

Data Description Sheet for

“Monitoring Quality of Mafia-Connected Accountants”

Pietro A. Bianchi*
Bocconi University

Jere R. Francis
Missouri University

Antonio Marra
Bocconi University

Nicola Pecchiari
Bocconi University

March 4, 2026

In compliance with the *JAR* Data Policy, we provide the following information regarding the empirical data used in the paper.

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

All the four authors were aware of the data collection process. Bianchi, Pecchiari, and Marra handled the data. The analyses were conducted by Bianchi.

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.

Bianchi, Marra, and Pecchiari vouch for the stated sources of the raw data. Bianchi handled the construction of the datasets with the help of Marra and Pecchiari. These three authors have access to all the programs and data.

Our empirical analysis relies on information from various sources:

- a. Financial data are obtained from **AIDA** database of Bureau Van Dijk, which contains financial and accounting information. In June 2014, we sampled the

entire population of private firms located in the Lombardy region of Italy for the period 2006 – 2013.

- b. **CERVED** provided governance data about board composition and identities of external accountants (i.e., members of Board of Statutory Auditors) for the firms we identified in the previous step. We received the data in June 2014.
- c. The Italian Internal Intelligence and Security Agency (“*Agenzia Informazioni e Sicurezza Interna*” or the **AISI**) provided confidential intelligence information on individuals under investigation for mafia-related activities. We submitted to the **AISI** the complete list of directors, executives, and shareholders obtained from **CERVED**. **AISI** returned the list with information about those individuals who were under investigation for mafia-related activities. We received the data in October 2016.
- d. Each firm in Italy has a unique identification number for VAT purposes that we used to merge the **AIDA** and **CERVED** databases. Further, each individual is also assigned a unique number (similar to the social security number in the U.S.) that we used to merge **CERVED** and the list of individuals under investigation provided by the **AISI**.

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 AIDA were accessed through a license obtained by Bocconi University. The data from CERVED were purchased by the authors. The data from AISI were obtained under a confidentiality agreement executed between Bocconi University and the Presidenza del Consiglio dei Ministri (Presidency of the Council of Ministers of the Italian Republic) on July 7, 2014. Pursuant to this agreement, the data may be used solely for research purposes and are subject to strict confidentiality and non-disclosure provisions, including limitations on the dissemination of sensitive or identifiable information. We have provided the editors with the contact person to verify our data and copy of the above confidentiality agreement.

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 our data in the “Data and Research Design” section of the paper and define all variables in Appendix 1. The building code file “01-Clean_Raw_Data_BFMP 2026” also contains comments on the construction and cleaning of the datasets we used.

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 manipulations of the data are completed via computer programs. The code used to convert the raw data into the final dataset, to execute the statistical and econometric analyses, and to generate the tables of the manuscript can be found in “01-Clean_Raw_Data_BFMP 2026”, “02-Build_Work_Data_BFMP 2026”, “03-Tables 2-8_BFMP 2026”, and “03-Tables 9 & 10_BFMP 2026”.

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 used to convert the raw data into the final dataset, to execute the statistical and econometric analyses, and to generate the tables of the manuscript can be found in “01-Clean_Raw_Data_BFMP 2026”, “02-Build_Work_Data_BFMP 2026”, “03-Tables 2-8_BFMP 2026”, and “03-Tables 9 & 10_BFMP 2026”. Additionally, we provide a file with the identifiers of the firms in our final sample in the excel file “identifiers”.

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 execution of the entire code can be found “logfile_BFMP2026”.

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 commit to retaining all data and code for at least six years, in line with NSF guidelines.