

Revenue Recognition Comparability and Analysts' Disclosure Processing Costs

Andrea Tillet
Wisconsin School of Business
University of Wisconsin-Madison
tillet@wisc.edu

Data Description Sheet

This data description sheet is provided in fulfillment of the *Journal of Accounting Research* data policy for submission of a manuscript.

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

I, Andrea Tillet, handled the construction of the datasets used in this manuscript. I have access to all the programs and data.

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).*

My empirical analysis relies on information from several data sources. I describe each of these data sources below. In addition, the "Sample Selection and Data" section of the manuscript and the Appendix provide variable definitions, as well as the corresponding data sources used to create the variables utilized in the empirical analysis.

I collected XBRL data in January and May 2021 and all other data between January-February 2021. Next, I updated all data in 2023 to extend my sample period to 2021.

- a. XBRL data was used to create the *RevPolicySim* measures as well as *PolicySim*. For additional information related to the XBRL data, including the specific XBRL tags that were used to construct these variables, please see the readme.txt file and part 5 of this Data Description Sheet.
 - b. Compustat Annual was used for the creation of firm's ASC 606 adoption date (*Post*) as well as additional control variables. Compustat Quarterly and Segments database were used for creation of control variables.
 - c. ASC 605 data was obtained from the FASB's Accounting Standard Codification and mapped to the SEC's Division of Corporate Finance SIC Code Classification List based on industry names.
 - d. Revenue recognition standards used to create *FASBSim* were collected from the FASB's website.
 - e. The product similarity TNIC data was obtained from the Hoberg-Phillips Data Library: <http://hobergphillips.tuck.dartmouth.edu/industryclass.htm>
 - f. I/B/E/S Detail was used for all analyst related data.
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).

Not applicable as I do not collect proprietary data.

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.*

I describe my data in the “Sample Selection and Data” section of the manuscript and define all variables in the Variable Appendix. The steps necessary to process the data are described in the computer programs referred to in part 6 of this Data Description Sheet.

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.*

After downloading the raw data, manipulations of the data were done in SAS and STATA. Three SAS and two STATA programs are available on the Journal of Accounting Research webpage.

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.*

I use Python, SAS, and STATA to process my data and conduct my empirical analysis. In addition to the Data Description Sheet, I am including a supplementary text file with my submission which will help researchers recreate the Python program. Other than the Python program, which I consider to be proprietary, I will submit other code and datasets with firm and analyst identifiers to the

Journal of Accounting Research to assist in the replication of my findings. Please see additional information below related to the computer programs used:

- Python program to download XBRL data and construct cosine similarity variables (e.g., *RevPolicySim* and *PolicySim*). Please see the “readme.txt” file for a description that enables readers to identify the relevant XBRL tags and recreate this program.
- SAS programs to download Compustat and I/B/E/S Detail raw data and construct key variables.
- STATA programs to clean and analyze data.
- Identifiers for the final sample of firms and analysts used in tests of H1/H2 and H3 are provided.
- Alternative comparability proxies (*RevOCFComp*) were created using the SAS code provided by Rodrigo Verdi: <https://mitmgmtfaculty.mit.edu/rverdi/publications/>

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.*

Comprehensive log files are provided for SAS and STATA programs.

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.*

I agree to maintain the data and programs used in this paper for at least six years, in line with NSF guidelines.