

Data Log for “The Effect of the Mandatory Disclosure of Corporate Tax Returns on Reporting Bias”

Eric Allen, Henry Friedman and Aydin Uysal

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

Aydin Uysal, Eric Allen and Henry Friedman collected the data. Eric Allen and Henry Friedman conducted the analysis using Sas, Stata and Python.

2. A detailed description of how the raw data were obtained or generated, including data sources, the 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.

All raw data was obtained by Aydin Uysal from the Borsa Istanbul Stock Exchange (BIST) (<https://www.borsaistanbul.com/en>) for stock transaction variables and pre-period financial statement data, and Public Disclosure Platform (PDP) (<https://www.kap.org.tr/>) for the book and tax financial statements. Eric Allen updated the data using disclosures from those sources as well as the Compustat Global Fundamentals Annual database. Henry Friedman obtained inflation adjustment factors from the St. Louis Fed (<https://fred.stlouisfed.org/tags/series?t=cpi%3Bturkey>) We note that the BIST removed public access to the financial statement archive in the fall of 2025. Access is still available upon request to members of the exchange and selected academics. Specific information is in the table below.

Variables	Source	Date Accessed
Book Financial Statement Variables, date of book earnings announcement	BIST (pre-2009), PDP (post 2009)	Downloaded starting in 2018. Last Accessed July 2025.
Tax Financial Statement Variables, date of tax earnings announcement	PDP	Downloaded starting in 2018. Last accessed July 2025.
Stock price, market adjusted returns, and other market variables	BIST	Downloaded daily starting in 2018. Last accessed November 2023
Capital expenditures, some of the book tax expense variables.	Compustat Global Fundamentals Annual	Last Accessed July 2025.
Inflation adjustment factors	St Louis Federal Reserve	Last Accessed August 2024

Note: The co-author who downloaded the majority of the raw data passed away in 2023 during the writing of the paper. He used his own web scraping software to download the raw variables over a period of years, starting before the co-authors joined the project. The remaining co-authors had access to his downloaded raw data sets, which they updated as necessary. The dates accessed reflect the estimated start date of the deceased co-author and the last time the remaining co-authors accessed the raw data source.

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 agreements, any restrictions imposed

by the organization on the authors, such as restrictions to publish certain results). In particular, the authors should indicate if an organization or data provider imposes restrictions on the publication of the results, has not given the authors full control of the relevant data, requires that the results must be reviewed or approved prior to public release of the paper or publication.

As noted in Internet Appendix 6 the BIST no longer allows public access to the financial statement archive which is where we obtained all pre-2009 financial statement data. At the time of the data collection the data was publicly available for anyone to access and we were not subject to any proprietary agreements restricting its use.

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.

See section 4 of the paper and Internet Appendix 6.

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.

Included in the replication package are the following files:

1. Merge raw data files.sas – merges the manually collected data described above into one combined data set
2. Merge raw data files log.txt – log of execution of previous file
3. Construct variables for analysis.sas – sas file converting raw data constructed in file 1 into the variables used in the paper
4. Construct variables for analysis.log – log file showing execution of previous file
5. Tables 1 and 2.do – Stata file used to construct tables Figure 2 and 1 and 2 in the paper
6. Tables 1 and 2 log.pdf – log file showing execution of .do file
7. run_all_notebook_final.ipynb – jupyter notebook file that generates remaining tables in main paper and online appendix. The output and log are shown in line with the code that executes it. This notebook incorporates the files below:
 - a. s_1_build_dataset.py – construct dataset used in structural analysis from samplenoscreens.xlsx output of ‘Construct variables for analysis.sas’ and ‘Turkey inflation by month – from FRED.xls’ downloaded from the Federal Reserve.
 - b. Estimator.py - method-of-moments estimator and firm-clustered jackknife for the structural model. Imported by make_tables.py and make_IA_tables.py
 - c. s_2_make_tables.py – construct tables 3-6 in the main draft
 - d. s_3_make_IA_tables.py – construct tables for the internet appendix
 - e. s_4.compile_pdf.py – bundles the tables generated by s_2 and s_3 into a single PDF file.

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 2 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. Authors should consult our FAQ Sheet on the JAR website for further details.

Please see the replication package noted above.

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.**

Please see the replication package noted above.

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.**

Eric Allen and Henry Friedman will maintain the data and programs in accordance with the guidelines.