

## Data Description

1. *A description of which author(s) handled the data and conducted the analyses*

Katarzyna Bilicka, Elisa Casi, Barbara Stage and Carol Seregni handled the data and conducted the analysis.

2. *A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.*

- Ownership data are gathered from the Bureau Van Dijk Orbis database. This database was purchased by the University of Mannheim, and we downloaded the data in January 2021 while Elisa Casi, Barbara Stage and Carol Seregni were PhD students at this institution.
- Firm information has been downloaded from Datastream. This database was purchased by the University of Mannheim, and we downloaded the data in November 2020 while Elisa Casi, Barbara Stage and Carol Seregni were PhD students at this institution. Carol Seregni downloaded additional Datastream data in July 2023 for the extended sample. This data was purchased by the Wharton School (University of Pennsylvania).
- Company media exposure comes from Ravenpack. This database was purchased by the University of Chicago, and Carol Seregni downloaded the data in January 2021 while she was a visiting PhD student at this institution. Carol Seregni downloaded additional Ravenpack data in July 2023. This data was purchased by the Wharton School (University of Pennsylvania).
- Annual reports are retrieved from the Perfect Information Filings. This database was purchased by the University of Mannheim, and we downloaded the data in November 2020 while Elisa Casi, Barbara Stage and Carol Seregni were PhD students at this institution. We manually collected annual reports for treated firms without coverage in Perfect Information through web searches and the Refinitiv Eikon database in November 2023.
- We manually collected tax strategy reports of the companies in our sample. For that, we searched in google for a company tax strategy report using the following key words “tax strategy report” or “approach to tax” or “tax contribution” together with the company name. We performed the tasks in January 2021. We performed a similar search in the summer of 2023 to collect the most recent tax strategy reports available on firms’ websites.
- We manually collected modern slavery statements of the treated companies in our sample. For that, we searched in google for a company modern slavery statement using the following key words “modern slavery” together with the company name. We performed the tasks in May 2023.
- We manually collected gender pay gap report of the treated companies in our sample. For that, we searched in google for a company modern slavery statement using the following key words “gender pay gap” together with the company name. We performed the tasks in May 2023.
- We obtained data on “impossible reporting” (Bailey et al. [2022]) in regard to the gender pay gap reports through email correspondence with Aneesh Raghunandan in July 2023.

- Unconsolidated financial data at subsidiary level are gathered from the Bureau Van Dijk Orbis database. This database was purchased by the National Bureau of Economic Research (NBER). Katarzyna Bilicka downloaded the data in November 2023.
  - ESG scores come from Refinitiv ESG. The database is part of the WRDS subscription of the Wharton School (University of Pennsylvania). Data was downloaded by Carol Seregni in June 2023.
  - Accounting board expertise is based on data from BoardEx. This database was accessible through the University of Mannheim. We downloaded the data in April 2020 while Elisa Casi, Barbara Stage and Carol Seregni were PhD students at this institution.
  - We manually collected media articles about the UK reform (Schedule 19 of the 2016 Finance Act) using Nexis Uni. This database was purchased by NHH Norwegian School of Economics, and we downloaded the data in September 2023.
3. *If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).*

Not applicable as we do not use proprietary data in the analysis in this study.

4. *A complete description of the steps necessary to download, obtain or collect as well as process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.*

Section IV, “Sample selection and variable measurement,” in the manuscript and the attached codes describe the steps necessary to gather and process the data used in the final analyses reported in the paper.

5. *After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.*

Apart from hand-classification of the training sample for the Naïve Bayes analysis (as detailed in appendix D), all data manipulations are performed via computer programs and provided in the attached codes. The raw data can be downloaded from the sources identified above.

6. *The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the*

*manuscript. A brief description that enables other researchers to understand and run the code should be provided. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, and which commands were used in the analysis, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors' perspective. Therefore, instead of disclosing the proprietary portion of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to provide the identifiers (e.g., CIK, CUSIP) for their final sample.*

We provide the four sets of codes necessary to process the raw data, construct the relevant variables and conduct the analyses reported in the paper. Please refer to the attached *read\_me* text file for a description of each code. Codes are commented to facilitate the understanding of each step. The ISINs for our disclosure and ETRs samples (analysis reported in Table 7) are provided in the *Identifiers.xlsx* file.

7. *A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.*

The log files corresponding to each code are included in the folder.

8. *An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.*

We assure that we will maintain the data and programs for at least six years, consistent with National Science Foundation guidelines.