

Data Description Sheet
“Disaggregating Cash Flows: The Effect of Linking and Labeling on Investors’
Understanding of the Statement of Cash Flows”

Richard Crandall, Shannon Garavaglia, and Cassie Mongold

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

All three coauthors handled the data and conducted the analysis.

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.

The data were obtained using an experimental instrument which was distributed to a pool of participants on the Prolific platform (Prolific.com). We screened participants, only allowing those who note they either “most of the time” or “always” make investments in common stock and evaluate a company’s financial statements when doing so. Additionally, we required a prior Prolific approval rating of 100%. Participants were not permitted to participate in multiple experiments in order to reduce the potential for demand effects across experiments.

Experiment One: We collected data on September 10, 2024. On September 10, 2024, we downloaded the data set from Qualtrics and collected no further data.

Experiment Two: we collected data on June 5, 2025. On June 5, 2025, we downloaded the data set from Qualtrics and collected no further data.

Experiment Three: we collected data on June 3, 2024. On June 4, 2024, we downloaded the data set from Qualtrics and collected no further data.

All authors have access to the Qualtrics surveys and can vouch for the source 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., 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.

Not applicable.

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.

The instruments we used to generate the data were included in the initial submission to the *Journal of Accounting Research*.

Using Prolific's participant screening functionality, we screened participants, only allowing those who note they either "most of the time" or "always" make investments in common stock and evaluate a company's financial statements when doing so. In the post-experimental questionnaire, we repeated the Prolific screening question regarding investment history. Participants whose responses contradicted their Prolific screening response were excluded from our analysis. Additionally, we required a prior Prolific approval rating of 100%. Participants were not permitted to participate in multiple experiments in order to reduce the potential for demand effects across experiments. For further details see Section 3 of the manuscript.

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 downloaded the data directly from the Qualtrics survey platform where it was collected. Raw data is available upon request from the authors. Attached to this datasheet please find the replication package which includes (1) a PDF file knitted from a Quarto Markdown file ("CGM-Analyses-R"), (2) two PDF files produced from SAS files ("CGM-CAT MOD direct vs theory Table 2" and "CGM-CATMOD for Understand"). Both the Markdown and SAS files contain the output following the code. Thus, the log and code are together in their respective 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 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. For Tables 1, 3, 4, and 5, please see “CGM-Analyses-R.” For Table 2, please see “CGM-CAT MOD direct vs theory Table 2” and “CGM-CATMOD for Understand.” Please note that R and SAS packages change over time which can cause slight differences in results. We have provided a brief description of our code’s functionality in the file “CGM-Readme.” As this study uses human participants recruited through Prolific rather than archival firm-level data, participant-level identifiers are not provided in order to protect participant privacy.

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 above-referenced PDFs that show the execution of the entire code.

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.

All three coauthors will ensure that at least one coauthor will maintain the data and programs according to the guidelines.