Does gender diversity impact performance and corporate social responsibility investment decisions in mutual funds?

Lidan Grossmass, Ming-Tsung Lin and Ser-Huang Poon*

Preliminary draft:

March 12, 2019

Abstract

Our paper examines the impact of gender diversity of mutual fund teams on fund performance and corporate social responsibility investment decisions. We use a rich and unique dataset of fund managers and match this with the detailed investments of their portfolio. The portfolio holdings are matched with their respective KLD scores to obtain numerical ratings of the social impact of the teams' investments. We find that the effect of gender diversity on social investing is positive but negative on fund performance. Both effects are significant and nonlinear. We further condition our findings on economic or fund conditions. We find that gender diverse teams tend to hold socially responsible investments even in poorer economic conditions at the cost of fund performance.

JEL-Classification: G23, J21, L22

Keywords: Gender, Mutual Funds, SRI, panel threshold regression

*Lidan Grossmass (lidan.grossmass@hhu.de) is at Düsseldorf Institute of Competition Economics, Germany. Ming-Tsung Lin (ming-tsung.lin@dmu.ac.uk) is at De Montfort University, UK. Ser-Huang Poon (ser-huang.poon@manchester.ac.uk and corresponding author) is at the University of Manchester, Alliance Manchester Business School, UK. Funding from Alliance Strategic Research Grant is gratefully acknowledged. We would also like to thank Dr. Chen Yiping for research assistance.

1 Introduction

A Financial Times article[1] observed that the numbers of female mutual fund managers have been decreasing since 2009 and argued that this imbalance "poses tangible investment issues", because females possess qualities that contribute to better decision making. This hence poses the intriguing argument of what the value of gender diversity have on mutual fund teams. Our paper attempts to answer this question, concentrating on aspects of mutual fund performances and more interestingly, on the social responsibility of investment decisions.

Social responsible investing (SRI) has gained traction in recent years, as investors become more socially conscious. As of 2018, the market for socially responsible investments has reached $20 trillion in terms of asset under management around the world[2]. Investors, institutional and retail alike, are increasingly made aware of ethical investment and issues related to environmental sustainability. As such, it will become increasingly important for the mutual fund industry to consider Corporate Social Responsibility (CSR) objectives in addition to investment risks and returns. Here, the measure of mutual fund's SRI level is determined by the average CSR in their investment portfolios.

In this paper, we seek to answer several intricately linked questions: what is the (conditional) gender effect on fund performance, what is the (conditional) effect of SRI on fund performance, and what is the (conditional) gender effect on SRI. To answer these questions, we use a rich and unique dataset of fund managers with comprehensive background information including gender, age, education and work experiences, and match this with the detailed investments of their portfolio. The investments in their portfolios are matched with scores from KLD, and then we are able to measure how socially responsible of a fund is in its investment.

Our study is linked to several strands in the current mutual fund literature. The most established strand of literature studies the relationship between fund performance and manager gender. It is widely perceived that there is gender difference in

[1]"Chris Newlands, 2015, 22 Feb. "The Fall of the Female Fund Manager", Financial Times

[2]Forbes, "The Remarkable Rise Of ESG", URL: https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/#7234676b1695.

investment decision making. Croson and Gneezy (2009) summarize the three main aspects of differences in gender: (1) emotion, (2) preferences in risk-taking, and (3) overconfidence. These aspects are linked: (1) women react with fear when faced with dangerous or stressful situations while men react with anger ((see Brody, 1993; Fujita et al., 1991)). (2) Their emotions with respect to danger or stress causes women to avoid risks, say when making investments in mutual funds (see national survey by Dwyer et al. (2002)), while men will view risk as a challenge and prefer to engage or participate ((Byrnes et al., 1999; Schubert et al., 1999)). (3) While Niederle and Vesterlund (2007) and Soll and Klayman (2004) find both genders to be overconfident, they find men to be more overconfident than women. (Soll and Klayman, 2004) for example finds that when making investment decisions under uncertainty, men are more likely to overestimate the return on investment as compared to women.

However the earlier strand of literature faces the criticism that female professionals and leaders most likely differ from the average females in terms of all three aspects of gender differences. Another strand of literature targets the professional financial environment and generally show that the investment decision made by individual managers is not affected by their gender. For example, Atkinson et al. (2003) study the fixed-income mutual fund performance between genders and find no gender difference in their returns and risk taking. However, they observed that investor's inflow to female-managed funds is lower than male-managed funds, on average, by 40% in 1990s. Bliss and Potter (2002) examine the equity mutual funds (both U.S. domestic and international funds) in the 1990s and find that female fund managers outperformed male fund managers. However, after controlling for risk and other fund characteristics, there is no gender difference for domestic funds. In a later study, Babalos et al. (2015) looked at quantile regressions in single-managed European equity mutual funds and find no performance and risk difference between male and female-managed funds, as consistent with prior studies. However, they find that female managers shows perverse market timing skill when funds perform poorly.

Our paper contributes to the current understanding of the later strand of literature targeting the effects of gender in the professional financial environment by linking it to a growing set of literature on mutual fund's social responsible investment. Several studies have explored the performance between SRI and non-SRI (conventional) funds, and the results are mixed. Kreander et al. (2005) study European mutual funds from 1995 to 2001 and find no difference in risk-adjusted performance between SRI and non-SRI funds. Later studies show that SRI mutual funds underperform conventional funds and argue that investors have to pay a additional cost for SRI. For example, Renneboog et al. (2008) compare SRI with non-SRI funds globally and find that SRI funds in several regions underperform their corresponding benchmark performances, which support the "investor's paying for ethics" thesis. They find that the underperformance is not driven by an ethics style factor but by other factors such as limited stock selection after removing sinful stocks. Both Munoz et al. (2014) and Belghitar et al. (2014) also document the underperformce of SRI funds. Munoz et al. (2014) looked at U.S. and European green mutual funds and find that SRI funds underperform in non-crisis periods but no statistical significance was found in the crisis period. Belghitar et al. (2014) came to similar conclusions by using the Marginal Conditional Stochastic Dominance to accommodate the non-normal distributions of the returns. Besides fund performance, other studies show that SRI investors are less sensitive to mutual fund performance (see, e.g. Benson and Humphrey, 2008; Bollen, 2007; El Ghoul and Karoui, 2017; Renneboog et al., 2011). Bollen (2007) also finds that investors to SRI funds are more sensitive to past positive return and less sensitive to past negative return.

Our work impacts the understanding of how gender diversity should be considered when hiring mutual fund managers or when investors choose mutual funds. We seek to quantify if there is a fund performance value in gender diversity, and if this value is dependent on economic conditions by using a panel threshold regression model. We also quantify the effect of gender diversity on the SRI scores of investments a mutual fund makes and further condition it on fund performance and fund flows. We find evidence of risk-adjusted underperformance of SRI funds as in Munoz et al. (2014) and Belghitar et al. (2014), and a nonlinear effect of increasing gender diversity on fund performance and SRI investment scores. These show that for low levels of gender diversity, the effect of female fund managers reduces fund performance and increases SRI scores. At more than 30% females, the effect of additional female fund managers become positive, but the marginal effect on SRI is also negative. Our panel threshold model further shows that the effect of female fund managers is higher during poorer economic conditions. A further conditional estimation showed that female fund managers tend to invest in higher SRI portfolios even in low fund performance or fund flows, which explains why female fund managers may reduce fund performance. All our findings are conditioned on fund manager age, education and experience.

The paper is organized as follows: Section 2 describes our hypotheses, regression models and the threshold panel regression model; Section 3 describes how we construct the dataset and gives the descriptive statistics of the data. Section 4 reports our estimation results and Section 5 concludes.

2 Hypotheses and Econometric Models

Studies about gender differences in risk-taking and preferences have been made in economics and experimental psychology, with evidence found that compared to men, women are more risk-averse (Schubert et al. (1999); Byrnes et al. (1999); Holt and Laury (2002); Hartog et al. (2002); Croson and Gneezy (2009)). However, the average female mutual fund manager may not necessarily have the same characteristics as the average female. Adams and Funk (2012) finds that women in boardrooms tend to be more risk-loving than men, and thus do not find evidence that gender diversity would contribute to more risk-aversion in decisions-making. Similarly, Sila et al. (2016) (Sila et al., 2016) do not find any effect of gender diversity in boards on firm equity risk.

The gender effect of mutual fund managers have also been the topic of several studies. Beckmann and Menkhoff (2008) analyzed survey responses and find that female fund managers are more risk averse and less confident than men. Qiong et al. (2014) and Bär et al. (2011) find that gender diverse teams negatively impact mutual fund performance. For single manager funds, Atkinson et al. (2003) and Niessen-Ruenzi and Ruenzi (2017) find female fund managers generate significantly lower inflows than male managers but do not find evidence of differences in risk-adjusted performance. We further explore this subject of area considering more complex situation of fund manager teams, which means other than the gender aspect, we have to consider how business decisions are made in teams[3]. Working in teams have been found to increase diversification of opinions through compromise and hence decision-making reflects the average-opinions made by the team (Sah and Stiglitz (1986), Sah and Stiglitz (1988)). Another competing effect is group-shift, where opinions of a team gravitates towards that of the dominant person(s) in the team (Moscovici and Zavalloni (1986),Hogg et al. (1990),Kerr (1992)). To allow for the two competing effects, we include a (nonlinear) squared term for gender composition variable in the regressions.

Our first set of hypotheses on mutual fund teams is focus on gender-diversity impact on fund performance and risk, allowing the possibility of team characteristics such as group-shift and diversification of opinions. We first formulate two hypotheses regarding the gender diversity effect on fund performance:

> *H1: Mutual fund teams' gender diversity has a linear relationship with fund performance, which means the gender composition has a nonlinear effect with performance.*

The effect of gender composition might be nonlinear due to the two competing effects of teams: group shift and diversification of opinions. At low levels of gender diversity, group shift effects might dominate, while at higher levels, diversification of opinions might dominate. Furthermore, several literature investigating firm gender diversity find a nonlinear impact from gender composition (see, e.g. Torchia et al., 2011; Joecks et al., 2013, for critical mass effect). Hence we hypothesize that the impact of gender composition in mutual fund industry is also nonlinear.

We use a pooled regression model to test the gender diversity effect on fund performances:

$$Y_{it} = \beta_0 + \beta_1 Fund_{i,t-1} + \beta_2 Female_{i,t-1} + \beta_3 Female^2_{i,t-1} + \varepsilon_{i,t-1} \tag{1}$$

where $Fund_{t-1}$ are the lagged fund characteristics including Fund Size (defined as logged total net assets, TNA), Fund Flows[4], Expense Ratio and Portfolio Turnover,

[3]We account for the fact that working in teams is different from working as single fund managers, but do not compare them. For such comparison, see Bär et al. (2011)

[4]We define fund flow as $FF_t = \frac{TNA_{t+1}}{1+r_{t+1}} - TNA_t$, following Rakowski (2010).

Fund Age and dummy variables for Fund Family funds[5], SRI funds, and investment style. $Female$ is the percentage of female fund managers in the management team. We also define an entropy-based gender diversity measure as in Bär et al. (2009), Jehn et al. (1999) and Teachman (1980):

$$Diversity = \sum_{i=1}^{2} -p_i \ln(p_i) \tag{2}$$

where i are categories male and female[6], and p_i is the proportion of team members that are male or female. In cases where the team comprises of only one gender, $Diversity$ is defined to be zero.

For dependent variables, we consider raw, excess returns as well as three risk-adjusted return measures: Jensen's alpha, Fama French alpha and Carhart Four-factor alpha. The Carhart (1997) four factor model for each fund i and year t is given by:

$$r_{i,m,t}-r_{f,m,t} = a_{i,t}+\beta_{1,i,t}(r_{M,m,t}-f_{f,m,t})+\beta_{2,i,t}SMB_{m,t}+\beta_{3,i,t}HML_{m,t}+\beta_{4,i,t}MOM_{m,t}+\epsilon_{i,m,t} \tag{3}$$

where $r_{i,m,t}$ is the monthly raw returns of fund i for month m and year t, $r_{M,m,t}$ the monthly market portfolio returns and $r_{f,m,t}$ the risk free rate. The dependent variable is hence the monthly excess returns for fund i in the period. $a_{i,t}$ is the risk-adjusted returns. Jensen's alpha refers to the CAPM alpha, where only one factor, excess market returns, is used in the model. Fama French alpha includes two additional factors: size factor SMB and value factor HML. Carhart four factor model includes the momentum factor MOM. The factor loadings, βs, are also informative. High/low factor loadings indicate specific strategy styles of the mutual fund team, e.g. a high β_1 is indicative of investment style with higher levels market risk.

We also consider two measures of market timing ability: Treynor-Mazuy ($T - M$) and Merton-Henriksson ($M - H$). Treynor and Mazuy (1966) market timing regression uses

$$r_{i,m,t} - r_{f,m,t} = a_{i,t} + \beta_{1,i,t}(r_{M,m,t} - f_{f,m,t}) + \gamma_{1,i,t}(r_{M,m,t} - f_{f,m,t})^2 + \epsilon_{i,m,t} \tag{4}$$

[5]The behavior of mutual fund families may differ from non-fund families due to differing incentives, e.g. "cross-fund subsidization". We define funds to belong to a fund family if the fund parent company has 5 funds and above.

[6]We assume just two genders for simplicity as our data does not contain LGBT information.

where coefficient $\gamma_{1,i}$ measures the ability of the team to increase/decrease beta when market is positive/negative. Merton and Henriksson (1981) proposed an alternative market timing measure using the regression

$$r_{i,m,t} - r_{f,m,t} = a_{i,t} + \beta_{1,i,t}(r_{M,m,t} - f_{f,m,t}) + \gamma_{1,i,t}[r_{M,m,t} - f_{f,m,t}]^{+} + \epsilon_{i,m,t} \tag{5}$$

where $[r_{M,m,t} - f_{f,m,t}]^{+}$ is defined as $max(0, r_{M,m,t} - f_{f,m,t})$. Here, a significantly positive $\gamma_{1,i,t}$ is indicative of a team's market timing ability, whereby the team would adjust its target beta when market forecast is optimistic.

Ferson and Schadt (1996) argue that *conditional* performance measures and market timing measures should be used to reduce confounding effects of time variation in risks and risk premiums onto these measures. Given a lagged public information variable Z_t, the conditional CAPM regression is

$$r_{i,m,t} - r_{f,m,t} = a_{i,t} + \beta_{1,i,t}(r_{M,m,t} - f_{f,m,t}) + \delta_{1,i,t}(r_{M,m,t} - f_{f,m,t})z_{m,t-1} + \epsilon_{i,m,t} \tag{6}$$

where $z_t = Z_t - E(Z)$ is the lagged demeaned conditional variable, $a_{i,t}$ the conditional Jensen's alpha and $\beta_{1,i,t}$ is the 'average beta'. We use the demeaned lagged S&P500 returns and VIX as conditioning variables in our estimations.

Similarly, conditional market timing measures are estimated by including the interaction term (dynamic factor). For Treynor-Mazuy (TM), we have

$$r_{i,m,t} - r_{f,m,t} = a_{i,t} + \beta_{1,i,t}(r_{M,m,t} - f_{f,m,t}) + \delta_{1,i,t}(r_{M,m,t} - f_{f,m,t})z_{m,t-1} + \gamma_{1,i,t}(r_{M,m,t} - f_{f,m,t})^2 + \epsilon_{i,m,t} \tag{7}$$

where $\gamma_{1,i,t}$ is the conditional TM measure that captures the team's market timing ability to *private* market signals (since the term $\delta_{1,i,t}(r_{M,m,t} - f_{f,m,t})z_{m,t-1}$ controls for public information effects).

The conditional Merton-Henriksson ($M - H$) is given by

$$r_{i,m,t} - r_{f,m,t} = \beta_{1,i,t}(r_{M,m,t} - f_{f,m,t}) + \delta_{1,i,t}(r_{M,m,t} - f_{f,m,t})z_{m,t-1} + \gamma_{1,i,t}r^{*}_{M,m,t} + \gamma_{2,i,t}(r^{*}_{M,m,t}z_{t-1}) + \epsilon_{i,m,t} \tag{8}$$

where $r^{*}_{M,m,t} = (r_{M,m,t} - f_{f,m,t})I((r_{M,m,t} - f_{f,m,t}) - E(r_{M,m,t} - f_{f,m,t})|Z_{t-1}) > 0)$. Here, $I(\cdot)$ is the indicator function. Positively significant $\gamma_1 + \gamma_2 z_{t-1}$ indicates superior market timing ability in a conditional setting, and this is our conditional MH measure.

H2: Changing the fund gender composition/diversity affects fund performance in the next period. This effect differs depending on market conditions.

We further explore the effect of fund gender composition and diversity in a panel regression with fund fixed effect. By using within-transformation, confounding factors that are constant over time within each fund is eliminated, which in the case of the gender variable, means that the effect of changes in gender composition of the fund is captured.

Furthermore, we apply the panel regression in a conditional threshold setting to allow the magnitude of gender diversity to differ in different market conditions. We use the methodology of Hansen (1999) to estimate the model in a panel setting with fixed effects, where the effects of a variable changes when a conditioning variable crosses one or multiple thresholds. To select the number of thresholds, we follow Gonzalo and Pitarakis (2002) for model selection in the case of multiple unknown regimes. We will describe the model in detail in section 2.1Threshold panel regressionssubsection.2.1.

Our conjecture is that a more gender-diverse mutual fund is more conservative when facing market uncertainty. However, some existing literature suggests that female fund managers do not have different risk appetite from male managers (e.g. Atkinson et al., 2003).

We next formulate our hypothesis regarding mutual fund gender composition and fund's CSR investment decision:

H3: Gender diversity of mutual funds' management team and the CSR of their investment portfolios are positively associated, which means the effect of gender composition is nonlinear on the CSR of the investment portfolio.

Using each fund's detailed monthly equity investment holdings information, we match them to the KLD scores of the investment. We then use a representative score for the year to obtain an annual measure of the CSR score. This score has seven components to it: Environment, Product, Corporate Governance, Community, Diversity, Employee Relations and Human Rights (see section 3Datasection.3 for a detailed description of what these entail). Each component scores can then further be broken down into 'strength' and 'weakness'. We compute the overall CSR score as an average of all seven category scores, as well as overall CSR strengths as average of all seven category strength scores, and overall CSR weakness as the average of seven category weakness scores. We then regress these different scores on fund and fund manager characteristics (including fund manager age, experience, tenure, and highest education level) as well as gender diversity measure, $Female$, to measure the effect of gender diversity on different aspects of CSR investments.

We also estimate the gender diversity and CSR investment on fund performance using a pooled regression, including the separate and cross effects of CSR and gender diversity as explanatory variables.

> *H4: Changing gender composition/diversity affects CSR of the investment portfolio. This effect depends on individual fund as well as overall market conditions.*

Here, we again explore the effect of fund gender composition and diversity on CSR in a panel regression with fund fixed effect, which means that in a within-transformation, the effect of changes in gender composition of the fund on CSR is captured. We do not pose specific direction regarding the magnitude of the gender-CSR effect. It could be highly plausible that the CSR aspect of investment decisions are affected by gender in a non-linear manner (e.g. too high a female ratio may be as detrimental as too low). We further use to threshold panel regression model to test if the effect of gender composition varies according to economic conditions.

2.1 Threshold panel regressions

For the conditional regressions in *H2* and *H4*, we use Hansen (1999)'s threshold panel regression models with fixed effects which divides individual observation into discrete classes depending on the value of a conditioning variable. A simple threshold model with only one threshold could be given as:

$$Y_{i,t} = \begin{cases} \beta_0 + \beta_{11} Fund_{i,t-1} + \beta_{21} Female_{i,t-1} + \beta_{31} Female^2_{i,t-1} + \beta_{41} D_{i,t-1} + \epsilon_{i,t-1} & \text{if } Z_t \leq \gamma \\ \beta_0 + \beta_{12} Fund_{i,t-1} + \beta_{22} Female_{i,t-1} + \beta_{32} Female^2_{i,t-1} + \beta_{42} D_{i,t-1} + \epsilon_{i,t-1} & \text{if } Z_t > \gamma \end{cases} \tag{9}$$

where $Fund_t$ are fund manager characteristics including Fund Size, Expense Ratio and Portfolio Turnover, D_t are dummy variables including dummies for fund styles, fund family and SRI funds, and Z_t is the threshold variable. By using fixed effects, it implies that the measured gender effect is due to a change in gender composition of the team, since within transformations for teams with static gender compositions will give a vector of zeros.

We consider both VIX and S&P500 as threshold variable Z_t to observe if the effect gender composition changes when conditioned on economic conditions. For a one-threshold model, the regression is separated into two regimes depending on whether Z_t is larger or smaller than threshold γ. For multiple thresholds, the threshold model can also be expressed as

$$Y_{i,t} = \sum_{j=1}^{m+1} \beta_j' X_t I(\gamma_{j-1} < Z_t \leq \gamma_j) + \epsilon_{i,t} \tag{10}$$

where X_t are the explanatory variables and $\gamma_1, \ldots, \gamma_m$ are the threshold parameters for m thresholds, with $\gamma_0 = -\infty$ and $\gamma_{m+1} = \infty$, and such that $\gamma_j \in \Gamma_m$. In the threshold model, the slope parameter estimates $\hat{\beta}_j$ are a function of the threshold estimates, i.e. $\hat{\beta}_j = \hat{\beta}_j(\hat{\gamma}_1, \ldots, \hat{\gamma}_m)$. Like in Hansen (1999) and Gonzalo and Pitarakis (2002), we estimate the threshold parameters by

$$(\hat{\gamma}_1, \ldots, \hat{\gamma}_m) = \arg\min S_T(\gamma_1, \ldots, \gamma_m) \tag{11}$$

where $S_T(\gamma_1, \ldots, \gamma_m)$ is the sum of squared errors of the regression. In practice, this minimization problem is done by searching over values of γ equalling to the distinct values of Z_t in the sample. It also requires imposing an ad-hoc lower bound for the number of observations in each regime. We set this at 5%, which eliminates the smallest and largest 5% of distinct threshold variables. The regression is estimated for 100 distinct quantiles of γ and the $\hat{\gamma}$ that gives the smallest value of S_T is chosen. We follow Hansen (1999) and use the likelihood ratio statistic to form the 99% confidence interval for $\hat{\gamma}$. Hansen (1999) propose using an F test to determine the statistical significance of a threshold and using a bootstrap procedure to obtain the critical values. This is computationally intensive so we follow Gonzalo and Pitarakis (2002) which propose the following model selection criteria to choose the number of thresholds

$$IC_T(\gamma_1, \ldots, \gamma_m) = \log S_T(\gamma_1, \ldots, \gamma_m) + \frac{\lambda_T}{T}(K(m+1)) \tag{12}$$

where λ_T is the penalty term and K is the number of regressors. For a one threshold model, a lower IC for the threshold model would mean rejecting a linear specification for the threshold model. Gonzalo and Pitarakis (2002) find BIC-type specifications ($\lambda_T = \log T$) for the penalty term to perform best in simulations.

To reduce computational load, we consider the one and two-threshold models. To estimate more than one threshold, we follow the procedure described in Hansen (1999). The procedure first involves deciding between a linear and one threshold model. If the data supports the threshold model, the estimate of the first threshold $\hat{\gamma}_1$ is obtained. Fixing the first-stage estimate $\hat{\gamma}_1$, the second threshold is then estimated to minimize

$$S_2(\gamma_2) = \begin{cases} S(\hat{\gamma}_1, \gamma_2) & \text{if } \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & \text{if } \gamma_2 < \hat{\gamma}_1 \end{cases} \tag{13}$$

Here $\hat{\gamma}_2$ is efficient but $\hat{\gamma}_1$ is not because the estimate was obtained in the presence of a neglected threshold. Hence a refinement estimator is made by fixing $\hat{\gamma}_2$ and reestimating $\hat{\gamma}_1$. Finally, while $\hat{\beta}(\hat{\gamma})$ depends on $\hat{\gamma}$, we follow Hansen (1999)'s argument that this dependence is not of first-order asymptotic importance and so inference on β can be made as if $\hat{\gamma}$ is the true value.

3 Data

We obtain US equity mutual funds data mainly from three resources: (1) CRSP, (2) Morningstar, and (3) MSCI KLD in order to construct the measures used in our estimations. In this section, we describe in detail how we construct the fund CSR investment scores as well as the fund manager details and fund characteristics. We finally give the descriptive statistics of our data.

3.1 Constructing the fund SRI investment scores

Over 31,000 Survivor-Bias-Free US open-end equity funds (of various share classes) have been obtained from WRDS (Wharton Research Data Services) CRSP (Center for Research in Security Prices) database for the sample period from 2002 to 2015. The key data items of each equity fund share class include the asset holdings, the net asset value (NAV) and the percentage weight of the NAV for each asset held in each reported period. Our sample contains more than 244.7 million number of asset holdings. There are 8,876 unique equity funds in our samples after combining different classes of the same fund.

We calculate the fund SRI score by linking the KLD scores of its holding assets. The MSCI ESG KLD Stats data, obtained from WRDS for the sample period of 2002 to 2015, provides information on how corporate social responsible a firm is. We obtain over 43,000 firm-year observations after we match mutual fund's holding and the KLD dataset.[7] The KLD stats data for each firm-year include 159 variables, namely "Controversial Business Involvement", and the strength and the concern indicators for "Governance", five subcategories of "Social" (i.e. "Community", "Diversity", "Employee Relations", "Human Right", and "Product"), and "Environment". The mutual fund SRI performance is measured by the value weighted KLD scores of firms held by the mutual fund. First, we note that comparing the KLD scores between firms and between years is impossible because the number of strength and concern indicators of each (sub)category vary cross-sectionally and over time. To facilitate comparison, we normalize the strength and concern scores of each (sub)category by the number of indicators for each firm year so that it falls within the 0-100% range. We also calculate an aggregated strength and aggregated concern score by summing up the normalized dimensional scores across dimensions. A net score is the difference between the strength score and the concern score for each dimension and a net aggregate score is the difference between the aggregated strength score and aggregated concern score.

[7] While our research focuses on US equity funds, these funds may invest in non-US equity that are listed or traded in the US, and MSCI does provide KLD scores for some of these non-US companies. The holding data from CRSP include company names, CUSIP for US firms and CINS for non-US firms. In the MSCI KLD database, company names and CUSIP are provided for US firms, ADR and North American companies (including Canadian firms), but company names only for non-US firms. Since a firm can potentially have many traded common equity in the form of ADR, overseas listed stocks, and international company typically have multiple security codes (CUSIP, ADR, CINS) that we have to trace to the same firm. We searched and found about 4,000 KLD international companies' CINS & ADR CUSIP codes through Historical SEC NP-X filings, and through manually checks via Bloomberg. A small number of unmatched stocks are discarded.

Since it is well known that CSR performance is strongly influenced by the industry sector, we also scale the company CSR performance by the industry's mean and standard deviation. The Standard Industry Classification (SIC) Codes for all the KLD constituents have been obtained from the WRDS COMPUSTAT database. Table 1 (not yet included) shows the number and distribution of firms held by the KLD constituents and their 1-digit top level SIC industry classification. It is clear that the sample firms are not evenly distributed with "Manufacturing" representing about one-third of the sample while "Public Administration" has the smallest number of companies for less than half of the sample period. The standard deviation for KLD scores is undefined and can be very unstable for a sector that has zero or very few firms. Hence, we created two variants of industry adjusted KLD score; the first is de-mean only, while the second is de-mean and scaled by the standard deviation. We set all standard deviations that are less than 1 (including zero) to equal to 1, which effectively means that the KLD score is not scaled if the standard deviation is less than 1.[8] We discard all funds where the KLD scores represent less than 90% of the funds' holding.

3.2 Constructing fund and fund manager characteristics

Mutual fund manager information is obtained from Morningstar for 5633 US equity mutual funds, of which 3394 funds have both fund manager information and Nasdaq tickers, which we use to match the information in CRSP. We can access two data-sets: one provides fund-manager tenure records, consisting of Morningstar fund identifier, manager identifier, manager starting date, and tenure end date for that fund; the other provides manager traits including birth year, level of education and

[8]Due to imperfect data quality, the total value weight of KLD score for each mutual fund varies cross funds and years. We have removed few cases of negative total weight as this representing the extreme case of short selling. We have also capped total KLD weight at 105% to remove firms that are heavily leveraged. The total weight is not capped at 100% due to rounding. The holding data appear on a monthly or a quarterly basis, but often with lots of variations and missing gaps between months/quarters. Since KLD is an annual data, we use the holding information that has the highest valid equity holding weight in the year to calculate weighted KLD score for the fund to make the score more representative of the fund's characteristics by assuming that the asset weight and KLD scores do not change drastically quickly.

professional qualification. Since these two sets of Morningstar data are static and received in July 2016, we assume fund manager's personal traits do not change throughout our sample period. In addition, if fund manager's end-of-the-tenure date is empty (meaning that the manager was still working in that fund), then we manually set that the tenure ends in July 2016. Finally, we identify the gender of the fund manager using proprietary data from Morningstar.

One manager can involve in multiple funds at the same time; likewise, one mutual fund can be managed by more than one manager. We identify the number of funds that are managed by a single manager and those by a team at each point in time, and only retain funds managed by teams. Funds may sometimes have singly managed and team managed at different points in time, and we retain only those observations when it is team-managed. For each month-fund observation, we locate the corresponding managers and construct the average managers' traits within that team, e.g. average age, proportion of gender, percentage of MA/MBA holders.

Finally, we then merge our CRSP and Morningstar data-sets by Nasdaq ticker. Mutual fund monthly net-of-fee return and fund's quarterly summary statistics e.g. management fee, fund style, and inception date are obtained from CRSP from Jan 2002-Dec 2015. Using the CRSP style codes[9], we identify 5 separate mutual fund styles for 2557 mutual funds. These comprise of "Growth" (1663 funds), "Growth and Income" (603 funds), "Hedged" (64 funds), "Short" (4 funds) and "Income" (223 funds). Due to the small sample in "Short", we disregard this style which leaves us with a sample of 2553 mutual funds.

Among our data-sets, the most granular frequency is on monthly basis. However, some of the data is available only on quarter/year-end. We then assume the information does not change over the quarter (year). Again, for static information, we assume it does not change over the sample period. We detail the mutual fund variables description and their sources in Table 9Variable Descriptionstable.caption.14 in

[9]CRSP identifies 5 separate mutual fund styles: Growth ("EDYG"), Growth and Income ("EDYB"), Hedged ("EDYH"), Short ("EDYS") and Income ("EDYI"). Under the Lipper code, the "Growth" style corresponds to Capital Appreciation and Growth funds; "Hedged" include the long/short equity funds, equity market neutral funds, absolute return funds and equity leverage funds; "Short" corresponds to Dedicated Short Bias Funds and "Income" to Equity Income Funds.

the Appendix.

3.3 Descriptive statistics

The descriptive statistics of our fund characteristics and fund manager characteristics is given in Table 2Empirical means of computed fund SRI scores based on investment holdings of the fund, divided into fund-family, non fund-family, by strategy types as well as by number of female fund managers in team. The computed scores have seven categories where each category score is computed as category strength minus category weakness. Overall CSR scores (sum of all categories), as well as overall CSR strengths and CRS weaknesses are computed as sum of all category strengths and category weakness respectively. Empirical standard deviations are reported in brackets.table.caption.2. Here, we exclude funds with teams of more than 10 managers as well as funds with observation lengths of less than 36 months (these are too short for constructing meaningful risk-adjusted performance measures). The funds are divided into fund and non-fund family, as well as by investment styles (sector-based funds are removed due to their small numbers), which leaves us with 2215 funds.

Table 2Empirical means of computed fund SRI scores based on investment holdings of the fund, divided into fund-family, non fund-family, by strategy types as well as by number of female fund managers in team. The computed scores have seven categories where each category score is computed as category strength minus category weakness. Overall CSR scores (sum of all categories), as well as overall CSR strengths and CRS weaknesses are computed as sum of all category strengths and category weakness respectively. Empirical standard deviations are reported in brackets.table.caption.2 shows that funds in fund families tend to be larger in terms of logged total net assets, older in terms of fund-year, but have lower mean returns and larger unconditional volatility[10] Interestingly, teams in fund families tend to have higher gender diversity as well as younger fund managers. While fund managers in fund families tend to have higher educational qualifications, they tend to have less

[10]This phenomenon that funds in mutual fund families engage in higher risks due to risk sharing benefits have been well document, see for example Gaspar et al. (2006).

work experience as compared to those in non-fund families.

We also split the sample up by the number of females in each team and compute their descriptive statistics. As compared to teams with no females, teams with one females have lower mean fund returns but the returns also have lower standard deviations and kurtosis. This is also true for teams with two females. With three or more females, mean returns are higher, however the sample size in this group is small. Overall, the larger the team is, the more females they tend to have. Teams with a female tend to be older and more experienced than those without females.

Table 2Empirical means of computed fund SRI scores based on investment holdings of the fund, divided into fund-family, non fund-family, by strategy types as well as by number of female fund managers in team. The computed scores have seven categories where each category score is computed as category strength minus category weakness. Overall CSR scores (sum of all categories), as well as overall CSR strengths and CRS weaknesses are computed as sum of all category strengths and category weakness respectively. Empirical standard deviations are reported in brackets.table.caption.2 gives the empirical means and standard deviations of the computed CSR scores of the funds by fund-year. The results are reported for subcategories of fund and non-fund family as well as by "strategy" and "capital funds". The mean scores of each of the seven categories are computed as means of $Strengths-Weakness$. The averages of overall strengths in the seven categories are reported under $CSRStrengths$ and average overall weakness as $CSRWeakness$. $CSRall$ gives the overall averages of the category scores. We also report the percentages of self-reported SRI funds in the sample. Here we see that most self-reported SRI funds fall in the category of non-fund family and as "strategy" funds. We also see that "strategy" funds tend to have much better CSR scores than "capital" funds in terms of overall CSR and CSR strengths. They however also have high "weakness" as compared to "capital" funds. Interestingly, funds in fund families also tend to have better overall CSR scores. In specific categories of "Environment", "Product", "Corporate Governance" and "Human Rights", non-fund families have better scores on average.

We further split the sample by number of female fund managers in the team. We find that teams with one female perform much better in overall CSR than those without females. For teams with two or more females, the scores are lower than those an all-male team. In terms of categories, the improvements with adding one female is highest for "Human Rights" and "Environment".

Table 1: Descriptive statistics (yearly) of the sample of 2215 US equity mutual funds† with team fund managers for the period 01.2002-12.2015, divided into those in fund families and non-fund families, and subdivided into two major styles: "Strategy-based" and "Capital-based". We also report the statistics based on number of female managers in the team.

	Fund Family		Non- Fund Family		No of females in team					
	Strategy	Capital	Strategy	Capital	All	0	1	2	3	>3
No of Funds	1025	711	399	170	2215	1881	661	126	26	6
Obs (fund-year)	7411	5542	2548	1134	16635	11993	2992	335	90	23
Avg Fund Age	16.20	12.73	14.20	11.69	14.43	14.20	15.51	13.07	11.63	9.138
Avg LTNA	5.167	4.962	4.433	4.534	4.947	4.927	5.055	5.153	3.665	5.825
Avg FundFlow	.0491	.0236	.0088	.0298	.0333	.0410	.0169	.0171	.0087	.0177
Avg Exp Ratio	.0108	.0117	.0129	.0130	.0116	.0117	.0112	.0112	.0118	.0106
Avg Turnover	.7990	.9286	.6238	.7707	.8147	.8276	.7011	1.046	.9470	.7910
Mean Returns	.0828	.0917	.0846	.1095	.0879	.0891	.0841	.0722	.0954	.0967
Std Dev Returns	.1976	.2220	.1938	.2213	.2073	.2073	.2043	.2126	.2134	.1451
Skew. Returns	-.7580	-.5102	-.6134	-.1817	-.5838	-.5570	-.5956	-.6358	-.3812	-1.321
Kurt. Returns	4.012	3.283	4.020	3.836	3.746	3.771	3.750	3.255	3.674	5.146
Avg team size	3.079	3.006	2.996	3.231	3.052	2.834	3.563	5.204	5.866	7.895
Avg %female	9.998	10.53	6.072	6.041	9.303	0.000	34.86	44.55	59.04	50.70
Avg Mgr Age	52.31	50.66	54.02	51.53	52.06	51.64	53.53	55.81	49.69	43.25
Avg Mgr Exp	8.513	7.909	9.057	9.682	8.475	8.514	8.571	7.655	8.258	4.622
Avg Mgr Tenure	5.022	4.792	6.360	6.149	5.227	5.326	5.618	4.785	4.863	2.876
% BA	83.12	84.04	77.82	83.12	82.61	83.42	82.06	82.64	63.81	87.83
% MA	14.49	12.20	11.68	12.17	12.14	13.27	12.53	16.74	11.36	10.37
% MBA	51.79	49.12	42.67	51.81	49.51	49.88	49.81	46.55	38.64	72.65
% PHD	3.756	2.890	1.886	2.411	3.089	3.044	3.337	3.194	3.146	1.331
% CFA	54.84	57.48	44.41	53.16	54.01	54.91	51.72	53.94	53.34	38.24

†We disregard sector-based funds due to their small numbers (34 funds). We also exclude funds with more than 10 managers (15 funds) and those with observations lengths of less than 36 months (773 funds).

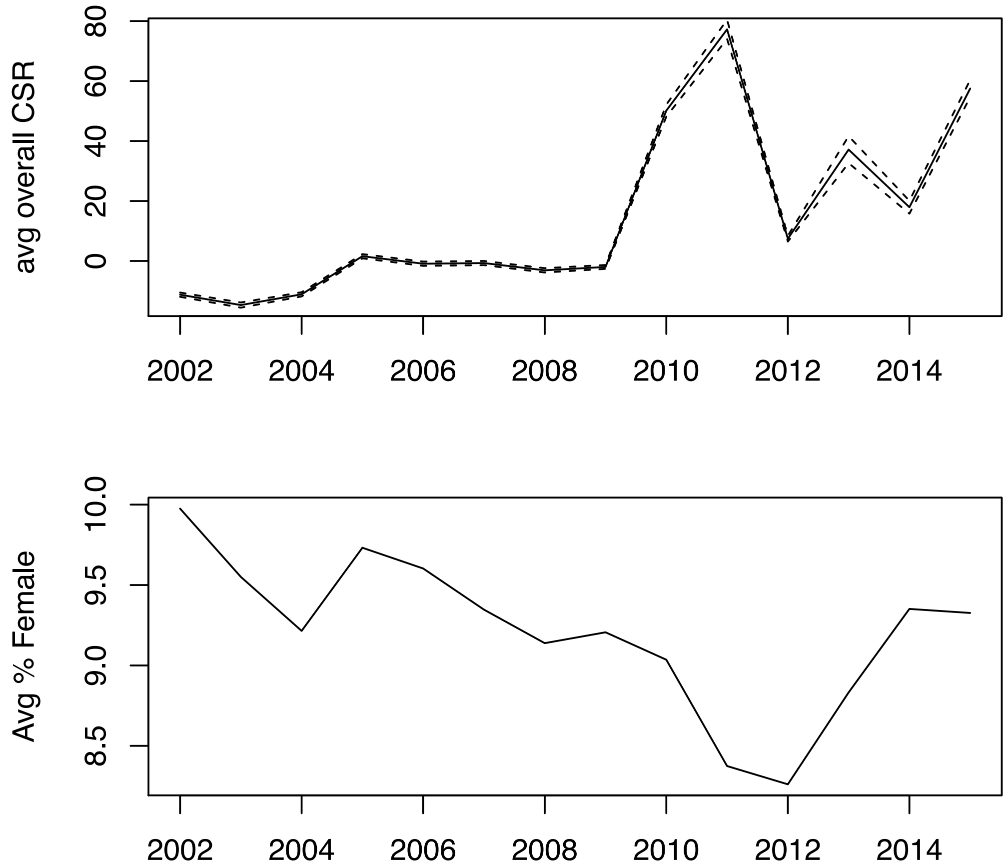
Table 2: Empirical means of computed fund SRI scores based on investment holdings of the fund, divided into fund-family, non fund-family, by strategy types as well as by number of female fund managers in team. The computed scores have seven categories where each category score is computed as category strength minus category weakness. Overall CSR scores (sum of all categories), as well as overall CSR strengths and CRS weaknesses are computed as sum of all category strengths and category weakness respectively. Empirical standard deviations are reported in brackets.

	Fund Family		Non- Fund Family		No of females in team					
	Strategy	Capital	Strategy	Capital	All	0	1	2	3	>3
% SRI funds	3.40	1.46	6.12	4.59	3.25	2.94	3.88	4.78	4.44	0.00
					Overall scores					
CSRall	39.55	-4.016	38.28	-8.955	21.54	21.25	24.60	9.589	17.26	5.749
	(58.39)	(36.10)	(60.00)	(35.54)	(55.34)	(55.16)	(57.00)	(49.53)	(49.90)	(44.21)
CSR Strength	84.36	-.5186	74.76	-8.190	48.27	47.22	52.67	36.90	38.74	67.89
	(62.57)	(44.18)	(67.18)	(39.01)	(69.93)	(68.99)	(73.23)	(67.44)	(66.62)	(41.90)
CSR Weakness	44.82	3.498	36.48	.7653	26.73	25.96	28.07	27.31	21.48	62.14
	(33.40)	(20.90)	(31.81)	(15.41)	(34.60)	(33.76)	(36.15)	(38.66)	(36.39)	(36.49)
					Category scores					
Environment	8.188	-.6808	8.290	-1.546	4.591	4.542	5.132	2.842	2.877	1.120
	(11.68)	(7.006)	(11.74)	(6.396)	(10.99)	(10.94)	(11.32)	(10.43)	(9.882)	(8.942)
Product	-5.668	-.9120	-3.636	-.6308	-3.415	-3.378	-2.970	-4.904	-3.034	-14.03
	(13.35)	(6.268)	(12.44)	(6.209)	(11.12)	(11.03)	(11.32)	(11.08)	(9.898)	(15.82)
Corporate Gov	3.722	-.8165	4.144	-1.324	1.937	1.926	2.740	-.3966	2.187	-1.803
	(17.69)	(8.766)	(17.64)	(8.688)	(14.95)	(14.95)	(15.34)	(13.21)	(13.32)	(13.84)
Community	3.637	-1.108	3.472	-1.228	1.702	1.755	1.670	-.0389	.5183	-.6933
	(8.047)	(5.734)	(9.015)	(6.814)	(7.810)	(7.876)	(7.666)	(6.802)	(6.875)	(2.995)
Diversity	22.91	2.158	19.72	-.0756	13.92	13.65	14.60	12.56	12.94	18.07
	(18.37)	(12.56)	(19.09)	(11.63)	(19.20)	(18.93)	(19.94)	(18.78)	(19.40)	(11.77)
Employee	4.522	-.0722	3.849	-.5212	2.541	2.503	2.701	1.485	1.199	4.561
	(7.464)	(4.575)	(7.537)	(4.763)	(6.856)	(6.834)	(6.969)	(6.713)	(6.345)	(5.536)
Human Rights	2.237	-2.584	2.439	-3.629	.2641	.2611	.7275	-1.960	.5711	-1.474
	(12.29)	(8.578)	(12.65)	(9.337)	(11.34)	(11.43)	(11.45)	(9.473)	(13.07)	(10.93)

Finally, Figure **??** (top) shows the time evolution of overall CSR scores of the 2215 mutual funds though our sample period. The dashed lines indicate the 95% confidence intervals. Overall CSR scores have improved substantially from 2009 onwards, with a peak in 2011. An explanation is that after the credit crises, there is a shift towards social responsible investments as an alternative fund strategy. Another explanation is greater social awareness of fund managers and investors.

Figure **??** (bottom) shows the average percentages of females in mutual fund teams. As noted by several sources[11], this has been falling steadily in our sample period, with the lowest trough in 2012.

Figure 1: Yearly average overall CSR scores of mutual fund equity investments and average percentage of female mutual fund managers. Dash lines indicate the 95% confidence interval.



4 Estimation Results

Table 3Pooled regressions of unconditional and conditional performance measures against fund manager characteristics using 14285 fund-year observations (excluding

[11]For example, "Chris Newlands, 2015, 22 Feb. "The Fall of the Female Fund Manager", Financial Times

funds with less than 36 month observations). $Female$ refers to the percentage of managers that are female in the team. TM and MH refers to the Treynor-Mazuy and Merton-Henriksson market timing models respectively. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level and † indicates significance at 10% level. table.caption.3 reports the pooled regression results of various measures of unconditional and conditional fund performance on fund and fund manager characteristics. $Experience$ tends to be positive and statistically significant while larger funds have negative impact on fund performance and positive impact on risk-adjusted fund performance. We find that the percentage of females in the team, $Female$, and its square term tends to be statistically significant. Not including the squared nonlinear term renders $Female$ to be statistically insignificant, as found by numerous other papers (e.g. Atkinson et al. (2003), Bliss and Potter (2002), Babalos et al. (2015)). Controlling for various other fund characteristics and fund manager characteristics, we find $Female$ to have a negative impact on fund performance. This is however nonlinear from the significant squared term and has a turning point at between 30-50 percent females. We find evidence that there is a U-shape relationship between fund's gender composition and fund's performance, since both $Female$ and $Female^2$ coefficients are significant in both measures. The effective turning point for the fund's excess return is approximately 40%, suggesting that the increasing performance starting from there (see Figure 1Nonlinear Gender-Return Relationshipfigure.caption.4). The results echo the findings in Joecks et al. (2013), where the authors argue a "critical mass" of female representatives to be reached, for which a more diverse board exceeds the one of a completely male board. On other hand, the critical mass for Jensen's alpha is larger. We repeat the regressions (Table 10Pooled regressions of unconditional and conditional performance measures against fund manager characteristics using 9161 fund-year observations (excluding funds with less than 36 month observations). $NFemale$ refers to number of female managers in the team. TM and MH refers to the Treynor-Mazuy and Merton-Henriksson market timing models respectively. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level and † indicates significance at 10% level. table.caption.15 in Appendix) using number of females $NFemales$ instead of percentage of females and find the same nonlinear effect. We confirm our first hypothesis that gender diversity has a nonlinear impact on performance.

Table 3: Pooled regressions of unconditional and conditional performance measures against fund manager characteristics using 14285 fund-year observations (excluding funds with less than 36 month observations). $Female$ refers to the percentage of managers that are female in the team. TM and MH refers to the Treynor-Mazuy and Merton-Henriksson market timing models respectively. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level and † indicates significance at 10% level.

	Returns		unconditional alpha			market timing		conditional alpha		conditional market timing			
	Raw	Excess	Jensen	Fama	Carhart	TM	MH	Z=SP500	Z=VIX	TM_{SP}	TM_{VIX}	MH_{SP}	MH_{VIX}
Intercept	11.75*	10.14*	-5.247*	-7.680*	-8.499*	-.3303*	-4.153*	-6.206*	-5.771*	-.3679*	-.1070	-2.673*	-1.713*
	(.5394)	(.5613)	(1.526)	(1.369)	(1.323)	(.0705)	(1.034)	(1.468)	(1.576)	(.0732)	(.0753)	(.8538)	(.8297)
Team Size	-.0204	.0428	.2565	.1692	.1258	.0140	.1710	.3326*	.3570*	.0142	.0074	.0633	-.0323
	(.0755)	(.0797)	(.1715)	(.1577)	(.1471)	(.0089)	(.1302)	(.1609)	(.1699)	(.0094)	(.0097)	(.1050)	(.1034)
Experience	.0568	.1507*	-.1622†	.1072	.2029*	.0165*	.2070*	-.0801	-.1271	.0142*	.0100*	.1205*	.0149
	(.0366)	(.0365)	(.0911)	(.0775)	(.0732)	(.0041)	(.0561)	(.0893)	(.0895)	(.0046)	(.0044)	(.0513)	(.0527)
Tenure	-.0222	-.0192	-.1432	-.2533*	-.2510*	.0046	.0074	-.2269*	-.1840†	.0034	.0026	-.1416*	-.0911
	(.0409)	(.0424)	(.1071)	(.0904)	(.0853)	(.0047)	(.0719)	(.1028)	(.1043)	(.0052)	(.0050)	(.0569)	(.0599)
Female	-.0361*	-.0344*	-.0666*	-.0653†	-.0448	-.0047†	-.0795*	-.0411	-.0614†	-.0048†	-.0035	-.0727*	-.0121
	(.0152)	(.0159)	(.0340)	(.0343)	(.0324)	(.0026)	(.0310)	(.0352)	(.0337)	(.0027)	(.0027)	(.0222)	(.0253)
Femalesq	.0006*	.0005*	.0008†	.0008	.0004	.0001†	.0013*	.0005	.0009†	.0001	.0001	.0007*	-.0001
	(.0003)	(.0003)	(.0005)	(.0005)	(.0005)	(.0000)	(.0006)	(.0005)	(.0005)	(.0001)	(.0001)	(.0003)	(.0005)
BA	.0071	.0063	.0342*	.0175†	.0191†	.0006	.0082	.0290*	.0291*	.0002	.0003	.0124†	.0104
	(.0043)	(.0046)	(.0116)	(.0104)	(.0101)	(.0006)	(.0088)	(.0108)	(.0118)	(.0006)	(.0006)	(.0065)	(.0065)
MA	.0004	.0000	.0098	-.0067	-.0222†	.0010	.0135	.0093	.0127	.0012†	.0009	.0042	-.0082
	(.0051)	(.0054)	(.0118)	(.0126)	(.0121)	(.0007)	(.0087)	(.0117)	(.0116)	(.0007)	(.0007)	(.0075)	(.0075)
MBA	.0023	.0008	-.0006	-.0010	-.0008	.0007	.0077	.0001	.0047	.0010*	.0003	.0061	-.0007
	(.0037)	(.0039)	(.0088)	(.0086)	(.0081)	(.0005)	(.0068)	(.0085)	(.0088)	(.0005)	(.0005)	(.0054)	(.0052)
PHD	.0030	.0027	.0103	.0128	.0159	-.0015	-.0293†	.0113	-.0036	-.0009	-.0018†	.0052	.0210
	(.0104)	(.0107)	(.0245)	(.0188)	(.0182)	(.0010)	(.0154)	(.0225)	(.0214)	(.0011)	(.0011)	(.0112)	(.0159)
CFA	-.0094*	-.0095*	-.0059	.0010	.0041	-.0006	-.0134*	-.0005	-.0050	-.0003	-.0008†	-.0031	-.0018
	(.0032)	(.0033)	(.0085)	(.0080)	(.0074)	(.0004)	(.0057)	(.0079)	(.0086)	(.0004)	(.0004)	(.0048)	(.0049)
Fund Size	-.3683*	-.4476*	.1858	.3360*	.2976*	-.0117	.1057	.2582†	.2343	-.0054	-.0125	.0438	.1349†
	(.0571)	(.0599)	(.1442)	(.1273)	(.1226)	(.0073)	(.1057)	(.1382)	(.1471)	(.0077)	(.0079)	(.0843)	(.0813)
Adj R-Squared	.0013	.0025	.0033	.0018	.0027	.0038	.0036	.0033	.0034	.0024	.0011	.0031	.0006
Observations	14285	14285	14285	14285	14285	14285	14285	14285	14285	14285	14285	14285	14285
Turn Pt(Female)	30.56	31.99	39.75	42.59	55.87	28.07	29.83	41.66	35.80	30.31	26.94	49.81	-

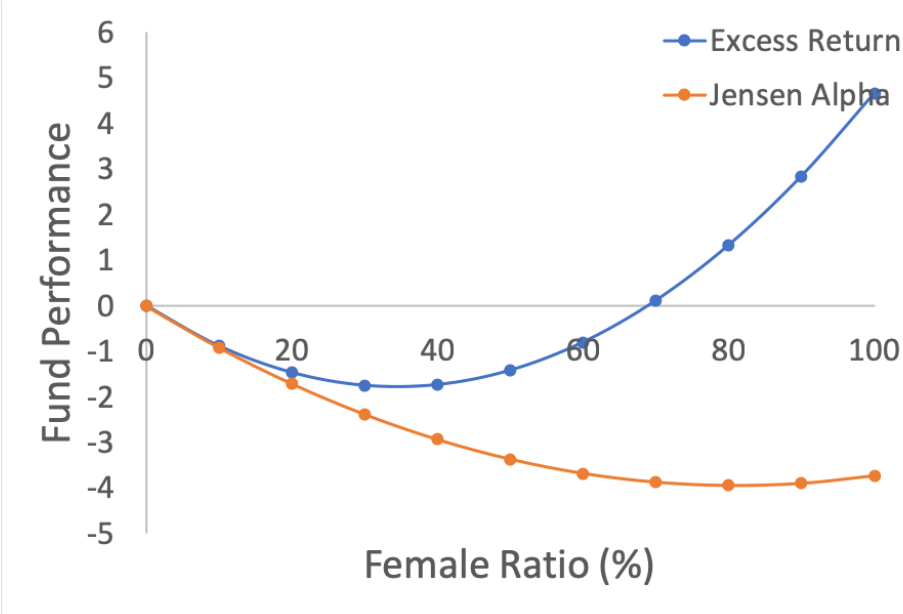


Figure 2: Nonlinear Gender-Return Relationship

Table 4Estimated parameters for the linear (0), 1 threshold (regimes=1A, 1B) and 2 threshold (regimes=2A, 2B, 2C) models conditioned on $Z = S\&P$ annual returns and annualized VIX. Dependent variables used are excess returns and Jensen's alpha (both one period ahead). Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level. 9 unique quantiles are used in the estimation with a trim ratio of 0.15. Estimated threshold is given by γ and the quantile of the estimated threshold $q(\gamma)$. Adjusted R^2 and BIC for each model is reported, where models with the lowest BIC are in bold. Estimated turning point for $Female(\%)$ in each regime is also given under $TurnPt$.table.caption.5 reports the gender diversity impact on fund's performance conditioned on economic conditions proxied using S&P500 index returns and VIX, the former representing the bullish or bearish of the stock market and the latter representing the uncertainty of the stock market (or investor's fear gauge). Two performance measures, excess return and Jensen's alpha, are used as dependent variables. The left (for S&P500) and the right (for VIX) panel report the threshold panel regression results. From the BICs, the one threshold model is selected for all the regressions, where the estimated thresholds are, $\hat{\gamma}$ is at 13.53% S&P returns and 1.86 annual VIX volatility, all close to the 45 percentiles, for both dependent variables. Overall, we find that the female ratio coefficients are significant in different market conditions. In addition, the coefficients show consistently that the gender diversity impact is more pronounced in a bearish market (lower S&P market returns, higher VIX volatility).

We now consider the interaction effects between gender diversity and CSR scores. Table 5Pooled regressions of raw and excess fund returns and risk-adjusted returns (Jensen's alpha and Fama French alpha) against fund characteristics (lagged at t-1): $FundSize$, measured by log of total net assets (LTNA), $FFlow$, fund flows, $Expense$-expense ratio (in %) , $Turnover$, portfolio turnover, $FundAge$, several dummy variables, D indicating SRI fund, Fund Family, and capital-style funds, overall CSR measures (all, strength and weakness) of the portfolio held, and percentage of female fund managers in team, as well as their interaction with CSR measures. Robust standard errors clustered by fund are given in brackets. $*$ indicates significance at 5% level or better and † indicates significance at 10% or better. table.caption.6 regress gender diversity, CSR scores (overall CSR, CSR strengths or CSR weakness) against fund returns, excess returns, risk-adjusted Jensen's alpha and Fama-French alpha. We find some evidence that improving CSR scores does significantly improve fund returns and excess returns, but this is not the case when using risk adjusted returns. It even negatively and significantly explains Fama-French alpha. Interestingly, investing in CSR strengths reduces fund performance consistently. Portfolios with high CSR weakness scores give highly negative returns. The gender composition terms are mainly negative but not significant except for Fama French alpha, where it is highly negative. The interaction terms between gender composition and CSR are mostly positive when significant. This means a positive relation between them and increasing gender diversity reduces the negative impact of social responsible investing.

We further confirm that gender diversity is positively related to CSR scores by directly regressing the percentage of females, $Female$, and its squared term against overall CSR, CSR strengths, CSR weakness and the seven components of the CSR scores. We report the results in Table 6Pooled regressions of CSR scores and its subcategories against fund and fund manager characteristics (using 5647 fund-year observations). Robust standard errors clustered by fund are given in brackets. $*$ indicates significance at 5% and † indicates significance at 10% level or better.table.caption.7.

Table 4: Estimated parameters for the linear (0), 1 threshold (regimes=1A, 1B) and 2 threshold (regimes=2A, 2B, 2C) models conditioned on $Z = S\&P$ annual returns and annualized VIX. Dependent variables used are excess returns and Jensen's alpha (both one period ahead). Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level. 9 unique quantiles are used in the estimation with a trim ratio of 0.15. Estimated threshold is given by γ and the quantile of the estimated threshold $q(\gamma)$. Adjusted R^2 and BIC for each model is reported, where models with the lowest BIC are in bold. Estimated turning point for $Female(\%)$ in each regime is also given under $TurnPt$.

	Dep var= excess returns						Dep var=Jensen alpha					
	0	1A	1B	2A	2B	2C	-	1A	1B	2A	2B	2C
						Z=S&P500 returns						
Female(%)	-.1034*	-.1444*	-.0685*	-.1439*	.0112	-.0724*	-.0973*	-.0978*	-.0856*	-.0971*	-.1623*	-.0820*
	(.0047)	(.0060)	(.0100)	(.0069)	(.0134)	(.0109)	(.0180)	(.0368)	(.0376)	(.0368)	(.0446)	(.0365)
Female^2	.0015*	.0021*	.0010*	.0022*	-.0008*	.0012*	.0006*	.0008*	.0002	.0008*	.0033	.0001
	(.0003)	(.0004)	(.0004)	(.0004)	(.0004)	(.0004)	(.0002)	(.0002)	(.0002)	(.0002)	(.0020)	(.0003)
Team Size	-.0115	-.2594	.0426	-.2078	-1.877*	.1130	-1.561*	-1.394	-1.666	-1.398	-1.021	-1.683
	(.2609)	(.6613)	(.6417)	(.6834)	(.7040)	(.6642)	(.4198)	(1.043)	(1.032)	(1.094)	(1.178)	(1.084)
Exper	.4900*	.3004*	.4299*	.3454*	.0558	.5134*	-1.026*	-.9033*	-1.063*	-.9018*	-1.456*	-1.055*
	(.0651)	(.1516)	(.1451)	(.1573)	(.1445)	(.1511)	(.1473)	(.2964)	(.2863)	(.3094)	(.3445)	(.2997)
BA	.0341*	-.0245	-.0332	-.0258	-.0974*	-.0365	.0529*	.0479	.0545	.0480	.0739	.0543
	(.0113)	(.0504)	(.0495)	(.0525)	(.0485)	(.0516)	(.0169)	(.0851)	(.0818)	(.0895)	(.0722)	(.0862)
CFA	-.0012	-.0152	.0099	-.0144	-.0041	.0111	-.0254†	-.0325	-.0179	-.0326	-.0244	-.0176
	(0.0095)	(.0204)	(.0206)	(.0210)	(.0235)	(.0212)	(.0153)	(.0343)	(.0354)	(.0360)	(.0474)	(.0371)
Thres,γ	-		.1353		.1353	.1515			.1353		.1353	.1515
$q(\gamma)$	-		.4773		.4773	.5568	-		.4773		.4773	.5568
adjR2	-.0007		.4365			.4372	.0172		.4569			.4577
BIC	15.42		**14.85**			14.86	15.44		**14.85**			14.86
TurnPt	33.46	33.67	32.93	33.21	7.178	31.09	81.91	60.93	227.8	62.27	24.55	516.5
		1A	1B	2A	2B	2C		1A	1B	2A	2B	2C
						Z=annual VIX						
Female(%)	-.1034*	-.0431*	-.1228*	-.0349*	-.0580*	-.1280*	-.0973*	-.1222*	-.0892*	-.1207*	-.0722	-.0989
	(.0047)	(.0039)	(.0061)	(.0067)	(.0120)	(.0116)	(.0180)	(.0449)	(.0416)	(.0595)	(.0604)	(.0610)
Female^2	.0015*	.0006	.0018*	.0004	-.0005	.0023*	.0006*	.0007*	.0006*	.0007*	.0007*	.0005*
	(.0003)	(.0004)	(.0004)	(.0007)	(.0013)	(.0006)	(.0002)	(.0003)	(.0002)	(.0003)	(.0002)	(.0001)
Team Size	-.0115	-.4037	.1835	-.3985	-.6544	.8256	-1.561*	-1.252	-1.645	-1.228	-1.659	-1.568
	(.2609)	(.5979)	(.5991)	(1.272)	(1.348)	(1.261)	(.4198)	(1.251)	(1.246)	(1.818)	(1.876)	(1.789)
Exper	.4900*	.1744	.6949*	.2972	1.634*	.2879	-1.026*	-1.141*	-.9352*	-1.128*	-1.273*	-.7191
	(.0651)	(.1382)	(.1374)	(.2972)	(.3200)	(.2958)	(.1473)	(.3481)	(.3319)	(.4759)	(.4624)	(.4585)
BA	-.0341*	-.0487	-.0358	-.0669	-.1894*	.0236	.0529*	.0271	.0580	.0231	.0532	.0533
	(.0113)	(.0449)	(.0455)	(.1000)	(.0981)	(.1047)	(.0169)	(.1097)	(.1133)	(.1611)	(.1626)	(.1695)
CFA	-.0012	.0097	-.0026	.0108	-.0309	.0204	-.0254†	-.0252	-.0240	-.0238	-.0185	-.0243
	(.0095)	(.0203)	(.0184)	(.0401)	(.0392)	(.0350)	(.0153)	(.0457)	(.0408)	(.0644)	(.0617)	(.0561)
Thres,γ			1.858		1.858	2.642			1.858		1.858	2.642
$q(\gamma)$			.2957		.2957	.5843			.2957		.2957	.5843
adjR2	-.0007		.4362			.4361	.0172		.4567			.4566
BIC	15.42		**14.85**			14.86	15.44		**14.85**			14.86
TurnPt	33.46	35.70	33.56	49.07	-61.85	28.30	81.91	83.04	80.59	81.84	55.06	95.32

Table 5: Pooled regressions of raw and excess fund returns and risk-adjusted returns (Jensen's alpha and Fama French alpha) against fund characteristics (lagged at t-1): $FundSize$, measured by log of total net assets (LTNA), $FFlow$, fund flows, $Expense$- expense ratio (in %) , $Turnover$, portfolio turnover, $FundAge$, several dummy variables, D indicating SRI fund, Fund Family, and capital-style funds, overall CSR measures (all, strength and weakness) of the portfolio held, and percentage of female fund managers in team, as well as their interaction with CSR measures. Robust standard errors clustered by fund are given in brackets. $*$ indicates significance at 5% level or better and † indicates significance at 10% or better.

	Raw Returns			Excess Returns			Jensen's alpha			FF alpha		
	1	2	3	1	2	3	1	2	3	1	2	3
Intercept	12.99*	16.01*	20.14*	12.50*	15.46*	20.38*	-.7679	-.6231	-.4614	.1277	.7168	-1.051
	(.6301)	(.6795)	(.6896)	(.6738)	(.7247)	(.7322)	(1.544)	(1.625)	(1.628)	(1.493)	(1.565)	(1.595)
Fund Size	-.6701*	-.6960*	-.7156*	-.7700*	-.7961*	-.8214*	.1044	.1044	.1023	-.0564	-.0575	-.0438
	(.0679)	(.0698)	(.0693)	(.0729)	(.0753)	(.0744)	(.1615)	(.1616)	(.1616)	(.1577)	(.1572)	(.1577)
FFlow	-.0209	-.0201	-.0167	-.0177	-.0170	-.0132	.0626†	.0631†	.0632	-.0873*	-.0863*	-.0867*
	(.0171)	(.0166)	(.0196)	(.0169)	(.0164)	(.0191)	(.0376)	(.0376)	(.0378)	(.0399)	(.0387)	(.0383)
Expense	-1.854*	-2.638*	-3.501*	-2.115*	-2.891*	-3.947*	-2.808*	-2.839*	-2.898†	-3.585*	-3.691*	-3.264*
	(.3223)	(.3357)	(.3515)	(.3424)	(.3562)	(.3722)	(.7971)	(.8045)	(.8072)	(.7891)	(.8051)	(.8067)
Turnover	.1612	.1414	.0923	.1314	.1135	.0577	-.0928	-.0929	-.1004	-1.114*	-1.119*	-1.110*
	(.1048)	(.1020)	(.1026)	(.1097)	(.1071)	(.1074)	(.2459)	(.2459)	(.2463)	(.3385)	(.3385)	(.3355)
FundAge	.0415*	.0487*	.0448*	.0532*	.0610*	.0582*	-.0655*	-.0647*	-.0645*	-.0265	-.0270	-.0297
	(.0089)	(.0093)	(.0089)	(.0100)	(.0105)	(.0099)	(.0208)	(.0207)	(.0209)	(.0195)	(.0194)	(.0196)
D(FFamily)	1.483*	1.598*	2.005*	1.533*	1.632*	2.087*	-.5221	-.5089	-.4710	.8028	.8660	.8030
	(.2629)	(.2691)	(.2778)	(.2826)	(.2900)	(.2967)	(.7827)	(.7810)	(.7833)	(.7975)	(.7967)	(.7998)
D(SRI)	.1192	.1552	-1.041†	.2449	.3523	-.9117	-3.374*	-3.301*	-3.331*	-.5685	-.7081	-.5827
	(.5346)	(.5681)	(.6054)	(.5833)	(.6189)	(.6633)	(1.400)	(1.394)	(1.395)	(1.225)	(1.218)	(1.225)
D(capital)	1.182*	-1.039*	-3.733*	1.472*	-.7108*	-3.970*	3.027*	2.936*	2.711*	1.793*	1.466*	2.652*
	(.2507)	(.2949)	(.3033)	(.2704)	(.3153)	(.3247)	(.6031)	(.6857)	(.7021)	(.5710)	(.6357)	(.6476)
CSRall	.0062*	-	-	.0130*			-.0036			-.0225*		
	(.0021)	-	-	(.0021)			(.0053)			(.0047)		
CSRstr	-	-.0258*			-.0222*			-.0035			-.0167*	
	-	(.0020)	-	-	(.0020)			(.0047)			(.0045)	
CSRweak	-	-	-.1187*			-.1260*			-.0061			.0020
			(.0047)			(.0049)			(.0097)			(.0092)
Female(%)	-.0059	-0.0051	-.0116	-.0085	-.0085	-.0155	-.0223	-.0272	-.0116	-.0325*	-.0428*	-.0325†
	(.0075)	(.0087)	(.0088)	(.0079)	(.0091)	(.0094)	(.0152)	(.0188)	(.0192)	(.0144)	(.0181)	(.0197)
Female*CSR	-.0001	.0000	.0004*	-.0001	.0000	.0004	.0006*	.0003	.0000	.0006*	.0004*	.0003
	(.0001)	(.0001)	(.0002)	(.0001)	(.0001)	(.0002)	(.0002)	(.0002)	(.0003)	(.0002)	(.0002)	(.0003)
Adj R-Squared	.0048	.0102	.0339	.0069	.0098	.0379	.0084	.0081	.0080	.0130	.0125	.0110
Observations	11373	11373	11373	11373	11373	11373	11373	11373	11373	11373	11373	11373

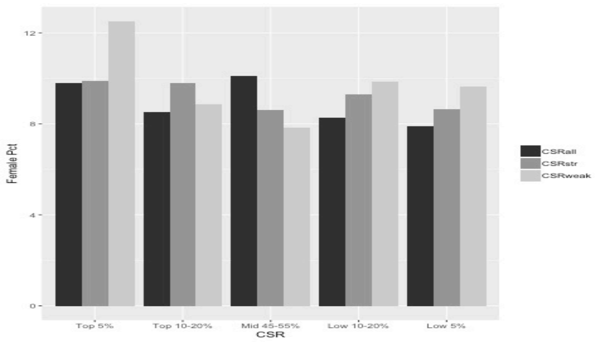
There is indeed a positive relationship against overall CSR, with a large and significant coefficient. The significant squared term also shows diminishing effects with increasing female percentage. Furthermore, we find that the positive relationship between female percentage is pronounced for "corporate governance", "employee relations" and "human rights" aspects in investment portfolios.

Figure 2The top left and right panels show the average female percentage and gender diversity for funds sorted according to their CSR investment portfolios. The lower left and right panels show the difference in female percentage and diversity respectively for funds with top 10% and bottom 10% investments in different CSR categories. figure.caption.8, top left and right panels, shows that funds with high CSR investments tend to have higher percentages of female managers as well as higher diversity scores. It however also shows that funds with investments that have high CSR weakness also have high female ratios and diversity. A further sort into the CSR categories (Figure 2The top left and right panels show the average female percentage and gender diversity for funds sorted according to their CSR investment portfolios. The lower left and right panels show the difference in female percentage and diversity respectively for funds with top 10% and bottom 10% investments in different CSR categories. figure.caption.8, bottom left and right) show that funds with investments in the employee aspect (EMP) of CSR have high levels of female managers compared to investments low in EMP. When using diversity, funds with investment high in the categories environment (ENV), corporate governance (CGOV), employee relations (EMP), and human rights (HUM) tend to have a more gender diverse team. Conversely, funds with investment high in product (PRO) and community (COM) aspects of CSR tend to have fewer female managers and less gender diverse teams.

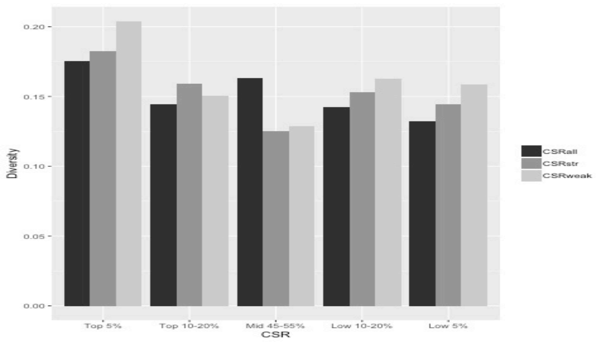
We repeat the estimations in the conditional threshold panel regressions model to observe if the gender diversity impact on CSR holds for differing economic conditions. Given that the idiosyncratic differences are captured by the fixed effects estimation, we leave out several fund manager team conditioning variables and retain $Exper$, which is correlated with Age and $Tenure$, and BA and CFA, proxies for higher education and continuing executive education. The conditioning variables used are again S&P500 returns and annual VIX. Table 7Results of threshold panel regression

Table 6: Pooled regressions of CSR scores and its sub-categories against fund and fund manager characteristics (using 5647 fund-year observations). Robust standard errors clustered by fund are given in brackets. $*$ indicates significance at 5% and † indicates significance at 10% level or better.

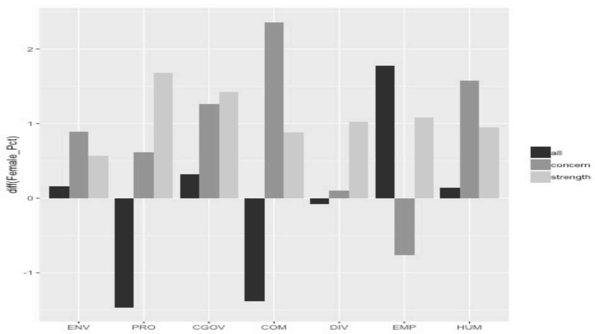
	overall	strengths	weakness	ENV	PRO	CGOV	COM	DIV	EMP	HUM
Intercept	-26.53*	2.900	29.43*	-2.493	-10.39*	-17.71*	2.985†	4.030	3.413*	-6.357*
	(9.785)	(16.86)	(11.73)	(2.021)	(2.847)	(1.672)	(7.690)	(4.786)	(1.566)	(2.125)
Fund Size	1.944*	3.680*	1.736*	.3928*	-.1584	.1897*	.3860*	.9625*	.1825*	.2349*
	(.5732)	(.9278)	(.4073)	(.1148)	(.1001)	(.1059)	(.0704)	(.2367)	(.0670)	(.0993)
Fund Flow	-1.736	-2.888	-1.152*	-.2449†	.0988	-.2583	-.2198	-.4372†	-.1500	-.5245
	(1.514)	(1.774)	(.4073)	(.1376)	(.1911)	(.1588)	(.2843)	(.2416)	(.2601)	(.5990)
SRI fund	-.4033	-8.058	-7.655†	.4352	2.68*	-1.173	-.1146	-1.343	.3308	-1.177
	(6.245)	(9.138)	(4.571)	(1.216)	(.9796)	(1.298)	(.7693)	(2.257)	(.6227)	(1.156)
Team Size	-.9290	-1.444	-.5151	-.2166†	-.0029	-.1175	-.0458	-.3053	-.2074*	-.0334
	(.5699)	(.9710)	(.5093)	(.1145)	(.1042)	(.1132)	(.0678)	(.2488)	(.0697)	(.0971)
Age	.2568†	.2422	.0206	.0297	.0578*	.0834*	.0301	.0097	.0075	.0445†
	(.1487)	(.2378)	(.1141)	(.0300)	(.0240)	(.0280)	(.0193)	(.0601)	(.0180)	(.0249)
Exp	-.0584	.0484	.1067	.0319	-.0796	.0081	-.0477	.0228	-.0251	.0312
	(.3825)	(.5930)	(.2747)	(.0741)	(.0588)	(.0752)	(.0468)	(.1409)	(.0408)	(.0708)
Tenure	-.5264	-1.070†	-.5434†	-.0922	.0575	-.1157	.0229	-.3006*	-.0499	-.0484
	(.3945)	(.6354)	(.2961)	(.0801)	(.0570)	(.0757)	(.0464)	(.1503)	(.0428)	(.0696)
Female	.3412*	.3612	.0200	.0520	.0392	.0936*	.0136	.0387	.0497*	.0543*
	(.1601)	(.2660)	(.1305)	(.0328)	(.0246)	(.0320)	(.0195)	(.0677)	(.0185)	(.0255)
Female^2	-.0080*	-.0075†	.0005	-.0014*	-.0010*	-.0020*	-.0006†	-.0006	-.0011*	-.0014*
	(.0025)	(.0041)	(.0020)	(.0005)	(.0004)	(.0005)	(.0003)	(.0010)	(.0003)	(.0004)
BA	.0465	-.0710	-.1176*	.0061	.0373*	.0223*	.0008	-.0301	.0049	.0052
	(.0567)	(.0952)	(.0493)	(.0115)	(.0100)	(.0108)	(.0071)	(.0236)	(.0069)	(.0105)
MA	.0297	.0478	.0180	.0024	-.0012	.0068	.0037	.0107	.0031	.0042
	(.0544)	(.0937)	(.0507)	(.0109)	(.0110)	(.0107)	(.0071)	(.0254)	(.0071)	(.0088)
MBA	.0220	.1472*	1.253*	.0085	-.0329*	-.0074	.0053	.0497*	.0030	-.0042
	(.0382)	(.0636)	(.0328)	(.0078)	(.0033)	(.0072)	(.0052)	(.0162)	(.0048)	(.0064)
PHD	.0108	.2659	.2552*	.0084	-.0500*	-.0159	-.0006	.0950†	-.0119	-.0141
	(.1227)	(.1973)	(.1071)	(.0239)	(.0249)	(.0293)	(.0133)	(.0491)	(.0132)	(.0188)
CFA	.0129	-.0093	-.0223	-.0021	.0050	-.0020	.0079†	.0025	.0043	-.0026
	(.0324)	(.0537)	(.0276)	(.0065)	(.0058)	(.0062)	(.0043)	(.0139)	(.0041)	(.0055)
Year F.E.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj R^2	.2856	.1641	.1776	.3288	.3272	.4706	.1086	.2713	.2040	.1538



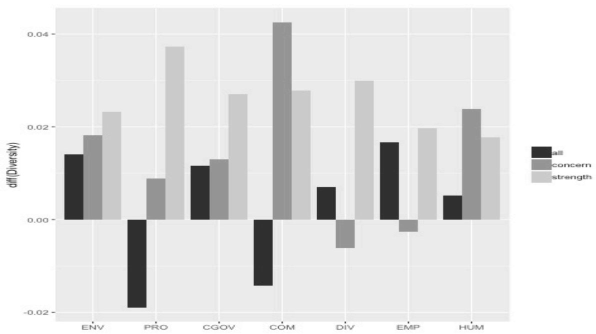
(a) % Female in funds sorted by CSR



(b) Diversity in funds sorted by CSR



(c) By CSR category: Top 10%-Low 10% average female



(d) By CSR category: Top 10%-Low 10% average diversity

Figure 3: The top left and right panels show the average female percentage and gender diversity for funds sorted according to their CSR investment portfolios. The lower left and right panels show the difference in female percentage and diversity respectively for funds with top 10% and bottom 10% investments in different CSR categories.

with (fund) fixed effects of CSR scores (overall and strength) against fund manager characteristics for matched funds with zero, one and two thresholds conditioned on demeaned Z=S&P500 and annual VIX. Numbers in brackets are robust fund clustered standard errors. † indicates statistical significance at 10% level while * indicates statistical significance at 5% or better.table.caption.9 reports the regression results where the dependent variable is the fund investments overall CSR score. The first column gives the fixed effects regression when no thresholds are applied. As in the previous pooled estimations, we observe that the percentage of female managers is positively linked to the fund's overall CSR score. The significant quadratic term $Female^2$ again shows that the marginal magnitude of the increase decreases with the proportion of the female fund managers.

Using threshold model, we find interestingly that impact of gender composition is higher on CSR scores when economic conditions are poor (lower S&P returns (regimes 1A, 2A, 2B), or higher market volatility (regimes 1B, 2B, 2C). Using BIC, the optimal threshold model chosen is the two threshold model for both conditioning variables. 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10 plots the impact of female ratio on overall CSR. Figure 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(a) plots the female fund manager's impact on fund's CSR investments under two S&P500 excess return regimes: bearish market (orange) and bullish market (grey). We find that female fund manager's effect is stronger in a bearish stock market, leading the fund to hold higher CSR stocks. On the contrary, funds tend to hold less CSR stocks in a bullish stock for the same proportion of female fund managers. It implies that a more gender-diverse fund management team, to some degree, tend to hold on to a higher CSR portfolio even when market sentiment is negative. Similarly, Figure 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(b) shows the results for two VIX regimes, supporting our previous findings that a more gender diverse fund tend to hold higher CSR stocks in a more uncertain market than in a less uncertain market. Figures 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(c) and 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(d) plot the threshold regression results for 3 regimes. The results are similar to our two-regime results, although we find the effect of female ratio is less apparent in extreme bearish stock (Figures 3S&P500 and VIX Threshold

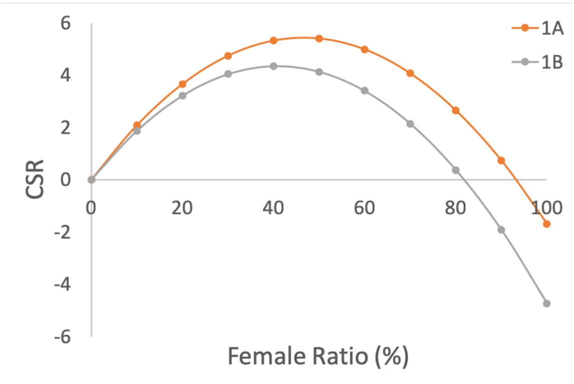
Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(c), blue) and very high uncertainty (Figure 3S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.figure.caption.10(d), grey) markets. The estimated thresholds for these regimes are however more extreme at the 8 percentile of S&P returns and 80 percentile for volatility[12]. We also report additional threshold regression results for where the dependent variables are CSR strengths and CSR weakness (see Table 11Results of threshold panel regression with (fund) fixed effects of CSR weakness against fund manager characteristics with zero (0), one (1A and 1B) and two (2A, 2B and 2C) thresholds conditioned on S&P500 annual returns (left) and annualized VIX (right). Numbers in brackets are robust fund clustered standard errors. † indicates statistical significance at 10% level while * indicates statistical significance at 5% or better.table.caption.16 in the Appendix). We find similar results, with the gender diversity effect much stronger for CSR strengths compared to CSR weakness. A more gender diverse team will actively choose to hold investments with higher CSR strengths during bearish markets rather than simply avoid investing in CSR weak portfolios.

Finally, we repeat our conditioning regressions, this time using fund performance as conditioning variables. We choose the last period's fund excess returns and fund

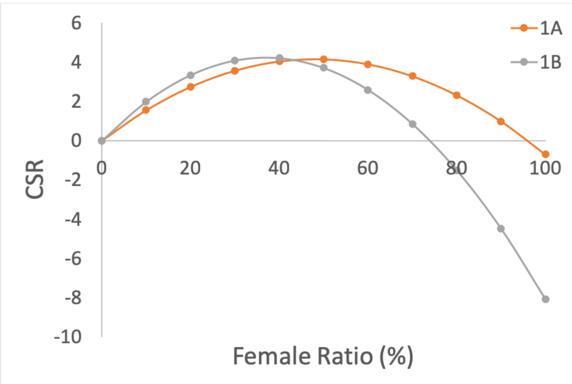
[12]Note that we cannot interpret the absolute threshold values for the conditioning variables as they have been demeaned.

Table 7: Results of threshold panel regression with (fund) fixed effects of CSR scores (overall and strength) against fund manager characteristics for matched funds with zero, one and two thresholds conditioned on demeaned Z=S&P500 and annual VIX. Numbers in brackets are robust fund clustered standard errors. † indicates statistical significance at 10% level while * indicates statistical significance at 5% or better.

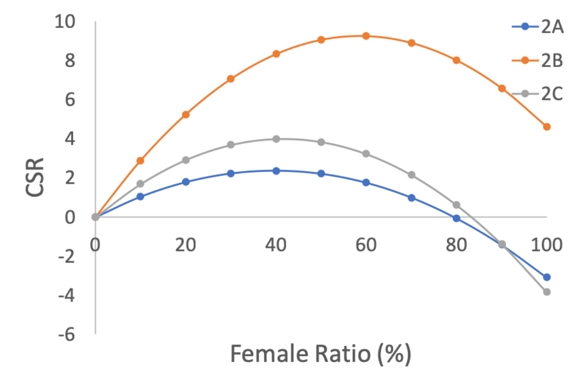
Variables		Z=S&P500 returns					Z=annual VIX				
	0	1A	1B	2A	2B	2C	1A	1B	2A	2B	2C
	Dependent variable= CSR overall										
Female(%)	.1755*	.2332*	.2127*	.1192	.3162*	.1917*	.1729*	.2292*	.1672*	.2993*	.1814*
	(.0416)	(.0414)	(.0459)	(.0886)	(.0934)	(.0745)	(.0553)	(.0601)	(.0600)	(.0516)	(.0599)
Female^2	-.0021*	-.0025*	-.0026*	-.0015*	-.0027*	-.0023*	-.0018*	-.0031*	-.0016*	-.0028*	-.0024*
	(.0005)	(.0008)	(.0007)	(.0006)	(.0004)	(.0006)	(.0008)	(.0007)	(.0009)	(.0012)	(.0010)
Team Size	2.285*	2.954	2.218	1.602	4.038	2.1567	1.712	2.954	1.726	4.224†	2.690
	(.9385)	(1.896)	(1.852)	(2.495)	(2.511)	(2.457)	(2.304)	(2.169)	(2.661)	(2.520)	(2.548)
Exper	3.364*	5.101*	3.238*	2.249*	5.080*	2.919*	3.149*	4.595*	3.100*	4.306*	4.461*
	(.3243)	(.5644)	(.5074)	(.7374)	(.6885)	(.6366)	(.5532)	(.6264)	(.6143)	(.6866)	(.6823)
BA	-.0940*	-.0976	-.1202	-.1540	.0327	-.1006	-.1670	-.1300	-.1718	-.0081	-.1681
	(.0368)	(.1276)	(.1293)	(.1829)	(.0311)	(.1851)	(.1678)	(.1671)	(.1966)	(.1996)	(.1960)
CFA	.0736*	.0718	.0678	.0517	.0816	.0641	.0630	.0820	.0713	.1446	.0758
	(.0377)	(.0642)	(.0652)	(.0926)	(.0894)	(.0891)	(.0781)	(.0746)	(.0904)	(.0875)	(.0871)
Threshold ($\hat{\gamma}$)			.0572		-.2178	.0572		2.711		2.711	2.884
$q(\hat{\gamma})$			.4209		.0794	.4209		.7137		.7137	.7945
$\overline{R}^2$	.0242		.0739			.1544		.0784			.0990
BIC	16.87		16.82			16.74		16.82			16.80



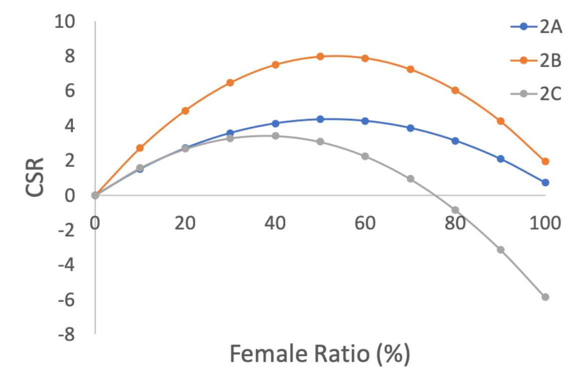
(a) 1-threshold S&P500



(b) 1-threshold VIX



(c) 2-threshold S&P500



(d) 2-threshold VIX

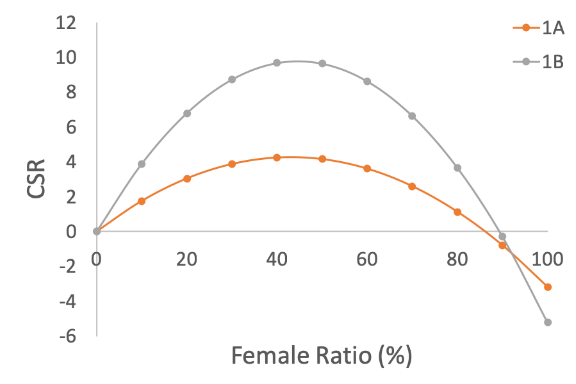
Figure 4: S&P500 and VIX Threshold Regressions: (a) and (c) plots the effect of female manager ratios on overall CSR investments conditioned on the S&P500 returns for the one and two threshold respectively. Regimes 1A, 2A and 2B indicate poorer market returns or economic sentiment, while regimes 1B and 2C indicate high market returns. (b) and (c) conditions on the annual VIX for the one and two threshold models respectively. Here, regimes 1A and 2A indicate lower market volatility, and regimes 1B, 2B and 2C indicate high market volatility.

flows as conditioning variable. This set of regressions tells us if fund management teams change their investment behavior according to the performance of the fund. Table 8Estimated parameters for the linear (0), 1 threshold (regimes=1A, 1B) and 2 threshold (regimes=2A, 2B, 2C) models conditioned on demeaned Z=Fund excess returns and Fund Flows. Dependent variable is the one period ahead overall CSR score. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level. 100 unique quantiles are used in the estimation with a trim ratio of 0.15. Estimated threshold is given by γ and the quantile of the estimated threshold $q(\gamma)$. Adjusted R^2 and BIC for each model is reported, where models with the lowest BIC are in bold. Estimated turning point for $Female(\%)$ in each regime is also given under $TurnPt$.table.caption.11 reports female ratio impact on fund's investment overall CSR score under different fund performance (left) and flows (right). BIC slightly favors the one threshold model over the two threshold model for both conditional variables. The results here are less surprising. We find that, for fund's excess return, a gender-diverse fund tend to hold higher CSR stocks if the fund has higher excess return in the last period, and if the fund has more fund inflow from investors in the last period. The results show that fund performance does constrain the behaviour of CSR investing by fund managers. A strong fund performance gives room for a more gender diverse team to invest more social responsibly compared to less gender diverse teams. Figures 4Panel Threshold Regression conditioned on fund excess returns and fund flows. The graphs show the effect of female ratio in the team on overall CSR scores conditioned on fund performance variables. (a) and (c) condition on excess returns for the one and two threshold models respectively, where regimes 1A, 2A and 2B show lower fund returns and regimes 1B and 2C shows high fund returns. (b) and (d) condition on fund flows for the one and two threshold models respectively, where 1A and 2A show lower fund flows while 1B, 2B and 2C shows higher fund flows.figure.caption.12 visualizes the gender ratio effects under different regimes. We also estimate the same threshold regression model using CSR strengths and CSR weakness as dependent variables (see Table 12Estimated parameters for the linear (0), 1 threshold (regimes=1A, 1B) and 2 threshold (regimes=2A, 2B, 2C) models conditioned on Z=Fund excess returns and Fund Flows. Dependent variable is the one period ahead overall CSR strengths and weaknesses. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level. 100 unique quantiles are used in the estimation with a trim ratio of 0.15. Estimated threshold is given by γ and the quantile of the estimated threshold $q(\gamma)$. Adjusted R^2 and BIC for each model is reported, where models with the lowest BIC are in bold. Estimated turning point for $Female(\%)$ in each regime is also given under $TurnPt$.table.caption.17 in Appendix). We obtain similar results, with the gender diversity effects on CSR strengths much higher than CSR weakness.

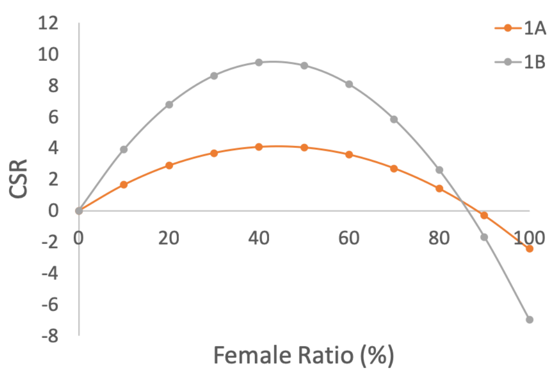
Table 8: Estimated parameters for the linear (0), 1 threshold (regimes=1A, 1B) and 2 threshold (regimes=2A, 2B, 2C) models conditioned on demeaned Z=Fund excess returns and Fund Flows. Dependent variable is the one period ahead overall CSR score. Robust standard errors clustered by fund are given in brackets. * indicates significance at 5% level. 100 unique quantiles are used in the estimation with a trim ratio of 0.15. Estimated threshold is given by γ and the quantile of the estimated threshold $q(\gamma)$. Adjusted R^2 and BIC for each model is reported, where models with the lowest BIC are in bold. Estimated turning point for $Female(\%)$ in each regime is also given under $TurnPt$.

		Z=excess returns					Z=fund flow				
	0	1A	1B	2A	2B	2C	1A	1B	2A	2B	2C
						Dep Var= CSR all					
Female(%)	.3186*	.1928	.4379*	.1763	.2155*	.4316*	.1856*	.4404*	.1869†	.3657*	.4747*
	(.0423)	(.1456)	(.1695)	(.1784)	(.0846)	(.1791)	(.0889)	(.1028)	(0.1009)	(.0899)	(.1257)
Female*	-.0036*	-.0023*	-.0049*	-.0020*	-.0029†	-.0048*	-.0021*	-.0051*	-.0021*	-.0040*	-.0056*
	(.0005)	(.0001)	(.0001)	(.0002)	(.0017)	(.0002)	(.0008)	(.0003)	(.0010)	(.0014)	(.0003)
Team Size	1.197	-.2223	1.681	-.2556	-.0779	1.754	1.745	.8037	1.687	.8918	.7119
	(.9768)	(4.474)	(4.329)	(4.955)	(5.243)	(4.883)	(3.069)	(3.069)	(4.097)	(4.119)	(4.160)
Exper	3.158*	1.602	2.748*	1.492	1.975	2.765	3.351*	3.128*	3.323*	3.273*	3.004*
	(.3478)	(1.114)	(1.135)	(1.273)	(1.295)	(1.277)	(.8351)	(.7730)	(1.101)	(1.190)	(1.004)
BA	-.0842*	-.1040	.0096	-.1223	-.0516	.0112	-.0342	-.1047	-.0326	-.0811	-.1087
	(.0408)	(.3585)	(.3492)	(.4067)	(.4033)	(.3963)	(.2253)	(.2400)	(.3065)	(.3036)	(.3301)
CFA	.0482	.0510	.0601	.0589	.0445	.0618	-.0188	.0732	-.0199	.0642	.0735
	(.0375)	(.1509)	(.1536)	(.1702)	(.1735)	(.1731)	(.1030)	(.0959)	(.1370)	(.1360)	(.1285)
Threshold			.0964		.0199	.0964		-.0086		-.0086	-.0036
q			.2230		.2209	.2230		.3587		.3587	.4951
adjR^2	.0201		.8605			.8606		.8605			.8605
BIC	16.80		**14.85**			14.86		**14.82**			14.82
TurnPt	43.92	41.08	45.11	44.17	36.63	45.24	43.54	43.33	43.54	45.75	42.27

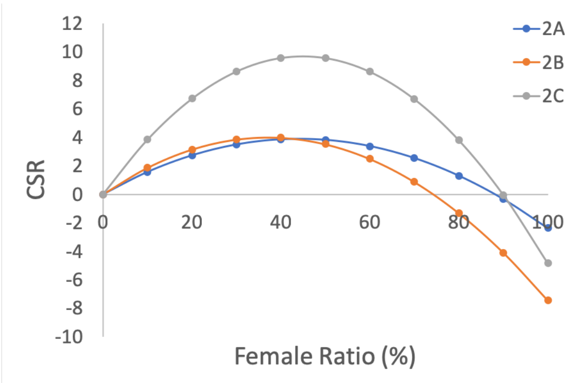
Overall, our results can be summarized as such: we find that gender diversity does negatively impact fund performance, but this is nonlinear and becomes marginally



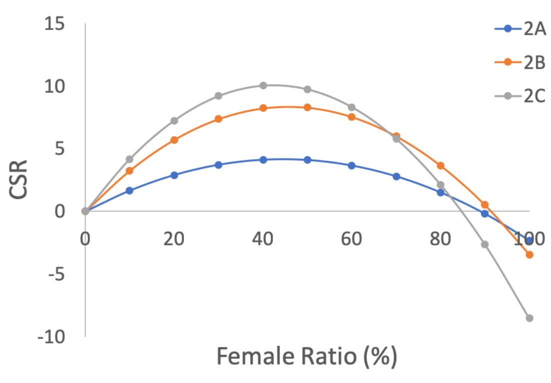
(a) 1-threshold Excess Return



(b) 1-threshold Fund Flow



(c) 2-threshold Excess Return



(d) 2-threshold Fund Flow

Figure 5: Panel Threshold Regression conditioned on fund excess returns and fund flows. The graphs show the effect of female ratio in the team on overall CSR scores conditioned on fund performance variables. (a) and (c) condition on excess returns for the one and two threshold models respectively, where regimes 1A, 2A and 2B show lower fund returns and regimes 1B and 2C shows high fund returns. (b) and (d) condition on fund flows for the one and two threshold models respectively, where 1A and 2A show lower fund flows while 1B, 2B and 2C shows higher fund flows.

positive when female ratios are above 40%. We find that investing in CSR does have a cost to risk-adjusted fund performance as well, and that there is a positive relationship between gender diversity and CSR scores of the investment portfolios, and that this effect is also nonlinear and peaks at a similar ratio of about 40%. Conditional threshold regressions tell us that gender diverse teams tend to invest more heavily in CSR portfolios, in particularly CSR strengths, even in bearish economic conditions. This might explain the negative performance of teams with more gender diverse portfolios. We also find, as expected, that fund performance impacts CSR investments. Gender diverse teams invest significantly more in CSR, especially CSR strengths, when funds are performing well (with high excess returns or investment fund flows) as compared to poor performance.

5 Conclusion

In this paper, we study the mutual fund gender diversity impact on fund's performance and its CSR investment decision. Using a very comprehensive dataset drawn from multiple sources about 2155 U.S. equity mutual funds with fund manager teams from 2002 to 2015, we obtained detailed information about the fund characteristics, fund managers characteristics including age, education, experience and genders, as well as what the equity investments they hold in their investment portfolios. We match these investments to their KLD scores and compute CSR scores for the investment portfolios. There are seven components to these scores and we consider these as well as overall scores in our estimations.

Considering only fund performance, we find that gender diversity, proxied in our case as percentage of female managers in the mutual fund team, does negatively affect both non-risk-adjusted and risk-adjusted fund performance. This effect is however non-linear and becomes marginally positive when there are approximately more than 40% females in the team. One explanation is the competing effects that teams have on decision making- group shift and diversity of opinions. At less than 40% females, it seems that group shift effects dominate, and the dominant gender of the team, males, causes decisions to gravitate to more extreme (negative) outcomes and when there are more females in the team, diversity of opinions take over to less extreme outcomes.

Our paper offers an alternative explanation for this result by considering how socially responsibly the team invests. We first find that there is a positive relationship between gender diversity and the social responsible investments, which we proxy using the estimated portfolio CSR score. This effect is interestingly nonlinear, becoming marginally negative when female ratios are above approximately 40%. We find evidence that investing in CSR does impose a cost on performance, as was found by several researchers (see e.g. Renneboog et al. (2008), Munoz et al. (2014), Belghitar et al. (2014)). Hence the negative performance from a higher female ratio could be explained by a gender diverse team being more likely to invest in CSR.

We further use conditional panel threshold regression models to allow the gender effect to differ under different market conditions. Interestingly, we find evidence that under poor market conditions (negative sentiment or greater uncertainty), the positive effect of fund gender diversity on CSR investments is much higher. This indicates that gender diverse teams still chooses to invest more social responsibly even in poor market conditions. Furthermore, we find this effect to be much more significant for CSR strengths rather than CSR weakness, indicating that gender diverse teams choose actively to invest in social responsible investments rather than just avoiding "sin" stocks. This further strengthens our argument that gender diverse teams have poorer fund performance because they tend to hold CSR investments even in market downturns.

Finally, we condition CSR investing on fund performance, proxied by the last period's fund returns and fund flows. We find, as expected, that fund performance does influence how much is invested in CSR, with the positive gender diversity effect on CSR investments much more pronounced when fund performance is good. Clearly, while gender diverse firms have greater tendency to invest social responsibly, they are limited by fund performance conditions, and have more room to do so when the fund is performing well.

Overall, our research has implications on gender research, social responsible investments and mutual fund research. We provide an argument for having a more gender diverse team for investment decisions when considerations are sustainability, environment and social impact rather than simply monetary returns.

References

Adams, R. B. and P. Funk (2012). Beyond the glass ceiling: Does gender matter? *Management Science 58*(2), 219–235.

Atkinson, S. M., S. B. Baird, and M. B. Frye (2003). Do female mutual fund managers manage differently? *Journal of Financial Research 26*(1), 1–18.

Babalos, V., G. M. Caporale, and N. Philippas (2015). Gender, style diversity, and their effect on fund performance. *Research in International Business and Finance 35*, 57–74.

Bär, M., A. Kempf, and S. Ruenzi (2011). Is a team different from the sum of its parts? evidence from mutual fund managers. *Review of Finance 15*, 359–396.

Bär, M., A. Niessen, and S. Ruenzi (2009). The impact of team diversity on mutual fund performance. Working paper, Center for Financial Research (CFR).

Beckmann, D. and L. Menkhoff (2008). Will women be women? analyzing the gender difference among financial experts. *Kyklos 61*, 364–384.

Belghitar, Y., E. Clark, and N. Deshmukh (2014). Does it pay to be ethical? evidence from the ftse4good. *Journal of Banking & Finance 47*, 54–62.

Benson, K. L. and J. E. Humphrey (2008). Socially responsible investment funds: Investor reaction to current and past returns. *Journal of Banking & Finance 32*(9), 1850–1859.

Bliss, R. T. and M. E. Potter (2002). Mutual fund managers: Does gender matter? *The Journal of Business and Economic Studies 8*(1), 1.

Bollen, N. P. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis 42*(3), 683–708.

Brody, L. R. (1993). On understanding gender differences in the expression of emotion. *Human feelings: Explorations in affect development and meaning* (87–121).

Byrnes, J. P., D. C. Miller, and W. D. Schafer (1999). Gender differences in risk-taking: A meta-analysis. *Psychological Bulletin 125*(3), 367–383.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*, 57–82.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–474.

Dwyer, P. D., J. H. Gilkeson, and J. A. List (2002). Gender differences in revealed risk taking: evidence from mutual fund investors. *Economics Letters 76*(2), 151–158.

El Ghoul, S. and A. Karoui (2017). Does corporate social responsibility affect mutual fund performance and flows? *Journal of Banking & Finance 77*, 53–63.

Ferson, W. E. and R. W. Schadt (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance 51*, 425–461.

Fujita, F., E. Diener, and E. Sandvik (1991). Gender differences in negative affect and well-being: the case for emotional intensity. *Journal of personality and social psychology 61*(3), 427.

Gaspar, J.-M., M. Massa, and P. Matos (2006). Favortism in mutual fund families? evidence of strategic cross-fund subsidization. *Journal of Finance 61*, 73–104.

Gonzalo, J. and J. Pitarakis (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics 110*, 319–352.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics 93*, 345–368.

Hartog, J., A. F. i Carbonell, and N. Jonker (2002). Linking measured risk aversion to individual characteristics. *Kyklos 55*(1), 3–26.

Hogg, M., J. Turner, and B. Davidson (1990). Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization. *Basic and Applied Social Psychology 11*, 77–100.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American Economic Review 92*(5), 1644–1655.

Jehn, K. A., G. B. Northcraft, and M. Neale (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly 44*, 741–763.

Joecks, J., K. Pull, and K. Vetter (2013). Gender diversity in the boardroom and firm performance: What exactly constitutes a "critical mass?". *Journal of business ethics 118*(1), 61–72.

Kerr, N. L. (1992). Group decision making at a multialternative task: Extremity, interfaction distance, pluralities and issue importance. *Organizational Behavior and Human Decision Processes 52*, 64–95.

Kreander, N., R. H. Gray, D. M. Power, and C. D. Sinclair (2005). Evaluating the performance of ethical and non-ethical funds: a matched pair analysis. *Journal of Business Finance & Accounting 32*(7–8), 1465–1493.

Merton, R. C. and R. D. Henriksson (1981). On market timing and investment performance ii: Statistical procedures for evaluating forecasting skills. *Journal of Business 54*, 513–534.

Moscovici, S. and M. Zavalloni (1986). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology 12*, 125–135.

Munoz, F., M. Vargas, and I. Marco (2014). Environmental mutual funds: Financial performance and managerial abilities. *Journal of Business Ethics 124*(4), 551–56.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niessen-Ruenzi, A. and S. Ruenzi (2017). Sex matters: Gender bias in the mutual fund industry. Working paper, University of Mannheim.

Qiong, L., K. Yang, and Y. Zhou (2014). The effect of management team characteristics on performance and style extremity of mutual fund portfolios. *Journal of Industrial Engineering and Management 7*(1), 294–3101.

Rakowski, D. (2010). Fund flow volatility and performance. *Journal of Financial and Quantitative Analysis 45*, 223–237.

Renneboog, L., J. T. Horst, and C. Zhang (2008). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance 14*(3), 302–322.

Renneboog, L., J. Ter Horst, and C. Zhang (2011). Is ethical money financially smart? nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation 20*(4), 562–588.

Sah, R. and J. Stiglitz (1986). The architecture of economic systems: Hierarchies and polyarchies. *American Economic Review 76*, 716–727.

Sah, R. and J. Stiglitz (1988). Committees, hierarchies and polyarchies. *The Economic Journal 98*, 451–470.

Schubert, R., M. Brown, M. Gysler, and H. W. Brachinger (1999). Financial decision-making: Are women really more risk averse? *American Economic Review 89*(2), 381–385.

Sila, V., A. Gonzalez, and J. Hagendorff (2016). Women on board: Does boardroom gender diversity affect firm risk? *Journal of Corporate Finance 36*, 26–53.

Soll, J. B. and J. Klayman (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30*(2), 299.

Teachman, J. D. (1980). Analysis of population diversity. *Sociological Methods and Research 8*, 341–362.

Torchia, M., A. Calabrò, and M. Huse (2011). Women directors on corporate boards: From tokenism to critical mass. *Journal of Business Ethics 102*(2), 299–317.

Treynor, J. and K. Mazuy (1966). Can mutual funds outguess the market? *Harvard Business Review 44*, 131–136.