

Financial Transparency of Private Firms: Evidence from a Randomized Field Experiment

DATASHEET

Joachim Gassen

Humboldt University of Berlin

Maximilian Muhn

Chicago Booth

March 2025

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

Maximilian Muhn handled the archival data and conducted most of the analyses. Both Joachim Gassen and Maximilian Muhn vouch for the accuracy of the data.

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.

As indicated in the paper, two main sources were used for the analyses:

- **Unternehmensregister** (www.undernehmensregister.de/ureg/) for the main outcome variable *Restricted* (i.e., whether a given financial statement was restricted or not) as well as the date of the release (e.g., used for *Filing Lag*).
 - The website was crawled in early 2017 (mostly from February until May), end of 2017 (mostly in December), and mid 2018 (mostly in June and July).
- **Dafne** for the financial statement data.
 - Pre-experimental data (e.g., total assets, firm age, tangibility, equity ratio) is mostly based on BvD Dafne data release n° 694 as of December 2, 2016
 - Data for additional cross-sectional tests were downloaded at a later time (in particular, BvD Dafne data release n° 708 as of March 3, 2017 as well as release n° 782 as of September 14, 2018)

For additional tests, we also obtain data on firms' activity:

- **Wayback machine** from Archive.org: Crawl activity status of firms' websites in May, October and November 2023.

- **Creditreform (VVC)** activity status obtained from Dafne in November 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 the authors obtained data. 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 agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).**

For the archival analysis, we either used publicly available data (in particular, Unternehmensregister) or commercially-licensed data (in particular, Dafne from Bureau van Dijk). With access to Dafne from Bureau van Dijk, similar data should be available for other researchers.

The field data was directly generated via the tool LimeSurvey hosted by the Humboldt University of Berlin. This tool served a dual function: i) Individualized treatment e-mails were sent via the built-in email template function and ii) responses were collected by the embedded survey. Similar to the randomly assigned treatment identifiers, we consider this field data confidential information (e.g., we assured the confidentiality to our survey participants). However, we retained archives of the LimeSurvey tool and can privately provide the editors with the respective data.

Additionally, we conducted a confidential Dynata survey in October and November 2023 as well as obtained confidential survey data (as specified in the paper) from two other sources.

4. **A complete description of the steps necessary to collect and 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.**

We will provide the underlying code upon acceptance of the paper. For the field part, we have also compiled an appendix containing detailed explanations as well as all the questions and answers to the survey instruments which are not explicitly tabulated in the main text. Together, these questions and answers make up the full survey instruments.

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.**

We generally conduct all substantive data processing, sample construction, variable construction, and statistical analyses using computer code. As indicated in the log file (see Item 7), some steps related to obtaining, exporting, or preparing source files could not be fully automated. In particular, the initial Dafne data had to be downloaded manually through the Bureau van Dijk web interface and, due to download size limits, in multiple batches. In addition, some source files and data inputs were manually obtained, vetted, or prepared before being processed by the code, such as checks to identify firms that may already have restricted their 2014 financial statements and corrections to email addresses before conducting the experiment. These manually prepared inputs were retained and subsequently processed by the code.

6. **The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. 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, 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 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 final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample. Authors should consult our FAQ Sheet on the JAR website for further details.**

The code used to construct the final dataset and conduct the analyses is attached. Due to the confidential nature of the field data (see above), providing an identifier list for the final sample would not be sensible. We will retain and be able to privately share the firm identifiers, treatment status indicators, and the crawled register data with the editors.

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.**

A comprehensive log file is attached.

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 will maintain code and data for at least six years.