

# Index Creation, Benchmarking, and External Finance

Vidhan K. Goyal, Daniel Urban, Wenting Zhao

## Data Description Sheet for the Journal of Accounting Research

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

Daniel Urban and Wenting Zhao handled the data and conducted the 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.*

- Wenting Zhao collected information on the index events from the webpages of 39 index providers by performing a news search on these webpages between June 6, 2016, and August 8, 2016. These index events are available in PDF form and have been verified by Daniel Urban.
- Wenting Zhao collected information on the index constituents from Datastream and Bloomberg between June 6, 2016, and September 2, 2016.
- Daniel Urban and Wenting Zhao downloaded financial accounting data from Worldscope on August 18, 2016, August 26, 2016, and January 11, 2021.
- Wenting Zhao downloaded stock market data from Datastream on September 8, 2016, and October 4, 2016.
- Wenting Zhao downloaded analyst data from I/B/E/S on September 21, 2016.

- Wenting Zhao downloaded debt structure data from Capital IQ on September 9, 2016.
- Daniel Urban obtained information on news coverage from Ravenpack on December 1, 2016.
- Daniel Urban obtained bond return data from Bloomberg and Capital IQ on November 26, 2016.
- Daniel Urban obtained data on the implied cost of capital on March 16, 2018.
- Daniel Urban obtained data on M&A transactions from Zephyr as well as the accompanying stock return data from Datastream on January 14, 2021.
- Daniel Urban obtained data on corporate boards from BoardEx and ThomsonONE on January 11, 2021.
- Daniel Urban calculated the Leuz et al. (2003) earnings management index as well as a measure of industry concentration using Worldscope data obtained on January 18, 2021, and January 26, 2021.
- Daniel Urban downloaded data on disclosure quality from the World Bank webpage on January 13, 2021.
- Daniel Urban gathered information on prior index membership from Eikon and web searches between September 2, 2021, and September 14, 2021.
- Daniel Urban gathered information on passive ownership from FactSet and Eikon between September 1, 2021, and September 6, 2021.
- Daniel Urban gathered information on ratings from Eikon and Capital IQ between January 10, 2022, and January 17, 2022.
- Vidhan Goyal gathered information on passive fund flows from EPFR between April 9, 2022, and April 11, 2022.
- Daniel Urban downloaded data on a firm's cross-listing status from Datastream on January 22, 2021.

- Daniel Urban gathered information on creditor rights from Djankov et al. (2007) on May 26, 2022.
  - Vidhan Goyal gathered information on credit default swaps and probabilities of default from the Credit Research Initiative (CRI) at the National University of Singapore on March 2, 2022.
  - Daniel Urban downloaded data on the revised antidirector rights index (Spamann, 2010) and the anti-self-dealing index (Djankov et al., 2008) on March 15, 2023.
  - Daniel Urban obtained stock market data to calculate the Amihud (2002) illiquidity measure from Datastream on December 12, 2023.
  - Daniel Urban downloaded data on creditor information and regulatory enforcement from the World Bank webpage on December 19, 2023.
  - Daniel Urban downloaded data on aggregate investor protection (La Porta et al., 2006) on December 19, 2023.
  - Daniel Urban downloaded data on earnings announcements from I/B/E/S on December 21, 2023.
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 to the paper as all sources of the data have been identified in the paper.

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.*

Section 2 of the paper and Appendix A describe the sample and provide variable definitions. The steps necessary to process the data are described in the computer programs referred to in 6.

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.*

All data manipulations are performed using the computer programs referenced in 6. Request templates for data downloads (e.g., Datastream) are available upon request.

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-bystep 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.*

All of the aforementioned computer programs are included in the data package. The file *Documentation.txt* provides an overview of all code files, which are designed to be self-explanatory. The first eight files contain the steps required to prepare the data, while the remaining files are labeled according to the tables they generate.

The Stata file *MainSample\_Identifiers.dta* contains the firm identifiers (i.e., Datastream Codes) for the final sample used in the main analysis. The Excel file *IndexEvents.xlsx* provides information on the sample of index events used in the research design

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.*

The log of all steps required to produce the data is recorded in the file *LogDataPreparation.smcl*, while all results are logged in *LogResults.smcl*.

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.*

Daniel Urban will maintain the data and programs for at least six years, con-

sistent with National Science Foundation guidelines.

## References

- Amihud, Yakov**, “Illiquidity and stock returns: Cross-section and time-series effects,” *Journal of Financial Markets*, 2002, *5*, 31–56.
- Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer**, “Private Credit in 129 Countries,” *Journal of Financial Economics*, 2007, *12*, 77–99.
- , **Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer**, “The Law and Economics of Self-Dealing,” *Journal of Financial Economics*, 2008, *88* (3), 430–465.
- Leuz, Christian, Dhananjay Nanda, and Peter D Wysocki**, “Earnings Management and Investor Protection: An International Comparison,” *Journal of Financial Economics*, 2003, *69*, 505–527.
- Porta, Rafael La, Florencio Lopez de Silanes, and Andrei Shleifer**, “What Works in Securities Laws?,” *Journal of Finance*, 2006, *61*, 1–32.
- Spamann, Holger**, “The “Antidirector Rights Index” Revisited,” *Review of Financial Studies*, 2010, *23* (2), 467–486.