

**Data Description Sheet Following *Journal of Accounting Research*
Data and Code Sharing Policy**

“Taxes and Investment: Evidence from the ‘Halloween Massacre’ of 2006”

By Matthew Boland, Khin Phyo Hlaing, and Petro Lisowsky

June 2026

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

Matthew Boland performed data handling and programming, and conducted the analyses closely working with co-authors.

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 hand-collected data on income trusts in July-August 2020 from Monthly & Quarterly Statistical Summary published by TSX Venture Exchange and Toronto Stock Exchange, specifically from TSX E-review, 9 and TSX E-review, 10, published in September 2006 and October 2006, respectively. The names of the income trusts were manually matched with Compustat firms and extracted GVKEYs and TICs.

Because headquarters-location data in Compustat can be stale, we identify firms’ headquarters locations during our sample period using a multi-step procedure. First, we use headquarters locations from SEC filings compiled by Gao et al. [2021], supplemented with Compustat location data where necessary. Second, we use GPT-4 to identify headquarters locations for Compustat firms in our sample. Finally, we compare the two sources and manually review all discrepancies.

To validate the accuracy of the income trust sample matching process, we identify all firms classified as Canadian in 2006 and use GPT-4 to assess whether each firm was structured as an income trust at that time. We compare these results with the manually matched data and investigate any discrepancies in detail.

For further analyses, we hand-collected M&A data for our sample of income trusts from Capital IQ; MD&A discussions used in sentiment analysis from income trust financial statements collected from Capital IQ; and Google Trend data for internet keyword search analysis (trends.google.com). We used ChatGPT as a mode of analyzing managerial sentiment in MD&A text, which we disclose and describe in Section 5.4 and Appendix C of the paper.

All other data used in our manuscript are available from the Compustat annual and quarterly database and Canadian Financial Markets Research Center (CFMRC).

All co-authors vouch for the stated raw data sources.

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 have to be reviewed or approved prior to public release of the paper or publication. This information should be provided to the editors upon submission.

All data used in this paper are available from public sources. We do not use any data on a proprietary basis, and thus have no disclosure issues or restrictions to report.

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.

Raw data is obtained from the sources listed in Item 2 above. We report our sample selection procedure in Section 4 of the manuscript.

5. The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. 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, 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 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 final dataset used in the analysis. In such cases, the authors should inform the editors *upon initial submission*, so that the editors can consider an exemption from the code sharing requirement. Whenever feasible, authors should also 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 describe the process used to construct our main dataset in Section 4 of the manuscript. We provide the computer code and programs used to process the raw data, construct the final analysis sample, define the study variables, and generate all tables reported in the paper (see 2026.06.08 H Tests.do). These materials are intended to enable other researchers to replicate our sample construction procedures and empirical results.

The attached replication materials also include the identifiers for the final sample observations.

Use of ChatGPT in sample refinement. To facilitate replication of the GPT-assisted steps in our sample construction, we provide the prompts used for the relevant firm-level classifications. For identifying Canadian firms, the prompt was: “Which firms in this list were Canadian companies in 2005/2006?” For identifying Canadian income trusts, the prompt was: “Which firms in this list were Canadian income trusts in 2005/2006?” In each case, the prompt was accompanied by the relevant list of firms to be classified. GPT was used as an aid in refining these firm-level classifications rather than as a source of financial data. The replication file Data_CAN_Income.dta contains the relevant firm identifiers associated with these classification steps, allowing researchers to identify the firms included in the resulting classifications and reproduce their use in the sample-construction process.

6. 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 agree to maintain the data and programs from the paper for at least six years.