Empirical Methods for Finance

First Graded Assignment
Do Investors Care about Sustainability?

Prof. Virginia Gianinazzi

Due: September 24, 2021

In an effort to increase financial flows to sustainable investments, policy makers often advocate for higher transparency about the sustainability of mutual fund portfolios. Among the various measures of sustainability that have become available, the Morningstar Sustainability Rating is a particularly popular one. The rating was launched in March 2016 and classifies mutual funds into five groups based on how the firms in their portfolios score on environmental, social, and governance (ESG) issues. Within each Morningstar category, funds are ranked from low sustainability (one globe) to high sustainability (five globes). The globe rating is prominently reported on each fund's Morningstar page[1].

[1] Check it out for yourself: Go to morningstar.com and look for a fund, e.g. PRBLX, then go to Portfolio

While you have frequently read that flows into ESG funds have increased dramatically in recent years, you turn to the data and to your knowledge of econometrics to address a fundamental question: Do mutual fund investors care about sustainability?

For this purpose, you are given an anonymized dataset provided by Morningstar that contains information on US domiciled, open-ended equity mutual funds. Variable definitions are at the end of this document.

1. Basic data manipulation [25%]

   After opening the data, perform the following tasks

   (a) Generate a variable *age* corresponding to the age of the fund. *inceptiondate* is the date of the fund's inception.

   (b) Compute flows in percentage of the fund size by scaling flows by the size of the fund in the prior quarter ($flowpct_{i,t} = flow_{i,t}/size_{i,t-1}$). Be careful: there may be gaps in the time series...

   (c) Generate a variable $flowpct\_lead$ as the percentage flow in the following quarter.

   (d) Keep only observations from 2019 Q3.

   (e) Keep only observations with non-missing ESG rating, star rating, size and lead percentage flow.

   (f) Winsorize lead flows and size at the 1% level.

   (g) The variable esg_rating now is a string. Replace string values with numerical values according to the rating definition (high sustainability = 5, etc.).

2. Summary Statistics and Plots [30%]

   (a) How many mutual funds are left in the sample?

   (b) Present summary statistics of funds by ESG rating. The table should report the number of observations, mean, standard deviation and median of fund size, age, star rating, expense ratio and return.

   (c) For each of the above variables, test if the difference between Low and High sustainability funds is statistically significant. Also, test if there is a statistically significant difference in average star ratings between mutual funds rated 4 and 5 globes.

   (d) Run a regression to test if the $fraction$ of 5-star ratings is significantly different across ESG ratings.

   (e) Produce a bar plot of the average lead percentage flow by ESG rating. Produce a bar plot with 95% confidence intervals of the average lead percentage flow by ESG rating (in Stata use the cibar command).

(f) Briefly comment on the plot.

3. OLS: ESTIMATION AND INTERPRETATION OF THE RESULTS [40%]

   (a) Regress $flowpct\_lead$ on esg_rating. Interpret the slope coefficient.

   (b) Now do the same, but this time include dummy variables for the different ESG ratings:

   $$flowpct\_lead = \beta_0 + \beta_1 I_{(ESG=2)} + \beta_2 I_{(ESG=3)} + \beta_3 I_{(ESG=4)} + \beta_4 I_{(ESG=5)} + \varepsilon$$

   Interpret the coefficients. What is the estimated change in flows from increasing the rating from 3 to 4? And from 4 to 5?

   (c) Briefly comment on the statement: "The model in point b) is more flexible than then model in point a), because the model in point a) is assuming a linear effect of the ESG rating on flows."

   (d) Repeat the above but select ESG=3 as the base level (in Stata, use the following factor notation to change the set the omitted category $ib3.esg\_rating$). Interpret the coefficients. How did the coefficients change with respect to the previous model? How do you explain that certain coefficients lose all or most of their statistical significance?

   (e) Under which assumptions can we interpret the estimated coefficients above as the causal effect of sustainability ratings on fund flows? What threats to the identification of the causal effect do you see in this case (your answer must be specific to this application)?

   (f) Regress again lead flows on the ESG dummies, but now control for $star\_rating$, $log(size)$, $log(age)$ and $log(exp\_ratio)$.

      i. Interpret the coefficient on $log(size)$.

      ii. How do the coefficients on the ESG dummies change? How do you explain it? ($Hint$: you should talk about omitted variable bias and correlation).

      iii. What do these last estimates suggest about whether sustainability ratings drive investors flows into mutual funds?

      iv. What seems to be the most important driver of flows? How do you interpret that? Does this surprise you? Explain and motivate (briefly, but clearly).

4. OLS MECHANICS [5%]

   (a) You are still worried that you are omitting important factors from your model. Compute the residuals and predicted values from the last estimated model. Compute the correlation between these two variables. Comment.

Datsets and Variables Description

The dataset *esg.dta* contains the following variables:

- $fundid$: unique fund identifier
- $return$: fund return (R_t)
- $size$: fund total net assets (TNA)
- $flow$: net flows, computed as $TNA_t - TNA_{t-1}(1 + R_t)$
- $star\_rating$: Morningstar Star Rating
- esg_rating: Morningstar Sustainability Rating
- $inceptiondate$: launch date of the fund
- exp_ratio: fund expense ratio