

Data Description Sheet

for the paper

“Beyond Old Boys’ Clubs: Financial Analysts’ Utilization of Professional Connections”

By Mengqiao Du and Rachel Xi Zhang

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

Both authors, Mengqiao Du and Rachel Xi Zhang, 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 can vouch for the stated source of the raw data.

The Appendix in the manuscript provides detailed information on how key variables in the study are constructed from various data sources.

- I/B/E/S database: Quarterly earnings forecasts from 2004 to 2022 were downloaded in April 2023.
- CRSP stock price data: Daily and monthly stock price data were downloaded in April 2023.
- Compustat quarterly financial statement data: Quarterly financial statement data was downloaded in April 2023.
- ExecuComp: Firm and executive linkage downloaded in April 2023.
- Refinitiv StreetEvents: Transcripts on or before 2021 were downloaded between March and April 2022. Transcripts after 2022 were downloaded in April 2023.
- Institutional Investor magazine: All-star analyst name lists were obtained from Institutional Investor LLC in June 2023.
- Capital IQ: People intelligence data were downloaded in May 2023, and earnings call transcripts were downloaded in April and June 2024.

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, 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.

The datasets used in this study are available to anyone with the necessary subscriptions.

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 complete description of the steps necessary to download, obtain, collect, and process the data used in the final analyses reported in the paper can be found in Section 3: Sample construction and variable definitions. The Appendix in the manuscript provides variable definitions and data sources.

In addition, the set of computer programs (i.e., codes) provides clear comments on each step of the data processing and statistical analysis.

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 conducted via computer programs, with the following exceptions.

- As described in Section 3.1 of the manuscript, IBES analysts' full names and gender are manually collected based on LinkedIn profiles, brokerage websites, and media coverage.
- We manually identify brokerage names in transcript text and in Capital IQ and link them to IBES brokerages, taking into account common name variants, name changes, abbreviations, and spelling errors.
- When linking analyst names from transcript text and in Capital IQ to IBES analysts and when linking executive names from transcript text to ExecuComp, we use a script to compute fuzzy name scores before manually inspecting each potential match, taking into account common name variants, abbreviations, spelling errors, suffixes, and titles. The scripted portion is provided in the set of computer programs.
- We manually inspect and remove duplicate transcripts that refer to the same event. This occurs when different programs of the same conference (i.e., the management presentation and the Q&A) are recorded as separate transcripts.

A detailed set of Python, SAS, and Stata codes used to process the raw data and to download the processed data is provided together with this datasheet.

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.

All programs used in this paper are provided alongside this datasheet. Each program provides clear comments and descriptions on the steps involved to either process the raw data into the final dataset or to execute the statistical analyses.

Specifically, the following programs are provided:

- Folder \investor_conference: a set of codes to parse information from conference transcripts
- Folder \professional_experience: a set of codes to collect information on analysts' professional experiences from Capital IQ
- Folder \earnings_call: a set of codes to collect information from Capital IQ earnings call transcripts
- Folder \analyses_code_log_Stata: a set of codes to construct the sample, merge various datasets, and conduct statistical analyses

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 have been provided alongside the computer programs.

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 will ensure that the data and programs are maintained for the six-year period.