

JAR Data Statement

Paper: Strategic (Inconsistent) Disclosures and Sophisticated Investors: Evidence from Hedge Funds

Authors: Yichang Liu, Joshua Madsen, and Frank S. Zhou

Date: June 26, 2025

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

All authors were involved with the data. Joshua and Yichang primarily conducted and oversaw the analysis of the Form ADV data and creation of our primary inconsistency measures. Frank and Yichang handled analysis of the hedge fund return data.

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 data used in this paper were downloaded/obtained the following sources, with approximate dates for when the data was either downloaded or initially processed:

- Form ADV Part 1: <https://www.sec.gov/help/foiadocsinvafoiahtm.html> and <https://adviserinfo.sec.gov/> (initially accessed August 2019)
- Form ADV Part 2: <https://adviserinfo.sec.gov/> and <https://www.sec.gov/foia/docs/form-adv-archive-data.htm#part2> (initially accessed September 2019)
- Thomson Reuter's (now LSEG) Lipper TASS Hedge Fund Database (initially accessed September 2019)
- WRDS Fama French Portfolios and Factors (initially accessed December 2019)
- David Hsieh's Data Library: <https://faculty.fuqua.duke.edu/~dah7/HFData.htm> (initially accessed December 2019)
- CRSP (initially accessed December 2019)
- Laughran and McDonald sentiment word list: <https://sraf.nd.edu/textual-analysis/resources/> (initially accessed November 2019)
- LinkedIn: (initially accessed September 2020)
- FINRA's Broker-Check website: <https://brokercheck.finra.org/> (initially accessed February 2023)

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.*

Not applicable.

To be provided *in the paper* or the *online appendix*:

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.*

Detailed steps used to obtain and process the data are outlined in Sections 2 through 6 of the paper. Variable definitions are available in Appendix A of the paper.

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.*

Please see item 6 below for the code that converts the raw data into the final datasets, executes the empirical analyses and generates the tables and figures of the paper. We are willing to share our measure of disclosure inconsistency upon request, but cannot publicly share data derived from the TASS Hedge Fund Database.

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.

Please refer to “LMZ 2025 Code_raw file.do” for code that converts the raw data into the final datasets used in the analyses, and “LMZ 2025 Code_main.do” for code that executes the empirical analyses and generates the tables and figures of the paper. Notes are provided in the code where necessary. We also provide the identifiers for the sample used in the main regression analyses of the paper.

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.*

Please refer to “LMZ 2025 Code_raw file log.log” and “LMZ 2025 Code_main log.log” for the Stata log files that show the execution of the entire code as mentioned in item 6.

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.*

We will maintain the data and programs for at least six years until 2031.