

Data Description Sheet
“Aggregated Compensation Peer Group Disclosure and Managerial Labor Market Competition: A Network Analysis”

by Ray Rui Gao and Yifei Lu

This data description sheet is provided to fulfill the requirements of the *Journal of Accounting Research* data policy for the submission of a manuscript.

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

Data collection and analyses were primarily handled by Ray Rui Gao and Yifei Lu.

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 list the data sources and dates on which data were downloaded or obtained in the table below. This data description sheet complements variable definitions and sample construction information presented in the paper. Ray Rui Gao and Yifei Lu vouch for the stated sources of the raw data.

Source	Date	Description
ISS Incentive Lab	Nov 2018, Nov 2019	Compensation peer group data, compensation consultant
Execucomp	Nov 2019	Executive compensation data (total pay and pay components), manager characteristics (e.g., age), manager tenure and turnover, job titles and descriptions,
Compustat	Nov 2019, Feb-Apr 2020	Firm characteristics used to calculate dependent and control variables, firm locations, firm segments, Dow Jones index membership
CRSP	Nov 2019, Feb-Apr 2020, Aug 2021, May 2022	Monthly and daily stock returns data, S&P500 index membership
Boardex	Nov 2019	Manager characteristics and board positions (e.g, board chair, age)

Hoberg-Phillips Data Library (https://hobergphillips.tuck.dartmouth.edu/)	Nov 2019	Hoberg and Phillips TNIC industries
Ken French's Data Library (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)	Nov 2019	Fama-French industry classifications
Claudia Custodio	Apr 2020	General Ability Index (Custodio et al., 2013)
I/B/E/S	July 2023	Analyst coverage based peers

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).*

No data were obtained from any organization on a proprietary basis. All data used in the paper are available from the publicly or commercially available sources noted above.

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.*

We describe the detailed steps to construct the sample and variables in Section 3, Section 4.1, and Section 5.1. Appendix A provides detailed variable definitions.

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 use SAS 9.4, UCINET 6.66, and STATA 18 to process the raw data, and use STATA 18, Gephi 0.10.1, and Tableau 2023.2 to conduct the empirical tests.

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.

The code is described in the file "Readme.docx" of the replication package. The programs to convert the raw data are in the "cleaning code" folder. The programs to execute the statistical or econometric analysis and to produce the output used in constructing tables of the manuscript are in the "analysis code" folder. Firm identifiers are listed in the "firm identifiers" folder.

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 file that shows the execution of the entire code is in the "log_files" folder of the replication package.

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 authors agree to maintain the data and programs used in this paper for the six-year time period following the suggestion of the National Science Foundation.