

Data Description Sheet

“Do Engagement Quality Reviewers’ Workplace Ties with Engagement Partners Influence Audit Quality?”

Yue Qi, Timothy A. Seidel, Joseph H. Zhang, Junsheng Zhang

March 2026

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

Yue Qi collected the raw historical data. All authors have jointly conducted all the empirical 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 raw data were obtained from the following sources:

- i. The China Stock Market and Accounting Research (CSMAR) database is publicly available. The data used in the paper were downloaded from the database on February 23, 2025. To address missing or potentially inaccurate observations, we cross-validate and supplement these data using the Chinese Research Data Services (CNRDS) database and individual auditor information disclosed on the Chinese Institute of Certified Public Accountants (CICPA) website. The CNRDS data were accessed in February 2025, and the CICPA data were manually collected from the official website on October 12, 2024; June 22, 2023; June 1, 2023; and April 16, 2018.
- ii. The information disclosure platform designated by the China Securities Regulatory Commission (CSRC) (<http://www.cninfo.com.cn/new/index>) is publicly available and serves as a centralized repository where listed companies in China disclose important information. Specifically, we collect listed companies’ announcements related to the reappointment or replacement of audit firms, changes of engagement partners or engagement quality reviewers (EQR), announcements on capital market projects (including annual audits, IPOs, convertible bonds, seasoned equity offerings, asset restructurings, and private placements), and supplementary disclosures on matters related to modified audit opinions. EQR related data were manually collected from these announcements during March-May 2022, March-May 2023, and September 2024. Audit materiality data were manually collected on June 11, 2025.
- iii. The official websites of the CSRC and its local branches, the Ministry of Finance, and the Shanghai and Shenzhen Stock Exchanges are publicly available and provide detailed information on audit violations. The data used in the paper were manually collected from these websites in February 2025.
- iv. The Easy-Board database is publicly available and provides standardized classifications of audit violations. The relevant data were accessed in February 2025.

- v. The Qichacha database is publicly available and provides firm-level information, including historical partner shareholding records. We identify auditor promotions to national equity partnerships by examining these records. The relevant data were manually collected in August 2025.

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.**

We do not use proprietary data.

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.**

The data collection procedures are described in Section 4.1 of the manuscript, and the construction and definition of all variables are described in Appendix D.

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 Stata programs along with limited manual checks to convert the raw data into the final dataset. We have provided the Stata code along with a description of the code.

6. **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 used to convert the raw data into the final dataset, conduct the statistical and econometric analyses, and generate the tables in the manuscript is organized across do-files 1-11. Additionally, we provide a file (identifiers.xlsx) that contains firm identifiers for the final sample.

- 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.**

Comprehensive log files documenting the execution of all Stata programs are provided. These log files cover the full process from raw data processing to final empirical analyses.

- 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 at least six years suggested by the National Science Foundation.