Missing Data in ESG Panels

Yanping Chen

April 15, 2024

Abstract

With the rise of ESG, its impact on stock returns has become an area of keen interest. However, the natural of this influence–whether positive, negative, or neural–remains a subject of contention. On the other hand, the ESG panels display a multivarite missing pattern and a missing at random (MAR) mechanism, which is seldom discussed in current ESG literature. Applying the multiple imputation method, we correct the selection bias arised from using only the part of firms that have ESG scores. Our results show a neural effect before correction and a negative effect after correction in the short term, and a positive effect in the long term.

1 Introduction

ESG stands for Environmental, Social, and Governance. It combines hundreds of performance indicators in these three areas [1] to evaluate a company's sustainability and ethical practices. In an era facing severe environmental and social challenges, ESG has been drawing increased attention from investors, analysts, regulators, researchers, stakeholders, etc, and is becoming pivotal in investment, corporate operations, and policy making. As a rising criterion, it is not surprisingly that many ESG issues are under discussion and debate (Starks, 2021; Edmans, 2023; Berg et al., 2022; Escrig-Olmedo et al., 2019), one of which is the motivations for ESG: Is it financial value driven or social values driven (Starks, 2023)?

With a growing belief that companies with higher social responsibilities and ethical considerations are more resilient, better managed, and exposed to lower potential risk, which can lead to sustainable

[1]The E (Environmental) pillar measures a company's impact on the environment, including considerations such as carbon emissions, energy efficiency, waste management, the use of natural resources, etc. The S (Social) pillar evaluates a company's relationship with its employees, customers, suppliers, and the communities in which it operates. The G (Governance) pillar assesses the quality of a company's leadership, management, shareholder rights, and adherence to ethical business practice.

and long-term financial performance, the empirical studies, however, have yielded mixed results. Luo (2022) examines the effect of ESG on stock returns in the United Kingdom and finds that firms with lower ESG earn higher returns than those with higher ESG, consistent with Hong and Kacperczyk (2009), Bolton and Kacperczyk (2021), and Pedersen et al. (2021). Serafeim (2020) and Kim and Li (2021), however, find a positive impact of ESG on corporate financial performance. On the other hand, Naffa and Fain (2022) and La Torre et al. (2020) get a neutral result. While this divergence of findings is considered to possibly come from the difference in ESG measurement approaches, temporal scopes, regional and industry-specific focuses, etc (Harris, 2022), we provide a new explanation from the data processing perspective–the sample selection bias arised from inappropriately handing missing data.

Current ESG literature typically use complete balanced panels without discussing the missngness in ESG. However, in our process of dealing with the ESG panels, we observe a non-negligible large proportion of missing values in ESG. Our analysis shows that large firms are much more likely to get rated by ESG agencies while many of the small firms don't have ESG scores, which belongs to the missing at random (MAR) mechanism and will result in a biased estimate if simply dropping the firms without ESG scores (Little and Rubin, 2019). On the other hand, missingness occurs not only in ESG but also in multiple variables across the whole panel dataset, which is a multivariate missing pattern where most imputation models for ESG scores using the remaining covariates as predictors can be problematic since the predictors themselves can contain missing values.

The missing data problem is actually ubiquitous in empirical economic studies. Abrevaya and Donald (2017) surveyed four top empirical economic journals [2] over three years from 2006 to 2008, and finds that nearly 40% of papers report the missing data problem, among which roughly 70% dropped observations due to missing values hence used the complete subset of data and around 20% used the imputation method. However, according to Little and Rubin (2019), simply discarding incomplete observations or imputing missing values without clarifying the relationship between the missingness and the values of covariates may affect the validity of inferences.

This paper first does a general and systematic literature review for the missing data problem, including the broadly used theoretical framework proposed by Rubin and current dealing methods. Second, we analyze the missingness in ESG panels and then employ the multiple imputation (MI) with predictive mean matching (PMM) due to its general missing pattern and the missing at random (MAR) mechanism. Finally, comparing the results from the complete subset of data and the results

[2]The four journals are American Economic Review(AER), Journal of Human Resources(JHR), Journal of Labor Economics(JLE), and Quarterly Journal of Economics(QJE)

with multiple imputation, we find that after correcting the selection bias, the ESG impact on stock returns is negative rather than neural in the short term, and positive in the long term.

The contribution of this paper is mainly in three aspects. First, it brings to our attention the importance of properly handling missing data and discusses the sample selection bias under this broad category. Second, we provide a new possible reason for the inconsistent results of ESG impact on stock returns, which adds to the ESG literature and can be extended to other ESG issues. Third, we do a thorough and broad literature review of the missing data problem, providing guidance for addressing the missing data problem in empirical economic research.

2 Literature Review

2.1 Theoretical framework

To address missing data, a usual way is to first identify the missing pattern and the missing mechanism. According to Little and Rubin (2019), the data missing pattern refers to the configuration of observed and missing values in the dataset, and the data missing mechanism concerns the relationship between missingness and the values of variables in the data matrix. The former tells what is missing and the latter talks about why it is missing.

The missing pattern is straightforward to identify if visualizing the dataset. The typical missing patterns include but not limited to

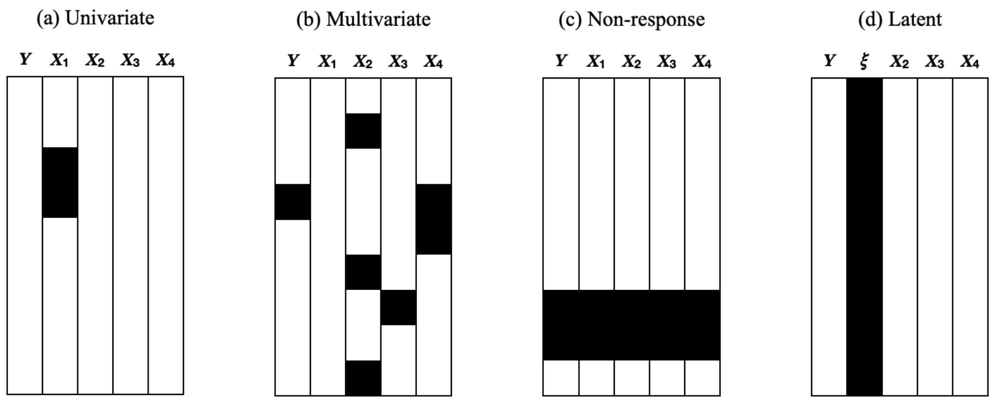

Figure 1: The Missing Patterns

(a) Univariate pattern: There is only one variable having missing values. It is the simplest form of missing data and involves only one dimension or features.

(b) Multivariate pattern: Missingness occurs in multiple variables and may be correlated across those variables. Multivariate missing patterns can complicate data analysis significantly since the standard imputation techniques that work for univariate patterns might not be effective and the correlations between different variables need to be considered when deciding how to handle the missing data.

(c) Non-response pattern: Non-response missing data occur when participants in a study or survey do not provide responses to some questions or sections, or just refuse to anticipate. This pattern can be systematic (when non-responses are not random and may be related to specific characteristics of respondents) or random. For example, people with low income may refuse to anticipate a labor survey.

(d) Latent pattern: Latent missing data involve situations where the missingness is related to unobserved or hidden variables. These latent variables influence the likelihood of missing data but are not directly measured themselves. For instance, in binary choice models, there is a latent variable that affect individual's decision.

The missing mechanism is extremely important as the properties of methods heavily depend on the nature of the dependencies in these mechanisms. However, its crucial role in the missing data analysis was largely ignored until this concept was first formalized by Rubin (1976), through the simple device of treating the missingness indicators as random variables and assigning them a distribution.

Define the complete data set (Y, X) where $Y = (y_{ij})$ have missing values and $X = (x_{ik})$ are all observed. The indicator matrix $M = (m_{ij})$ denotes the missingness of Y where $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is observed. Assume for simplicity that (x_i, y_i, m_i) are independently and identically distributed over i. Rubin (1976) characterizes the missing mechanisms by the conditional distribution of m_i given (y_i, x_i), i.e. $f_{M|Y,X}(m_i|y_i, x_i, \phi)$ where ϕ denotes unknown parameters, and classifies them into three categories.

(a) Missing completely at random (MACR): If missingness does not depend on the values of the data no matter missing or observed, that is, if for all i and any distinct values $(y_i, x_i), (y_i^*, x_i^*)$ in the sample space of (Y, X),

$$f_{M|Y,X}(m_i|y_i, x_i, \phi) = f_{M|Y,X}(m_i|y_i^*, x_i^*, \phi). \tag{1}$$

(b) Missing at random (MAR): A less restrictive case where missingness only depends on the observed components x_i, that is, for all i and any distinct values y_i, y_i^* in the sample space of Y,

$$f_{M|Y,X}(m_i|y_i, x_i, \phi) = f_{M|Y,X}(m_i|y_i^*, x_i, \phi). \tag{2}$$

(c) Missing not at random (MNAR): If the distribution of m_i depends on the missing components y_i, that is, Eq.(2) does not hold for some i and some values y_i, y_i^* of the missing components.

Let Z denote the component of the causes of missingness unrelated to X and Y. Schafer and Graham (2002) graphically represents the above missing mechanisms as follows.

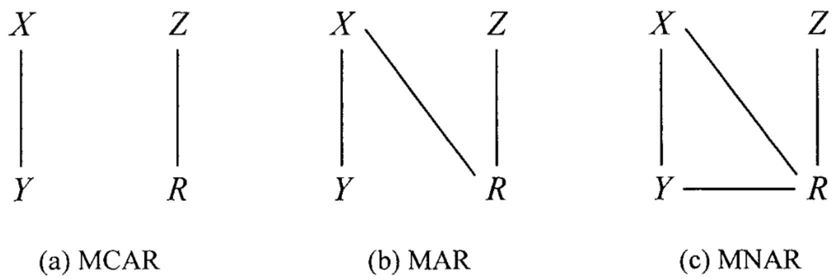

Figure 2: The Missing Mechanisms

The economics literature mainly focus on the selection problem in the category of missing not at random (MNAR). In Verbeek and Nijman (1996), the concept of ignorability is formalized, saying that a selection mechanism is ignorable if conditioning on the distribution of missingness does not affect the properties of the estimators under concern. Define a data set (y, x) and a response indicator variable r where $r = 1$ if both y and x are observed and $r = 0$ if either y is missing or both y and x are missing. Then a selection mechanism is defined to be ignorable if conditioning on the response indicator variable r does not affect the joint distribution of y and x, i.e.

$$f(y, x|\phi) = f(y, x|r, \phi). \tag{3}$$

In this case, estimators for parameters in marginal or conditional distribution involving y and x, whose consistency holds if $f(y, x|\phi)$ is the true distribution, are consistent. However, in more general cases, whether the estimators are consistent for the parameters of interest not only depends on the properties of the selection process, but also on the estimator which is used and on the parameters of interest. There are many cases that the selection rule is ignorable for some parameters but non-ignorable for other parameters.

Identifying the missing mechanisms is however a big challenge. Commonly used methods include:

- Visual inspection: Examine the patterns of missing data across variables. If there is a systematic pattern or certain variables are more likely to be missing than others, it could indicate non-random missingness.

- Subgroup analysis: Analyze subgroups to see if there are differences in missingness patterns. If certain subgroups consistently have more missing data, it may indicate a non-ignorable missing data mechanism.

- Subject-matter knowledge: Gain insights into whether the missing data are likely to be related to unobserved characteristics or other factors.

There are also some more rigorous statistical approaches to test some specific types of missing mechanisms. Little (1988) proposes a single global test statistic for MCAR that uses all of the available data and shows that its null distribution is asymptotically chi-squared. Based on the multiple imputation approach developed by Rubin (1987), the ignorability of the missing mechanism can be tested by imputing missing values in different imputation models and then checking whether the results are sensitive to the choice of imputation model. In Daniels and Hogan (2008), the data missing mechanism is tested by assessing the impact of different missing data assumptions on the results, that is, conduct sensitivity analyses to see how robust the results are to different assumptions about the missing data mechanism.

2.2 Existing Methods

2.2.1 Deleting Methods

Among traditional methods, the most commonly used one is to omit any records that are not fully complete and to conduct analyses solely on records with complete data. This approach is widely known as complete-case analysis (CC), also referred to as case deletion or listwise deletion. It is straightforward to use and might be adequate when the volume of missing data is minimal. Nevertheless, it can introduce significant biases and generally lacks efficiency, particularly when making inferences about specific subpopulations. On the other hand, available-case analysis (AC) employs varying sets of sample units for estimating different parameters. This method utilizes all the data that is available; however, its drawback is that the sample varies from one variable to another depending on the missing data pattern. Such variability complicates the verification of whether tables calculated for various conceptual sample bases are accurately defined and poses challenges in ensuring comparability across variables if the missing data mechanism depends on the variables being studied, i.e., if it is not MCAR. Kim and Curry (1977) demonstrated that AC methods are more efficient than CC methods when the data are MACR and the correlations are modest. Yet, other studies have shown that CC analysis may be superior when correlations are high (Haitovsky, 1968; Azen and Van Guilder, 1981). Nonetheless, according to (Hagenaars and Van Praag, 1985), neither method is generally satisfactory.

2.2.2 Weighting Methods

This type of methods re-weight the observed units by some designed weights in an attempt to adjust the estimates for incomplete units as if they were included. One widely used method is inverse probability weighting (IPW), which creates a pseudo-subpopulation by weighting each observation inversely to its probability of being observed. However, the effectiveness of IPW heavily relies on the correct specification of the model for the data-missing mechanism and IPW can lead to inflated variances, especially when the probability weights have high variance.

Wooldridge (2007) studies inverse probability weighted M-estimation under a broad missing data framework, including MAR and some MNAR cases (for example, the probability of observing depends on the stratum that falls into). Graham et al. (2011) introduces a variant of IPW known as inverse probability tilting (IPT). This approach aligns with Wooldridge's IPW estimator, but it replaces the conditional maximum likelihood estimate of the propensity score with a method-of-moments estimate. They demonstrate that if the unconditional moments selected for estimating the propensity score parameter are carefully chosen, this technique is locally efficient and remains robust even when the propensity score is inaccurately specified.

2.2.3 Imputation Methods

Imputation serves as a broad and adaptable approach for addressing missing data by filling in the missing values, allowing the use of standard analytical methods on the now complete dataset. This process involves creating a predictive distribution based on the observed data and then drawing values for the missing data from this distribution. Common imputation methods include:

- Mean imputation: This method involves substituting missing values with the mean of observed data points. Although simple, it significantly distorts statistical estimates, even when data are

MCAR. Simulation studies suggest that mean imputation is possibly the worst missing data handling methods available hence not recommended.

- Regression imputation: In this approach, missing values are replaced with estimates from a regression equation based on available data. The imputation process can be complicated if the data set is multivariate missing since each missing pattern requires a unique regression equation. This approach is superior to mean imputation but still has predictable bias. Opposite to the problem of mean imputation, the regression imputed values have high correlations with other variables thus overestimating correlations and even when the data are MCAR. Another drawback is the lack of variability as the imputed data fall directly on the predicted line.
- Stochastic regression imputation: This method enhances regression imputation by adding a random residual term to each predicted score, which helps restore variability and reduces bias, making it the only unbiased method under the MAR (Missing at Random) scenario. Also, there is not need to consider a variable's role in the subsequent statistical analysis when specifying an imputation model. However, this approach remains complicated with multivariate missing data and it also attenuates standard errors, leading to an increased risk of type I errors. The bootstrap resampling approach can correct this bias but it requires a lot more efforts.

An important limitation of single imputation is that standard sampling variance formulas applied to the filled-in data systematically underestimate the true sampling variance of estimates, even if the model used to generate the imputations is correct.

Multiple imputation (MI) creates multiple plausible imputed dataset, analyze each dataset separately, and combining the results to obtain more robust and unbiased parameter estimates. The idea behind this is to mimic the uncertainty associated with missing data and to provide more accurate estimates compared to a single imputation or complete-case analysis. The algorithm is iterative, and the number of imputations is usually determined based on statistical considerations and the desired balance between precision and computational efficiency.

2.2.4 Model-based Methods

Procedures of model-based methods are based on inferences for specific models. A typically example is the Expectation-Maximization (EM) algorithm for maximum-likelihood (ML) estimation. Each iteration of EM consists of an expectation step (E step) and a maximization step (M step), or more specifically, the EM steps are as follows: (i) replace missing values by estimated values; (ii) estimate parameters; (iii) re-estimate the missing values assuming the new parameter estimates are correct; (iv) re-estimate parameters; and so forth, iterating until apparent convergence.

For the mechanism of missing not at random (MNAR), the selection model proposed by (Heckman, 1979) describes how the probability of response to a sensitive questionnaire item (e.g. personal income) may depend on that item. It assumes a two-step process: the first step models the probability of selection (participation), and the second step models the outcome of interest given selection. Many econometrics literature have been focusing on this idea. For example, Kim and Curry (1977) derives semi-parametric estimators for the linear panel data model under sample selection when the explanatory variables are strictly exogenous. Semykina and Wooldridge (2010) shows how to estimate linear unobserved effects panel data models with endogenous explanatory variables and nonrandom sample selection. Semykina and Wooldridge (2018) considers estimating binary response panel data models in the presence of nonrandom selection. The primary advantage of selection models is their ability to account for selection bias by explicitly modeling the selection process. However, the effectiveness of selection models heavily relies on the correctness of assumptions about the selection process.

Alternatively, Little (1993) describes the pattern-mixture models which do not describe individuals' propensities to respond but classify individuals by their missingness and describe the observed data within each missingness group. The pattern-mixture models provide a flexible framework for explicitly modeling different missing data patterns, allowing researchers to account for the diversity of missingness mechanisms. They are often easy to fit, given assumptions to render the parameters estimable. However, the flexibility of pattern-mixture models comes at the cost of increased complexity. Specifying and estimating these models may be challenging, particularly when dealing with numerous patterns and parameters. And the results can be challenging to interpret due to the increased complexity and the potential presence of multiple parameters associated with different missing data patterns.

3 Data

Among various ESG rating agencies, we use the ESG dataset from MSCI (originally known as Morgan Stanley Capital International Inc.)–one of the leading providers for stock indexes whose ESG standards and methodologies have become a benchmark in the industry. After merging the ESG scores into the dataset of firms' characteristics from Compustat, we get a panel dataset with $841,625$ monthly observations from Dec 2012 to Dec 2022 (121 periods) for a total of $12,505$ firms. As is shown in Figure 3 in Appendix, our dataset confronts a severe missing data problem–a non-ignorably large proportion of missing data in multiple variables. Figure 4 visualizes the missingness of our dependent (stock returns) and independent variables (ESG scores), and some typical control variables. Figure 5 shows that the missing percentage in each variable is overall smooth, i.e. no sudden collapse in some year. Specifically, for the missingness in ESG scores, Figure 6 tells that the number of firms with ESG ratings is stable along time although the total number of firms is increasing. This is consistent with the suggestion of Figure 8 that the large missing proportion of ESG ratings mainly comes from the exclusion of some firms rather than the unexpected drop-out of rated firms in the interim.

As is shown in Figure 8, where the existing periods denote the number of periods that a firm exists in our dataset (at least one characteristic is observed), and the observed periods denote the the number of periods that a firm's MSCI ESG score is observed, around a half (even more) of firms don't have ESG ratings even once during the whole existing periods. To check whether there exists selection of firms by the ESG rating agency, we compare the characteristics of the ESG sample (i.e. firms with ESG ratings) and the full sample. As displayed in Figure 7, the distribution of $log(marketvalue)$ is centered around 8 for the ESG sample while it is around 5 in the full sample, suggesting that larger firms are more likely to be rated by the ESG rating agency. Besides, there is a noticeable difference in the distribution of industries in the ESG sample and the full sample: firms in the industry "Holding and Other Investment Offices" are much less likely to be rated than firms in other industries. Therefore, ESG is missing at random (MAR) and simply dropping the firms without ESG, which is the typical strategy adopted by current ESG literature, brings about selection bias.

Among our total of $12,505$ firms, $3,607$ firms exist throughout the whole 121 time periods which take up 51.86% (436,447 observations) of the panel dataset, 78 firms (0.624%) exist in the first and last period but miss somewhere in the middle, and the remaining $8,820$ firms don't exist either at the beginning or in the end of our time periods. As is shown in Figure 8, firms existing throughout the total 121 periods holds a dominant quantity than others, and Figure 6 suggests that firms existing less than 121 periods probably have market entry/exit behavior which involves firm's management and decision-making and is not the focus of this study. Therefore, we only focus on the balanced part of our dataset in this paper. Comparing the ESG sample and the full sample of the balanced subset of our panel data, Figure 9 shows that the data structure remains and the sample selection still exists.

4 Methods

Considering the MAR mechanism and the multivariate missing pattern, the methods we adopt are the multiple imputation (MI) which produces the unbiased estimate under MAR and the predictive mean matching (PMM) which is robust and simple enough to deal with the multivariate missing pattern.

4.1 Multiple Imputation (MI)

Multiple imputation is built on the Bayesian framework such that it relates the observed-data posterior distribution to the "complete-data" posterior distribution. Let θ denote the parameters we are interested in, Y_{obs} denote the observed part of data and Y_{mis} denote the missing data, we have

$$p(\theta|Y_{obs}) = \int p(\theta, Y_{mis}|Y_{obs})dY_{mis} = \int p(\theta|Y_{mis}, Y_{obs})p(Y_{mis}|Y_{obs})dY_{mis}.$$

The bias idea is to simulate the observed-data posterior distribution $p(\theta|Y_{obs})$ by first drawing the missing values $Y_{mis}^{(d)}$ from the conditional distribution $p(Y_{mis}|Y_{obs})$, imputing the drawn values to complete the dataset, and then drawing θ from the "complete-data" posterior distribution $p(\theta|Y_{mis}^{(d)}, Y_{obs})$. The mean and variance of the posterior distribution can be written as

$$\begin{aligned} \mathbb{E}[\theta|Y_{obs}] &= \mathbb{E}\Big[\mathbb{E}[\theta|Y_{mis}, Y_{obs}]\big|Y_{obs}\Big], \\ var(\theta|Y_{obs}) &= \mathbb{E}\Big[var(\theta|Y_{mis}, Y_{obs})\big|Y_{obs}\Big] + var\Big(\mathbb{E}[\theta|Y_{mis}, Y_{obs}]\big|Y_{obs}\Big). \end{aligned}$$

Multiple impution approximates the integral over the missing values as the average

$$\begin{aligned} \widehat{p}(\theta|Y_{obs}) &= \frac{1}{D}\sum_{d=1}^{D} p(\theta|Y_{mis}^{(d)}, Y_{obs}), \\ \widehat{\mathbb{E}}[\theta|Y_{obs}] &= \int \theta\widehat{p}(\theta|Y_{obs})d\theta = \frac{1}{D}\sum_{d=1}^{D}\int \theta p(\theta|Y_{obs}^{(d)}, Y_{mis})d\theta = \frac{1}{D}\sum_{d=1}^{D}\widehat{\theta}_d \end{aligned}$$

4.2 Predictive Mean Matching (PMM)

5 Results

6 Conclusions

Missing data are there, whether we like it or not.

References

Jason Abrevaya and Stephen G Donald. A gmm approach for dealing with missing data on regressors. *Review of Economics and Statistics*, 99(4):657–662, 2017.

S Azen and M Van Guilder. Conclusions regarding algorithms for handling incomplete data. In *Proceedings of the Statistical Computing Section, American Statistical Association*, volume 1981, pages 53–56, 1981.

Badi H Baltagi and Seuck Heun Song. Unbalanced panel data: A survey. *Statistical Papers*, 47(4): 493–523, 2006.

Florian Berg, Julian F Koelbel, and Roberto Rigobon. Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344, 2022.

Patrick Bolton and Marcin Kacperczyk. Do investors care about carbon risk? *Journal of financial economics*, 142(2):517–549, 2021.

Michael J Daniels and Joseph W Hogan. *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC press, 2008.

Peter Diggle and Michael G Kenward. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 43(1):49–73, 1994.

Gregor Dorfleitner, Gerhard Halbritter, and Mai Nguyen. Measuring the level and risk of corporate responsibility–an empirical comparison of different esg rating approaches. *Journal of Asset Management*, 16:450–466, 2015.

Alex Edmans. The end of esg. *Financial Management*, 52(1):3–17, 2023.

Elena Escrig-Olmedo, María Ángeles Fernández-Izquierdo, Idoya Ferrero-Ferrero, Juana María Rivera-Lirio, and María Jesús Muñoz-Torres. Rating the raters: Evaluating how esg rating agencies integrate sustainability principles. *Sustainability*, 11(3):915, 2019.

Bryan S Graham, C Campos De Xavier Pinto, and Daniel Egel. Inverse probability tilting estimation of average treatment effects in stata. *The Stata Journal*, 1(1):1–16, 2011.

Aldi JM Hagenaars and Bernard MS Van Praag. A synthesis of poverty line definitions. *Review of Income and Wealth*, 31(2):139–154, 1985.

Yoel Haitovsky. Missing data in regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(1):67–82, 1968.

Genese Caitlyn Harris. *The Relationship Between Environmental, Social, and Governance (ESG) Ratings and Firms' Return as Expressed by Corporations' Profitability*. PhD thesis, Capella University, 2022.

James J Heckman. Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161, 1979.

Harrison Hong and Marcin Kacperczyk. The price of sin: The effects of social norms on markets. *Journal of financial economics*, 93(1):15–36, 2009.

Jae-On Kim and James Curry. The treatment of missing data in multivariate analysis. *Sociological Methods & Research*, 6(2):215–240, 1977.

Sang Kim and Zhichuan Li. Understanding the impact of esg practices in corporate finance. *Sustainability*, 13(7):3746, 2021.

Ekaterini Kyriazidou. Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society*, pages 1335–1364, 1997.

Mario La Torre, Fabiomassimo Mango, Arturo Cafaro, and Sabrina Leo. Does the esg index affect stock return? evidence from the eurostoxx50. *Sustainability*, 12(16):6387, 2020.

RJA Little and DB Rubin. Multiple imputation for nonresponse in surveys. *John Wiley & Sons, Inc.. doi*, 10:9780470316696, 1987.

Roderick JA Little. A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association*, 83(404):1198–1202, 1988.

Roderick JA Little. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134, 1993.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.

Di Luo. Esg, liquidity, and stock returns. *Journal of International Financial Markets, Institutions and Money*, 78:101526, 2022.

Helena Naffa and Máté Fain. A factor approach to the performance of esg leaders and laggards. *Finance Research Letters*, 44:102073, 2022.

Lasse Heje Pedersen, Shaun Fitzgibbons, and Lukasz Pomorski. Responsible investing: The esg-efficient frontier. *Journal of financial economics*, 142(2):572–597, 2021.

Donald B Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

Donald B Rubin. Multiple imputation after 18+ years. *Journal of the American statistical Association*, 91(434):473–489, 1996.

Joseph L Schafer and John W Graham. Missing data: our view of the state of the art. *Psychological methods*, 7(2):147, 2002.

Anastasia Semykina and Jeffrey M Wooldridge. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380, 2010.

Anastasia Semykina and Jeffrey M Wooldridge. Binary response panel data models with sample selection and self-selection. *Journal of Applied Econometrics*, 33(2):179–197, 2018.

George Serafeim. Public sentiment and the price of corporate sustainability. *Financial Analysts Journal*, 76(2):26–46, 2020.

Laura T Starks. Environmental, social, and governance issues and the financial analysts journal. *Financial Analysts Journal*, 77(4):5–21, 2021.

Laura T Starks. Presidential address: Sustainable finance and esg issues—value versus values. *The Journal of Finance*, 78(4):1837–1872, 2023.

Marno Verbeek and Theo Nijman. Incomplete panels and selection bias. In *The econometrics of panel data: A handbook of the theory with applications*, pages 449–490. Springer, 1996.

Jeffrey M Wooldridge. Inverse probability weighted estimation for general missing data problems. *Journal of econometrics*, 141(2):1281–1301, 2007.

Jeffrey M Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2010.

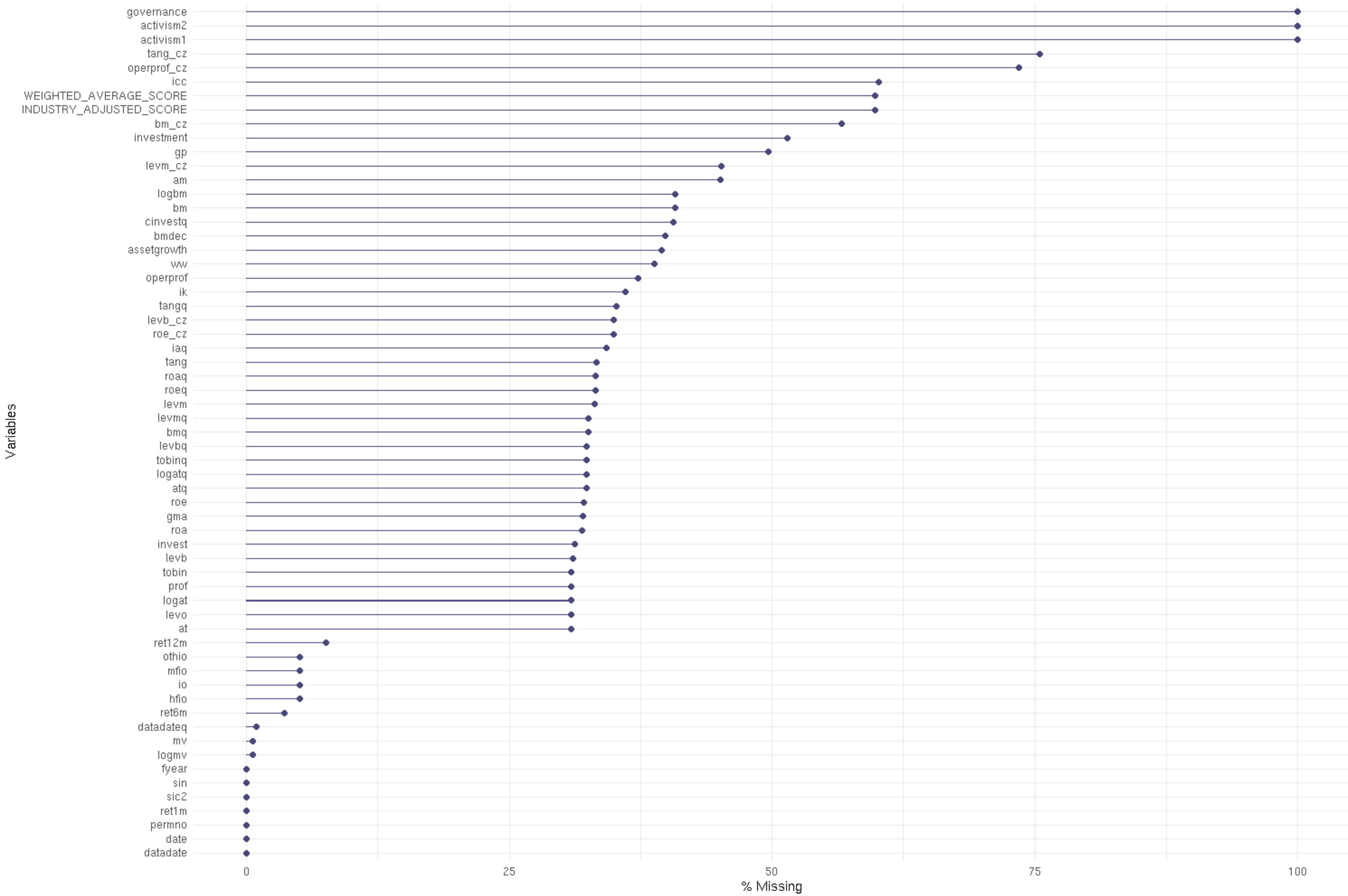

Figure 3: Missing Percentages of Variables

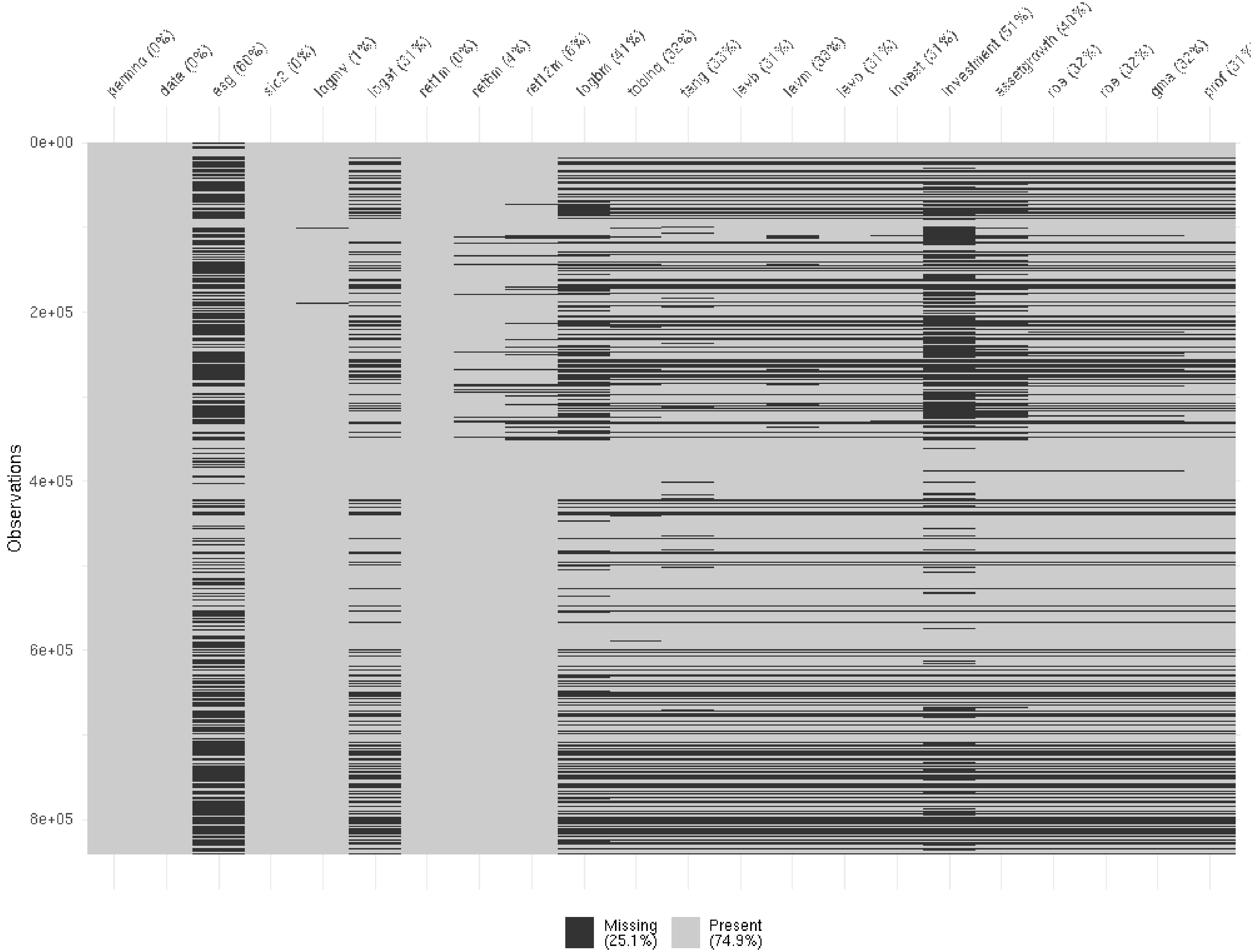

Figure 4: Visualization of Missingness in Panel

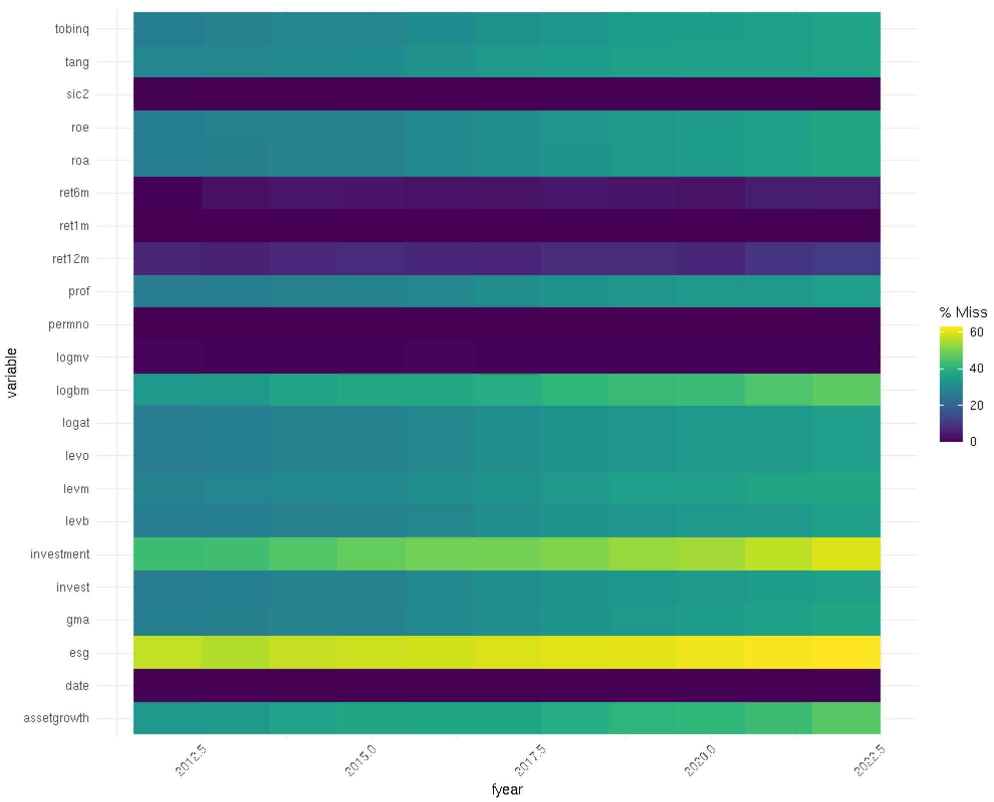

Figure 5: Missing Percentage in Variables along Time

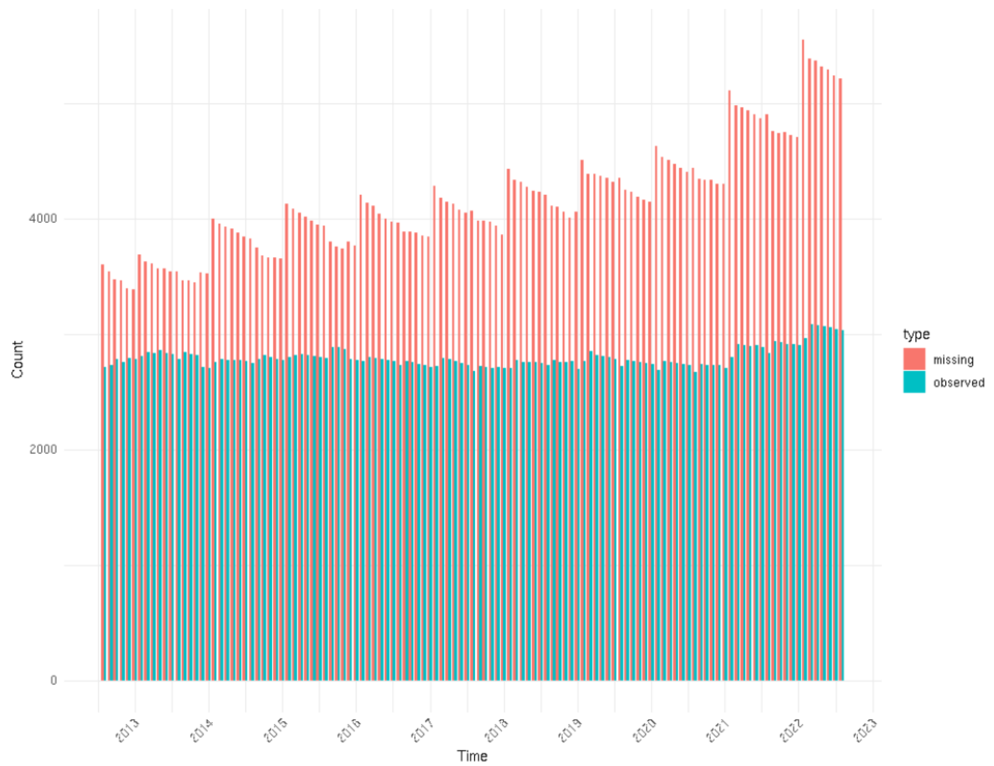

Figure 6: Number of Firms with/without ESG along Time

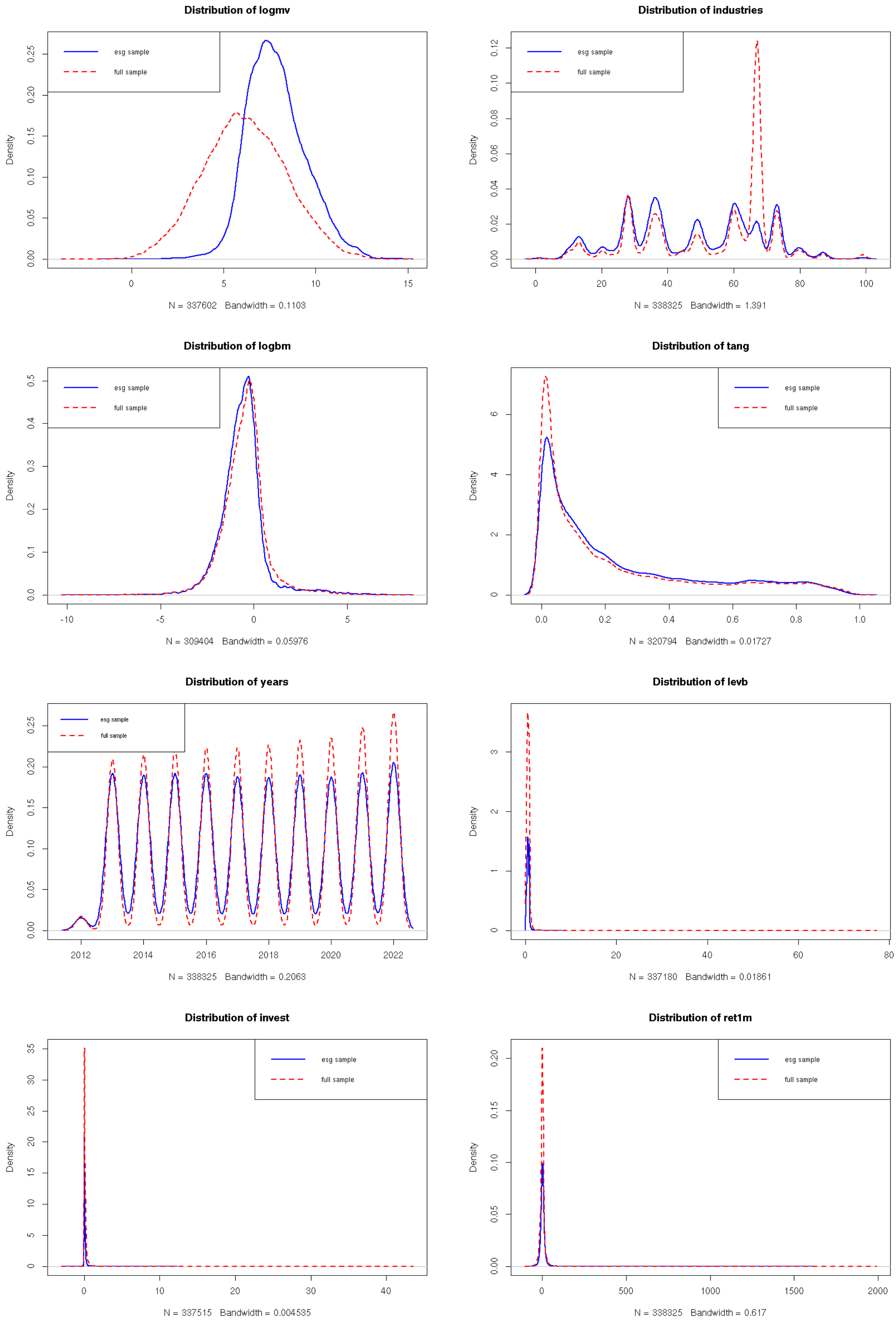

Figure 7: Characteristic Distributions in ESG Sample and Full Sample

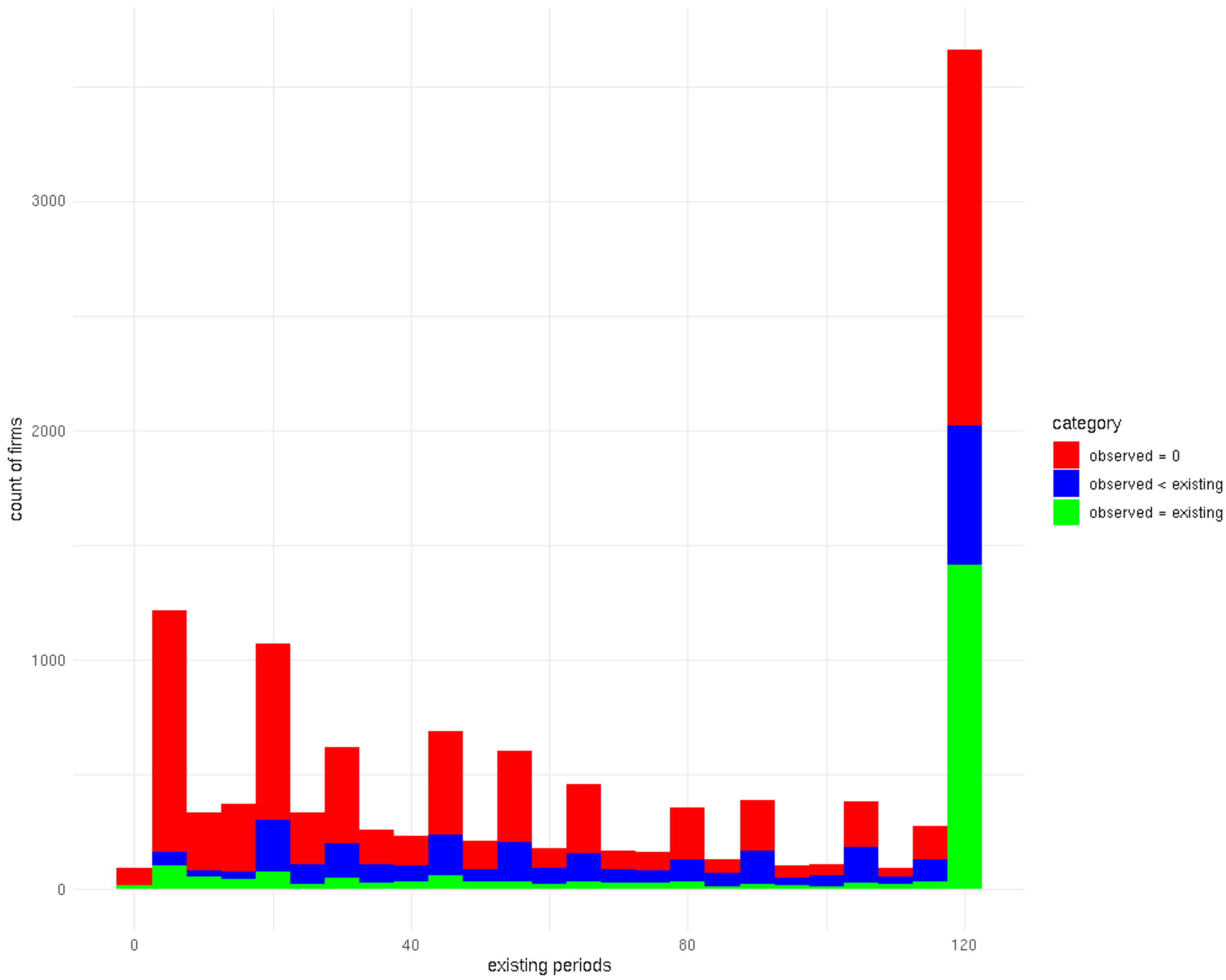

Figure 8: Existing periods and ESG Observed Periods of Firms

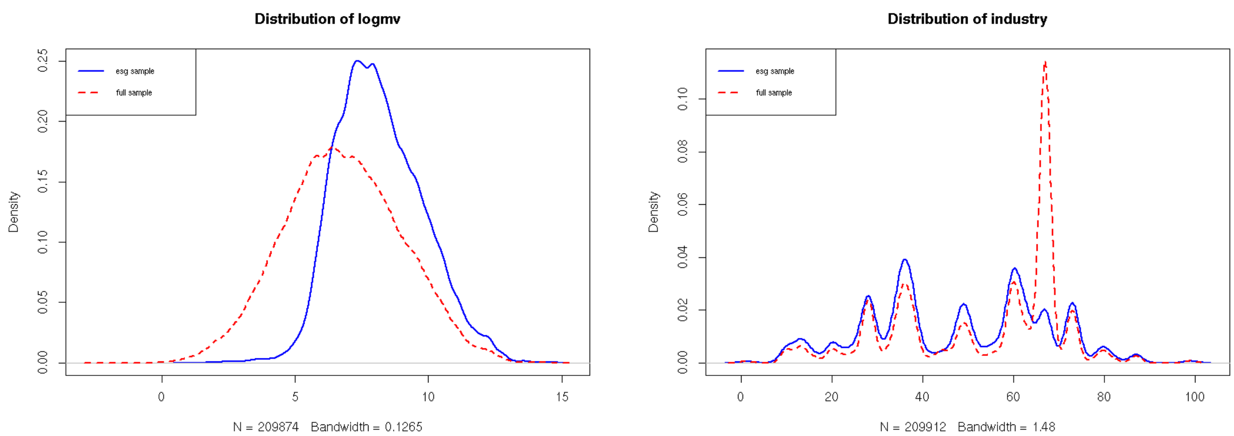

Figure 9: Comparison of ESG Sample and Full Sample in Balanced Panel

Table 1: Results from the ESG Sample

	Dependent variable:		
	ret1m		
	(1) $logmv <= 9$	(2) $logmv > 9$	(3) full-sample
esg	-0.059^{*}	0.046	-0.029
	(0.034)	(0.041)	(0.026)
momentum	0.043^{***}	0.070^{***}	0.055^{***}
	(0.001)	(0.001)	(0.0004)
logmv	0.360^{***}	0.011	0.113^{***}
	(0.031)	(0.043)	(0.017)
logbm	0.050^{*}	-0.005	0.016
	(0.028)	(0.051)	(0.023)
tang	0.004	-0.641^{**}	-0.084
	(0.193)	(0.309)	(0.161)
levb	0.102	-0.243	0.132
	(0.163)	(0.272)	(0.136)
invest	1.072^{***}	0.838	1.293^{***}
	(0.416)	(0.794)	(0.362)
roa	-0.398	0.527	0.152
	(0.280)	(0.594)	(0.242)
Constant	0.395	0.667	1.335^{**}
	(0.725)	(0.741)	(0.663)
time-fixed effect	✓	✓	✓
industry-fixed effect	✓	✓	✓
Observations	136,129	55,724	191,853
R^2	0.261	0.353	0.272
Adjusted R^2	0.260	0.351	0.271
Residual Std. Error	10.364	8.699	9.986
	(df = 135935)	(df = 55535)	(df = 191658)
F Statistic	248.237^{***}	161.182^{***}	369.169^{***}
	(df = 193; 135935)	(df = 188; 55535)	(df = 194; 191658)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 2: Results from Multiple Imputation

	Dependent variable:		
	ret1m		
	(1) $logmv <= 9$	(2) $logmv > 9$	(3) full-sample
esg	-0.059^{*}	0.046	-0.064^{*}
	(0.034)	(0.041)	(0.031)
momentum	0.043^{***}	0.070^{***}	0.048^{***}
	(0.0005)	(0.0006)	(0.0003)
logmv	0.360^{***}	0.011	0.085^{***}
	(0.031)	(0.043)	(0.010)
logbm	0.050^{*}	-0.005	-0.015
	(0.028)	(0.051)	(0.019)
tang	0.004	-0.641^{**}	0.098
	(0.193)	(0.309)	(0.103)
levb	0.102	-0.243	0.039
	(0.163)	(0.272)	(0.065)
invest	1.072^{***}	0.838	0.154
	(0.416)	(0.794)	(0.132)
roa	-0.398	0.527	-0.006
	(0.280)	(0.594)	(0.007)
Constant	0.395	0.667	0.916^{*}
	(0.725)	(0.741)	(0.495)
time-fixed effect	✓	✓	✓
industry-fixed effect	✓	✓	✓

Note: *p<0.1; **p<0.05; ***p<0.01