

DATA DESCRIPTION SHEET

Tax Policy Expectations and Investment

John Gallemore
University of North
Carolina

Stephan Hollander
Tilburg University

Martin Jacob
IESE Business School

Xiang Zheng
Nanyang Technological
University

August 26, 2024

This data description sheet is part of our submission to the *Journal of Accounting Research* and provides the following information in accordance with *JAR*'s data policy:

Data handling and programming were performed as described in more detail below. All authors vouch for the stated sources of the raw data.

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

Markus Schwedeler, who generously contributed to this project, processed the raw Refinitiv EIKON data (earnings conference call transcripts and metadata) and constructed the tax policy sentiment measure. Stephan Hollander conducted other analyses using the earnings conference call data. Martin Jacob and Xiang Zheng handled the other raw data and merged datasets to arrive at the final sample used in the analysis. Martin Jacob and Xiang Zheng conducted all the empirical analyses. All authors have full access to the data and programs, with the exception of the code developed by Hassan et al (2019).

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 can vouch for the stated source of the raw data.

Our main analyses are based on data from the following sources: earnings conference call transcripts from the Refinitiv EIKON database (see 1, above), financial accounting data from Standard & Poor's Compustat, stock market data from CRSP, corporate headquarters location data from WRDS SEC Analytics Suite, data on European-based subsidiaries of U.S.-based parent firms from the Amadeus database which we directly obtained from Bureau van Dijk, data on tax haven subsidiaries from Subsidiary Data by WRDS and SEC EDGAR, data on unrepatriated foreign earnings from Audit Analytics, and data on sector-level investment from the U.S. Bureau of Economic Analysis. Data on tax policy uncertainty, general policy uncertainty, general policy sentiment, and non-political risk are from Hassan et al. (2019), made available on www.firmlevelrisk.com. Data were downloaded in January 2019, February 2020, April 2020, September 2020, October 2021, May 2022, June 2022, and July 2023. The ownership data in Amadeus are based on information from November 2017.

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, and any restrictions imposed by the organization on the authors). 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.

No proprietary data have been used in the analyses. All data can be accessed with a subscription (to the databases listed above in 2) or via <https://www.firmlevelrisk.com> and <https://www.johngallemore.com>.

- 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.**

All data have been obtained directly from the sources described in (2) above. Processing the data for use in the final analyses involves merging the raw data across sources into a firm-year panel, as explained 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.**

Attached to this datasheet is the replication package, which includes: (1) Stata code files and (2) a README file explaining the execution sequence and directory structure.

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

Please refer to the replication package. For sample construction, see “0_data_preparation.do,” “1_data_merge.do,” “2_rank_variables.do,” and “3_data_for_regression.do.” For statistical analysis and output generation, see “4_paper_figures_and_tables.do.” The identifiers are available in the “identifiers” subfolder.

- 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 refer to the log file, which shows that all code files were executed.

- 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 plan to maintain the data and programs from the paper for at least 6 years.