

Do Investors Value Auditor Involvement in Non-GAAP Reporting?

Phillip T. Lamoreaux
Arizona State University
phillip.lamoreaux@asu.edu

Lauren Matkaluk
University of San Diego
lmatkaluk@sandiego.edu

Amy G. Sheneman
The Ohio State University
sheneman.2@osu.edu

Data Description Sheet

April 2026

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

For the primary sample (Tables 1-4, 8, 10, 11), Lauren Matkaluk and Amy Sheneman handled the merging of raw datasets and the construction of variables. Amy Sheneman conducted the empirical analyses.

For the quarterly sample (Table 7), Amy Sheneman handled the merging of raw datasets, the construction of variables, and the empirical analyses.

For the S&P 500 sample (Tables 5-6, 9), Lauren Matkaluk and Amy Sheneman hand collected the data. Amy Sheneman conducted the empirical analyses.

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.

We have obtained data from various sources as outlined in Section 3.1 (Sample), 5.1 (Alternative Identification Strategy), 5.2 (Quality of Non-GAAP Information), 5.3 (Alternative Dependent Variable), 5.4.3 (Nature of Non-GAAP Exclusions), Appendix A (Variable Definitions), Appendix B (Description of the Methodology to Identify Firm-Years with Non-GAAP Reporting), Appendix C (Determinants Analysis), and Appendix D (Description of the Methodology to Construct the Alternative Measure of Auditor Involvement). The table below provides further detail on the dates that the raw data was obtained or generated:

Data Source	Data Obtained / Generated
Compustat	January 2022; updated December 2024 and March 2026 (for quarterly sample)
CRSP	February 2024; updated December 2024
I/B/E/S	February 2024
Audit Analytics	November 2022
Non-GAAP EPS from Bentley et al. (2018)	January 2022
directEDGAR	March 2022
SEC website for S&P 500 sample	July 2023-March 2024

All authors vouch for the stated sources of the raw 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., nondisclosure 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.*

All data used in this study are obtained from commercial databases or hand collected from publicly available documents (e.g., 10-K filings).

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.*

A description of the steps necessary to download, obtain, collect, and process the data used in the final analyses of the paper can be found in Section 3 (Research Design) and Appendices A (variable definitions and data sources), Appendix B (methodology for identifying firm-years with non-GAAP reporting), Appendix C (process for conducting the determinants analysis), and Appendix D (methodology for constructing our alternative measure of auditor involvement, %NG_RECONCILED).

We also provide detailed step-by-step instructions for running the provided code below:

For the primary sample

- LMS_PrimarySample.sas is the primary SAS code file to download Compustat, CRSP, I/B/E/S data. It relies upon the following four source files:

- LMS.2022.FINAL.sas is the SAS code to pull Compustat and Audit Analytics data from the WRDS servers. This is used as a source file in LMS_PrimarySample.sas.
- iclink_macro.sas is the SAS code to link CRSP to I/B/E/S and is referenced in LMS_PrimarySample.sas.
- institutional.sas is the SAS code to construct the institutional ownership variables. This is used as a source file in LMS_PrimarySample.sas.
- LMS_addlcontrols.sas, LMS_audchg.sas, LMS_avol.sas, LMS_dochg.sas, LMS_dvs.sas, LMS_loc.sas, and LMS_persistence.sas are individual SAS code files that construct Stata files used in the analyses of the primary sample.
- LMS_PrimarySample_T1.do is the Stata code that constructs the sample selection table (Table 1).
- LMS_PrimarySample.do is the Stata code that constructs the analyses for the primary sample, which includes Tables 2-4, 8, 10, and 11.

For the quarterly sample

- LMS_T7.sas is the SAS code to pull Compustat and CRSP data from the WRDS servers. This code also incorporates the input data from Bentley et al. (2018).
- iclink_macro.sas is the SAS code to link CRSP to I/B/E/S and is referenced in LMS_T7.sas.
- LMS_QuarterlySample is the Stata code to run the analyses tabulated in Table 7.

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.*

Manipulations of data were completed using the code provided with this datasheet for Tables 1-4, 7, 8, 10, 11, Table C1.

We are unable to provide code for the hand collected sample of S&P 500 firms (Tables 5-6, 9) as analysis using code was not practicable. The variables *%NG_VERIFIED*, used in Tables 5-6, and *EFFECTIVELY_RECONCILED*, used in Table 9, were constructed with excel after hand collecting data from non-GAAP reconciliations in Forms 10-K. All authors had access to this data at all times.

Because the hand collected sample identifies those firm-years that report non-GAAP information in their 10-K filings, which is the heart of our contribution and identification strategy, we have provided a detailed discussion of these variables in Appendix C and

Sections 5.1 and 5.4.3. Consistent with the editor's suggestion, we will make the %NG_VERIFIED data available 18 months from the date of this publication.

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.*

We have provided the code to (1) convert the raw data into the primary and quarterly samples used in the manuscript; (2) execute the statistical analyses; and (3) produce the tables. Identifiers for both the primary and quarterly samples are also provided.

For the hand collected sample of S&P 500 firms (Tables 5-6, 9), we provide a detailed discussion of the variable construction in Appendix C and Sections 5.1 and 5.4.3, to allow others to replicate our data collection process. We provide the log file for the regression results reported in Table 6. Consistent with the editor's suggestion, we will make the %NG_VERIFIED data available 18 months from the date of this publication.

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.*

Log files showing the execution of the SAS and STATA code have been provided for the primary sample (Tables 1, 2-4, 8, 10-11, C1) and the quarterly sample (Table 7). The descriptive statistics in Tables 5 and 9 were created with Excel using the hand collected

data; therefore, no log file is available. The log file for the regression analysis for the S&P 500 sample reported in Table 6 is also included.

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.*

The data and programs will be maintained by Amy Sheneman for at least six years, consistent with National Science Foundation guidelines.