

Data Description Sheet for “Consequences for Culpable Auditors”

This data description sheet is provided in fulfillment of the *Journal of Accounting Research* data policy for submission of a manuscript.

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

- *Li and Park were the primary authors responsible for handling the data. Krishnan and Mehta contributed to the data hand collection.*

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).

- *Data were obtained from public sources, with some obtained through subscription services from Audit Analytics, AAER database, and Compustat. Data from Lexis Nexis Personal Records were obtained by Temple University, with Li designated as the sole licensed user of the database. Data from LinkedIn, Facebook, Google, and enforcement documents were hand-collected by Li. Data from Revelio Labs were handled by Park. All data used in this project were collected between September 2021 and June 2024. All co-authors can vouch for the stated source of the raw data.*

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).

- *The data from LinkedIn and Facebook are proprietary. To access or replicate, researchers would need to create accounts on both platforms. All data are either publicly available or commercially licensed.*

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 followed the steps described below to create a sample of culpable auditors:
1) We first download all enforcement orders directly from the SEC and PCAOB website. We then read each enforcement order to identify instances in which a respondent named in the enforcement action was employed by an accounting firm at the time of the alleged misconduct, and the enforcement was against audit related misconduct.*

- When individual auditors are named in more than one enforcement action related to the same event, we only retain one observation for each person. When multiple individuals are named in the same enforcement order, we treat each individual as a unique observation. We collect details including the enforcement date, the violation year, the nature of the violation, and for each culpable auditor: name, age, job title, employer, and the penalty imposed by the regulator.*
- 2) We use the enforcement orders to identify culpable auditor sample from step (1) that are CPAs and the state where the culpable auditor is located. We search State Board of Accountancy websites for disciplinary orders against each individual in our sample. We review State Board of Accountancy Consent Agreement and Agreed Order and meeting minutes (if available) to identify the sanctioned and financial costs levied on the individual.*
 - 3) To obtain labor market consequences for culpable auditor sample in step (1), we gather employment details from LinkedIn, LexisNexis, Facebook, and Google searches or (in a few cases) calling the audit firm to confirm employment. We use the culpable individual's name, age, job title, employer, state where the culpable auditor is located (all obtained directly from the enforcement orders) as searching parameters. If we cannot verify the individual's post-enforcement employment, we classify the auditor as "retired" if the auditor is within the mandatory retirement age bracket and otherwise as "unknown".*
 - 4) To obtain personal outcomes for culpable auditor sample in step (1), we use the culpable individual's name, age, state where the culpable auditor is located (all obtained directly from the enforcement orders) as searching parameters in LexisNexis. We use individual's CPA license number, employer, and job title to narrow down the search results and identify the right person. We then review the culpable auditors' real estate ownership holdings and identify buying and selling transactions for the three-year window before and after the misconduct enforcement order release date. We collect the details about cash flow effects of these purchases (repurchases) and sales.*
- We followed the steps described below to create a comparison sample of non-culpable auditors using raw data from Revelio:*
- 1) We first download all Position Files from Revelio which contain information on individuals' employer name, region, country, state, employment start date, employment end date, and title.*
 - 2) We then merge all Position Files by limiting individuals that work in United States using the 'country' field.*
 - 3) We then fuzzy match and only keep those that ever work at the top 100 largest professional service firms from Inside Public Accounting (IPA 100).*
 - 4) We then only retain all individuals with the title including the words "Audit",*

“Assurance”, or “Accountant”.

- 5) *We then create indicators for “Partner”, “Manager”, or “Staff” to match with our culpable auditors.*
 - 6) *We then merge data from step (5) with our culpable auditor dataset based on job title, Big 4 or Non-Big 4 membership, state location, and employment at the firm for at least two years before the year in which the culpable auditor faces enforcement.*
 - 7) *Data from step (6) is used to calculate the turnover rate for the comparison sample.*
 - 8) *Data from step (6) is used as a beginning point to provide descriptive evidence on labor market outcome for the comparison sample. We only keep those comparison auditors that depart their firm.*
 - 9) *We then merge those departing comparison auditors’ full employment information from Position Files from Revelio.*
 - 10) *We then classify those departing comparison auditors’ subsequent career outcomes. Please refer to Online Appendix 1 for more details. Data from this step is used to provide evidence on labor market outcomes for the comparison sample.*
 - 11) *Finally, to create a dataset for real estate comparison sample, we start from step (3) above. From this dataset, we create indicators for “Partner”, “Manager”, or “Staff” to match with our culpable auditors.*
 - 12) *We then use User Files from Revelio to merge information on first name, last name, and gender.*
 - 13) *We then use Education Files from Revelio to merge information on the years during which individuals attend their undergraduate institution which we use to calculate their age.*
 - 14) *We then match data from step (13) with our culpable auditor dataset based on job title, Big 4 or Non-Big 4 membership, state location, employment at the firm in the year in which the culpable auditor faces enforcement, gender, and age.*
 - 15) *We use data from step (14) to identify a comparison sample for real-estate holdings.*
 - 16) *We use Lexis Nexis manually to look at their real-estate holdings and transactions.*
- 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 manually collect raw data on culpable auditors from the various sources identified above and organize it into an Excel spreadsheet. This final dataset is then used to provide descriptive evidence on a sample of culpable auditors.*
 - *We used Stata and SAS to convert the raw data into the final dataset used to provide descriptive evidence on a comparison sample of non-culpable auditors. The Stata and SAS codes that transfer raw Revelio data used to create and calculate descriptive evidence on comparison sample is provided.*
- 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 for at least six years.*