owned, average number of individual investors who own the firms in one's portfolio (squared, cubed, and to the fourth power). The sample of individuals consists of a simple randomly drawn subset. Robust standard errors are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

	(1) Simple	(2) IPW	(3) IPW	(4) IPW	(5) Heckman	(6) Heckman	(7) Heckman
Log Account Value		0.818***	0.700***	-0.795*	1.152***	1.065***	-0.873**
		(0.062)	(0.113)	(0.377)	(0.017)	(0.019)	(0.309)
Log Number of Firms Owned		-0.728***	-0.498***	-1.877***	-0.859***	-0.768***	-2.198***
		(0.076)	(0.133)	(0.217)	(0.022)	(0.021)	(0.142)
Average Log Market Cap of Firms Owned			0.678***	4.122***		1.158***	3.300***
			(0.107)	(0.257)		(0.026)	(0.156)
County 2016 Republican Presidential Share			6.041***	6.949***		5.605***	7.077***
			(0.295)	(0.997)		(0.120)	(0.639)
Zip Code Fraction Over Age 65			10.800***	25.794***		6.865***	27.352***
			(0.696)	(1.340)		(0.231)	(0.891)
Average Log Number of Individual Owners, across Firms Owned			-1.106***	-0.644***		-1.441***	-0.979***
			(0.112)	(0.052)		(0.028)	(0.035)
Intercept	87.988***	87.109***	86.907***	86.941***	86.448***	86.014***	86.476***
	(0.017)	(0.025)	(0.028)	(0.024)	(0.201)	(0.203)	(0.211)
N	1,756,803	1,756,803	1,756,803	1,756,803	12,848,583	12,848,583	12,848,583
Additional Covariates	No	No	No	Yes	No	No	Yes
Sample Bias-Adjusted Average	**87.988*** **	**86.879*** **	**86.673*** **	**86.898*** **	**86.873*** **	**85.689*** **	**86.376*** **
	(0.017)	(0.032)	(0.050)	(0.026)	(0.017)	(0.018)	(0.018)

Table 5. Individual and Fund Variation in Voting

In this table, I estimate the following specification:

$$R_{ipct} = \delta_0 + (\delta_1 X_{pct}) + (\phi_p) + (\psi_i) + \phi_t + \varepsilon_{ipct}$$

i indexes investors, p indexes proposals, c indexes firms, t indexes years, Y_{ipct} is i's vote on proposal p (multiplied by 100), X_{pct} is a vector of firm-month and proposal-level covariates, and ϕ_p, ψ_i, and ϕ_t are proposal, investor, and year fixed effects, respectively. Columns 1, 3, and 5 are for individuals; Columns 2, 4, and 6 are for funds. The sample of individuals consists of a randomly drawn subset of 500,000 individuals with SRI votes in consecutive years at different firms. All regressions are weighted so that each investor has equal weight. All right-hand-side variables are demeaned over the regression sample, so that the intercept term reflects the average value of the dependent variable in the sample (no other coefficient is affected). Standard errors clustered at the account and proposal level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

	(1) Individuals	(2) Funds	(3) Individuals	(4) Funds	(5) Individuals	(6) Funds
Total q	-0.16	0.01				
	(0.14)	(0.11)				
ROA	3.07	6.94				
	(6.02)	(3.74)				
Market to Book Ratio	-0.02	-0.04				
	(0.11)	(0.07)				
ISS in Favor	6.32***	31.34***				
	(1.39)	(1.20)				
Glass Lewis in Favor	0.34	14.89***				
	(1.41)	(1.46)				
Constant	27.71***	27.48***	27.86***	28.63***	27.18***	28.09***
	(0.58)	(0.61)	(0.04)	(0.41)	(0.34)	(0.90)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Proposal Fixed Effects	No	No	Yes	Yes	No	No
Investor Fixed Effects	No	No	No	No	Yes	Yes
R^2	0.01	0.19	0.03	0.22	0.66	0.42
N	6,425,963	228,962	7,230,501	264,932	7,190,605	264,799
Number of Clusters	397	389	466	458	466	458

Table 6. Systematic Heterogeneity in Voting

In this table, I estimate the following specification, following Matvos and Ostrovsky (2010):

$$R_{ipct} = \beta_0 + \beta_1 Y_{i(-c)(t-1)} + \phi_t + \varepsilon_{ipct}$$

In which i indexes investors, c indexes firms (and $-c$ denotes firms other than c), and t indexes time. The left hand side is the investor's vote on a particular proposal. The right hand side contains the same investor's vote (randomly selected) for a different firm in the previous year. Column 1 contains individual votes; column 2 contains fund votes; and column 3 contains both, with an interaction term capturing the difference between them. The sample of individuals consists of a randomly drawn subset of 500,000 individuals with SRI votes in consecutive years at different firms. Standard errors clustered at the account level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

	(1)	(2)	(3)
Data Universe:	Individuals	Funds	Both
Last year's vote	47.7***	36.2***	37.4***
	(0.1)	(0.7)	(0.8)
Individual			-2.5***
			(0.4)
Last year's vote × Individual			10.3***
			(0.8)
Constant	15.1***	19.4***	17.6***
	(0.0)	(0.4)	(0.4)
Proposal Fixed Effects	Yes	Yes	Yes
R^2	0.25	0.34	0.25
N	7,230,501	264,932	7,495,433
Number of Clusters	500,000	3,792	503,792

Table 7. Representation

This table presents a regression of fund ideology on the mean ideology of its underlying individual investors. Specifically, I estimate:

$$\bar{R}_f = \beta_0 + \beta_1 \sum_{a \in \Theta_f} \bar{R}_a + \nu_f$$

In which f indexes funds, a indexes individual accounts, Θ_f is the set of accounts at fund f, and $\bar{R}_f$ and $\bar{R}_a$ are fund f's and account a's fractions of SRI proposals voted in favor. Observations are at the fund level, with accounts aggregated within a fund. The sample is limited to funds and individuals with at least 10 SRI votes, and to funds with at least 15 qualifying individual investors. Below the table, I include the impact on β_1 of including coefficients, along with standard errors, calculated using the *b1x2* STATA function created in Gelbach (2016). Standard errors clustered at fund shareholder meeting level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

	Funds Weighted Equally				Funds Weighted by Number of Accounts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction in Favor of SRI, Fund Mean Owner	0.895***	0.006	0.158	0.221	0.495	-0.703	0.312	0.210
	(0.256)	(0.192)	(0.176)	(0.189)	(0.649)	(0.528)	(0.405)	(0.283)
ESG Indicator		0.573***	0.534***	0.522***		0.806***	0.646***	0.710***
		(0.092)	(0.085)	(0.088)		(0.085)	(0.058)	(0.050)
Index Indicator			-0.117*	-0.099			-0.127***	-0.105**
			(0.047)	(0.054)			(0.028)	(0.037)
Intercept	-0.039	0.195***	0.181**	0.151*	0.011	0.326*	0.093	0.096
	(0.077)	(0.053)	(0.055)	(0.060)	(0.167)	(0.136)	(0.111)	(0.082)
Fund Type FE	No	No	No	Yes	No	No	No	Yes
R^2	0.05	0.12	0.15	0.22	0.00	0.09	0.15	0.29
N	800	800	800	671	800	800	800	671
Num Clusters	291	291	291	193	291	291	291	193
Impact of Including *ESG* on Coefficient		0.888**	0.829***			1.198**	0.960**	
		(0.273)	(0.243)			(0.393)	(0.366)	
Impact of Including *Index* on Coefficient			-0.092				-0.776*	
			(0.066)				(0.313)	

Table 8. Ownership on Ideology Match

In this table, I estimate the relationship between ownership and the match between individual and fund ideology. Specifically, I estimate account a's decision whether to own fund f as follows:

$$Q_{af} = \kappa_0 + \kappa_1\left(\bar{R}_f - \bar{R}_a\right)^2 + \kappa_2 X_{af} + \phi_f + \theta_a + \eta_{af}$$

In which Q_{af} equals 1 (multiplied by 100) if account a owns fund f, $\bar{R}_f$ and $\bar{R}_a$ are fund and account fraction of SRI proposals voted in favor, X_{af} is a vector of additional variables concerning the match between a and f, ϕ_f is fund fixed effects and θ_a is account fixed effects. X_{af} is distance between individual and fund headquarters. I select, for each account, twenty control group funds at random that it does not own, and calculate the sampling weight as the inverse of the probability of being selected. Standard errors clustered at the account and fund shareholder meeting level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

Dependent variable: Ownership*100	(1)	(2)	(3)	(4)
(Fund SRI Fraction - Account SRI Fraction)^2	0.005	0.014	0.005	0.007
	(0.009)	(0.008)	(0.005)	(0.005)
ESG Indicator × Account SRI Fraction		0.072***	0.060**	0.067**
		(0.020)	(0.020)	(0.022)
Index Indicator × Account SRI Fraction			0.059***	0.054***
			(0.015)	(0.015)
Log Distance in Miles				-0.015***
				(0.004)
Intercept	0.240***	0.238***	0.236***	0.351***
	(0.002)	(0.002)	(0.001)	(0.026)
Account Fixed Effects	Yes	Yes	Yes	Yes
Fund Fixed Effects	Yes	Yes	Yes	Yes
R^2	0.01	0.01	0.01	0.01
N	7,884,474	7,884,474	7,884,474	6,521,915
Num Clusters	346	346	346	308

Table 9. Non-Salient Fund and Firm Characteristics

This table presents estimates designed to capture whether features of an individual's fund or firm holdings predict her ideology. In Panel A, I estimate:

$$\sum_{a \in \Theta_f} \bar{R}_a = \delta_0 + \delta_1 X_f + \varepsilon_f$$

In which f indexes funds, a indexes individual accounts, Θ_f is the set of accounts at fund f, $\bar{R}_a$ is account a's percentage of SRI proposals voted in favor, and X_f is a vector of fund features. Observations are at the fund level, with accounts aggregated within a fund. The sample is limited to individuals with at least 10 SRI votes, and to funds with at least 15 qualifying individual investors. Standard errors clustered at fund shareholder meeting level are in parentheses. Panel B contains results from the following regression:

$$\bar{R}_a = \beta_0 + \beta_1 X_c + \phi_t + \varepsilon_{it}$$

In which a indexes investors, t indexes years, c indexes firms, and X_c is a vector of industry categories. Standard errors clustered at the account and firm level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

Panel A: Individual Selection of Funds

Dep. Variable: Fund Mean Account SRI	(1)	(2)	(3)	(4)	(5)
ESG Indicator		0.264***		0.258***	0.258***
		(0.039)		(0.030)	(0.030)
Morningstar Sustainability Rating	0.006	-0.003			
	(0.005)	(0.002)			
Fossil Fuels			-0.016	-0.008	0.003
			(0.010)	(0.010)	(0.010)
			(0.045)	(0.031)	(0.032)
Renewable Energy			0.356***	0.085	0.087
			(0.054)	(0.063)	(0.064)
Energy Fund Name					-0.025*
					(0.011)
Constant	0.263***	0.282***	0.283***	0.277***	0.277***
	(0.011)	(0.007)	(0.004)	(0.003)	(0.003)
R^2	0.01	0.42	0.03	0.36	0.36
N	745	745	1,140	1,140	1,140
Number of Clusters	203	203	288	288	288

Panel B: Individual Selection of Firms

	(1)	(2)	(3)	(4)	(5)
Dep. Variable: SRI Percentage of Holder					
Fossil Fuels	-0.99**		-1.10**		-1.30***
	(0.36)		(0.40)		(0.39)
Renewable Energy	7.19***		7.32***		7.37***
	(1.72)		(1.81)		(1.77)
KLD SRI Score		-0.06	-0.07		
		(0.04)	(0.04)		
KLD Environmental Score				-0.21*	-0.23*
				(0.10)	(0.11)
Constant	28.71***	28.81***	28.81***	28.95***	28.98***
	(0.16)	(0.21)	(0.21)	(0.20)	(0.21)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
R^2	0.00	0.00	0.00	0.00	0.00
N	5,000,000	4,376,471	4,376,471	4,376,471	4,376,471
Number of Clusters	5,591	2,596	2,596	2,596	2,596

Table 10. Investment Size

This table presents estimations related to an account's average stake size. In Panel A, similar to Table 8, I estimate the relationship between ownership and the match between individual and fund ideology. Specifically, I estimate account a's decision whether to own fund f as follows:

$$Q_{af} = \kappa_0 + \kappa_1\left(\bar{R}_f - \bar{R}_a\right)^2 + \kappa_2 LogAverageAmount_a + \kappa_3\left(\bar{R}_f - \bar{R}_a\right)^2 * LogAverageAmount_a + \kappa_4 X_{af} + \phi_f + \theta_a + \eta_{af} \quad (4B)$$

Where $LogAverageAmount_a$ is the log of the average stake value in account a's securities, Q_{af} equals 1 (multiplied by 100) if account a owns fund f, $\bar{R}_f$ and $\bar{R}_a$ are fund and account fraction of SRI proposals voted in favor, $\left(\bar{R}_f - \bar{R}_a\right)^2$ is a measure of ideological distance, ϕ_f is fund fixed effects and θ_a is account fixed effects. X_{af} includes the distance between individual and fund headquarters and the interaction of distance with the account's (log) number of SRI votes cast. I exclude all ESG funds. I select, for each account, twenty control group funds at random that it does not own, and calculate the sampling weight as the inverse of the probability of being selected. Standard errors clustered at the account and fund shareholder meeting level are in parentheses.

Panel B recreates Table 7 weighting accounts by ownership of the fund. Specifically, I estimate:

$$\bar{R}_f = \beta_0 + \beta_1 \sum_{a \in \Theta_f} w_{af} \bar{R}_a + \nu_f$$

In which f indexes funds, a indexes individual accounts, w_{af} is account a's ownership fraction of fund f, Θ_f is the set of accounts at fund f, $\bar{R}_f$ is the fund's fraction in favor of SRI proposals, and R_a is account a's fraction in favor of SRI proposals. Observations are at the fund level, with accounts aggregated within a fund. The sample is limited to funds and individuals with at least 10 SRI votes, and to funds with at least 15 qualifying individual investors. Standard errors clustered at fund shareholder meeting level are in parentheses. *, **, and *** represent significance at the 0.05, 0.01, and 0.001 levels, respectively.

Panel A: Ownership on Ideology Match, Interacted with Account's Average Investment Value

Dependent variable: Ownership*100	(1)	(2)	(3)
(Fund SRI Fraction - Account SRI Fraction)^2	0.256**	0.254**	0.237**
	(0.090)	(0.090)	(0.090)
Account Log Average Investment Value	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
(Fund SRI Fraction - Account SRI Fraction)^2 × Account Log Average Investment Value	-0.024**	-0.025**	-0.025**
	(0.009)	(0.009)	(0.009)
Log Distance in Miles	-0.015***	-0.015***	-0.015***
	(0.004)	(0.004)	(0.004)
Account Number of SRI Votes		0.000	
		(0.000)	
(Fund SRI Fraction - Account SRI Fraction)^2 × Account Number of SRI Votes		0.000	
		(0.000)	
Log Account Number of SRI Votes			0.000
			(0.000)
(Fund SRI Fraction - Account SRI Fraction)^2 × Log Account Number of SRI Votes			0.007
			(0.005)
Intercept	0.359***	0.359***	0.359***
	(0.027)	(0.027)	(0.027)
Account Fixed Effects	Yes	Yes	Yes
Fund Fixed Effects	Yes	Yes	Yes
R^2	0.01	0.01	0.01
N	6,292,362	6,292,362	6,292,362
Num Clusters	300	300	300

Panel B: Representation, Accounts Weighted by Ownership Fraction of the Fund

Dependent variable: Fund For % on SRI Props	(1) Funds Weighted Equally	(2) Funds Weighted Equally	(3) Funds Weighted Equally	(4) Funds Weighted Equally	(5) Funds Weighted by Number of Accounts	(6) Funds Weighted by Number of Accounts	(7) Funds Weighted by Number of Accounts	(8) Funds Weighted by Number of Accounts
For % on SRI Props, Fund Mean Owner	0.912***	0.473**	0.440**	0.487**	1.333***	0.848***	0.472**	0.375*
	(0.164)	(0.152)	(0.147)	(0.161)	(0.265)	(0.171)	(0.179)	(0.150)
ESG Indicator		0.439***	0.447***	0.437***		0.565***	0.606***	0.666***
		(0.101)	(0.095)	(0.098)		(0.072)	(0.049)	(0.051)
Index Indicator			-0.112*	-0.096			-0.109***	-0.095*
			(0.047)	(0.054)			(0.029)	(0.039)
Intercept	-0.014	0.084	0.118**	0.094	-0.170**	-0.062	0.060	0.063
	(0.051)	(0.049)	(0.044)	(0.049)	(0.055)	(0.043)	(0.048)	(0.044)
Fund Type FE	No	No	No	Yes	No	No	No	Yes
R^2	0.08	0.13	0.16	0.23	0.07	0.12	0.17	0.30
N	798	798	798	669	798	798	798	669
Num Clusters	290	290	290	192	290	290	290	192

Data Appendix

In this Data Appendix, I describe in greater detail the cleaning and merging of mutual fund ownership data from Broadridge.

Overview

The Broadridge mutual fund dataset consists of observations at the account-mutual fund meeting level. Each observation contains the CUSIP of fund mutual fund, a code identifying the meeting, the meeting date, the record date, an anonymized account identifier, and anonymized broker identifier (each of which identities are anonymous with non-anonymous data retained by Broadridge), the account's number of shares, and the account's 5-digit zip code.[31]

Hand-cleaning of meeting codes

Each mutual fund shareholder meeting (that is, a shareholder meeting held by the mutual fund, as opposed to one that the mutual fund is participating in as an owner) is identified by a Broadridge code. Multiple meeting codes frequently refer to the same meeting. An account that owns shares of the mutual fund might be recorded as owning shares for one meeting code of a fund and not the others. This multiplicity of meeting codes arises because meetings sometimes require separate mailings to different shareholders, or because mutual fund shareholder elections are frequently extended or delayed. I classify multiple Broadridge meeting codes as a single meeting if their meeting date is within 6 months of each other and if their accounts have less than 2% overlap. Using this method reclassifies nearly 9,000 meetings. All but a small handful of reclassified meetings have no account overlap with the meetings they are re-categorized into.

Other datasets

I merge mutual fund shareholder meetings from Broadridge to CRSP open-end mutual funds (crsp.fund_names) by 9-digit CUSIP, limited to matches in which the meeting date of the shareholder meeting is within the date range for which the CRSP CUSIP is valid. For those funds where there is no match, I obtain additional matches (generally cases where the CUSIP has changed over time) by hand-matching funds based on information from public filings.

I merge to CRSP monthly total net asset values and net asset value per share (crspq.fund_summary) using the CRSP fund number, matching the last monthly CRSP record

[31] There are also several additional fields regarding the account's participation in the meeting that I do not make use of.

prior to the shareholder meeting's record date, so long as the monthly CRSP record is within 35 days of the shareholder meeting's record date. CRSP fund numbers refer to a share class of the fund, not the fund itself (referred to as the "portfolio" within CRSP); therefore, to determine a fund's size, I aggregate across share classes within a fund.

For closed-end mutual funds, which have a single share class and trade on securities markets, I match to CRSP's monthly securities update by 9-digit CUSIP, and hand-code to add additional matches and correct mistaken matches.

I merge at the fund level to ISS fund voting results. I use the ISS-CRSP match from Brav et al. (2020) for the open-end funds, with additional hand-coded matches, and hand-match the closed-end funds.

As in Brav et al. (2020), I match the direct ownership of equities data with CRSP monthly securities and Compustat. I obtain information on proposals from FactSet SharkRepellent and ISS Voting Analytics, the latter of which I also obtain open- and closed-end fund votes disclosed on Form NPX.

I match by zip code with zip code-level Internal Revenue Service data, zip code-level Census data, and Bureau of Labor Statistics data on educational attainment. Each of these sets is used in Brav et al. (2020) and merging procedures are described in detail therein.

I use 2016 county presidential election results from MIT's election lab, then map counties to zip codes.

I add in zip-code longitude and latitude from the U.S. Census Bureau via CivicSpace Labs.

Key Variables

I calculate an account's mutual fund stake value by multiplying the account's number of shares at the time of the mutual fund's meeting by the share price at the same time. If the fund has multiple meetings, I average the account's stake value across meetings.

To identify index funds, I use the flag from CRSP's Open-End fund dataset, counting funds with flag "D" or "B" as an index fund. I identify closed-end funds as non-index funds.

To identify ESG funds, I start with the 235 funds identified in a Morningstar report from January 2018 (Hale (2018)), which "defined the U.S. sustainable funds universe as those open-end funds and exchange-traded portfolios that, by prospectus, state that they incorporate ESG criteria into their investment processes, or indicate that they pursue a sustainability-related

theme, or seek measurable sustainable impact alongside financial return." To those I add funds with the word "Sustainable", "ESG", "Social", or "Clean Energy" in the name, as well as funds from five institutions that are explicitly mission-driven: Calvert, Pax, Parnassus, Trillium, and Praxis.

I obtain ISS recommendations from ISS Voting Analytics, as is standard. For Glass Lewis recommendations, I modify a procedure set forth in Bubb and Catan (2020). The procedure uses a dataset from ProxyInsights that contains, for each mutual fund, the percentage of time it agrees with ISS and Glass Lewis. I use this data to back out the Glass Lewis recommendation.

Coverage

Most mutual funds are organized as Maryland corporations or Delaware or Massachusetts business trusts.[32] Open-end funds, including exchange-traded funds, do not hold annual shareholder meetings unless required by their charter documents.[33] Therefore, the dataset only includes ownership and votes in open-end mutual funds that happened to hold a shareholder meeting in the three-year period 2015-2017, generally due to a proposed change in the shareholder advisory agreement. Those funds that do not have a meeting in the three-year period generally did not have one because they had no proposed changes to the shareholder advisory agreement, because they had a majority shareholder, or because they had an exemptive order from the SEC allowing them to alter subadvisory agreements without a shareholder vote.

Data Appendix, Table 1, Panel A shows coverage of open-end mutual funds that appear in the CRSP Survivor-Bias Free Open-End Mutual Fund Database. In total, I have 29.8% of open-end funds in CRSP (and a greater portion of large open-end funds), constituting $32.4T in assets, slightly more than half of total assets in the CRSP open-end mutual fund dataset.

Only mutual funds that own and vote on equities file form NPX, so only a subset of CRSP funds appear in ISS. Data Appendix, Table 1, Panel B shows the universe of CRSP fund share classes that I successfully matched to ISS's fund-level NPX dataset. This intersection comprises my main sample. I successfully match 32.8% of funds in this sample.

Unlike open-end funds, closed-end funds are required to hold annual meetings. I designate a closed-end fund in CRSP as one with share code 14 or 44. Data Appendix Table 1,

[32] K&L Gates 2013

[33] K&L Gates 2013, Investment Lawyer 2015 p. 7. Closed-end funds are required to hold regular director elections.

Panel C show coverage in my dataset of closed-end mutual funds. Coverage is nearly comprehensive, with 94.7% of CRSP closed-end funds appearing in the retail dataset. Panel D shows the subset that match to ISS's NPX dataset—I have 100% coverage of such funds.

Finally, in Data Appendix, Table 1, Panel E, I evaluate what portion of funds in the individual mutual fund ownership dataset from Broadridge are usable (by being matched to CRSP fund data and ISS voting data). Because the retail dataset is at the CUSIP level, which corresponds to a share class, I can only evaluate the portion of matching share classes, not the portion of matching funds. I successfully match the vast majority of funds in the Broadridge dataset to CRSP mutual funds. Because many of those funds do not own and vote on U.S. equity securities, only a smaller fraction is matched to ISS.

In Data Appendix, Table 5, I present the numbers of accounts in the two different datasets from Broadridge.

Data Appendix, Table 1. Coverage Statistics

Panel A. Coverage of CRSP Open-End Mutual Fund Set by Broadridge Retail Dataset

Size Quintile:	Share Classes			Funds			Value (in billions)		
	Number of CRSP Share Classes	Number of Share classes in CRSP and Retail Dataset	Coverage Percent	Number of CRSP Funds	Number of Funds in CRSP and Retail Dataset	Coverage Percent	Total CRSP NAV	Total NAV Matched to Retail	Coverage Percent
Smallest	8860	1879	21.2	3583	617	17.2	74	15.6	21.1
2	7856	2366	30.1	2802	777	27.7	441.6	127.6	28.9
3	7436	2237	30.1	2679	809	30.2	1585.2	479.6	30.3
4	7193	2348	32.6	2615	906	34.6	5137.9	1827.2	35.6
Largest	7101	2746	38.7	2579	1141	44.2	56510.9	30023.1	53.1
Total	38446	11576	30.1	14258	4250	29.8	63749.5	32473.1	50.9

Panel B. Coverage of CRSP Open-End Mutual Fund Set/NPX Overlap by Broadridge Retail Dataset

Size Quintile:	Share Classes			Funds			Value (in billions)		
	Number of CRSP Share Classes in ISS NPX	Number of Share classes in CRSP, NPX and Retail Dataset	Coverage Percent	Number of CRSP Funds in NPX	Number of Funds in CRSP, NPX and Retail Dataset	Coverage Percent	Total CRSP /NPX NAV	Total NAV Matched to Retail	Coverage Percent
Smallest	4883	1155	23.7	968	205	21.2	23.9	5.8	24.3
2	4584	1521	33.2	1136	310	27.3	181.2	50.6	27.9
3	4268	1394	32.7	1189	363	30.5	705.8	215	30.5
4	4437	1534	34.6	1271	451	35.5	2535	931.4	36.7
Largest	4615	1900	41.2	1354	611	45.1	30617.9	16108.1	52.6
Total	22787	7504	32.9	5918	1940	32.8	34063.8	17310.8	50.8

Panel C. Coverage of CRSP Closed-End Mutual Fund Set by Broadridge Retail Dataset

Size Quintile:	Funds			Value (in billions)		
	Number of CRSP Funds	Number of Funds in CRSP and Retail Dataset	Coverage Percent	Total CRSP Market Equity	Total Market Equity Matched to Retail	Coverage Percent
Smallest	136	130	95.6	9.8	9.4	95.9
2	128	118	92.2	20.5	18.9	92.2
3	125	120	96	33.2	32	96.4
4	121	113	93.4	60.5	56.9	94
Largest	114	110	96.5	190.9	187	98
Total	624	591	94.7	314.9	304.2	96.6

Panel D. Coverage of CRSP Closed-End Mutual Fund Set/NPX Overlap by Broadridge Retail Dataset

Size Quintile:	Funds			Value (in billions)		
	Number of CRSP Funds in NPX	Number of Funds in CRSP, NPX and Retail Dataset	Coverage Percent	Total CRSP Market Equity in NPX	Total Market Equity Matched to Retail	Coverage Percent
Smallest	57	57	100	4.5	4.5	100
2	59	59	100	9.5	9.5	100
3	64	64	100	17.3	17.3	100
4	71	71	100	37.8	37.8	100
Largest	66	66	100	73.9	73.9	100
Total	317	317	100	143	143	100

Panel E. Coverage of Retail-Owned Funds by CRSP Open-End and Closed-End Mutual Fund

Number of Accounts Quintile:	Number of Share Classes in Retail Dataset	Number of Share Classes in Retail Set and CRSP Open-End Dataset	Number of Share Classes in Retail Set and CRSP Closed End	Coverage Percent	Number of Share Classes in Retail Set, CRSP, and ISS	Coverage Percent
Fewest	2742	2267	195	89.8	1454	53
2	2706	2366	16	88	1464	54.1
3	2716	2474	60	93.3	1542	56.8
4	2721	2326	302	96.6	1737	63.8
Most	2720	2399	259	97.7	2014	74
Total	13605	11832	832	93.1	8211	60.4

Table 5. Numbers of Observations

	Portfolios of Equities		Portfolios of Funds		Total Portfolios
	All	Conditional on Matching to Fund	All	Conditional on Matching to Equity	All
Number of Unique Accounts	46,686,015	13,414,912	80,209,211	13,414,912	113,480,314
Number of Unique Accounts Matched to CRSP	41,886,035	12,110,222	78,800,757	13,298,142	108,666,339
Percentage of Accounts Owning Retail	100	100	16.7%	100	41.1%
Percentage of Accounts Owning Fund	28.7%	100	100	100	70.7%