

**Data Description Sheet**  
**“Hedge Fund Regulation and Fund Governance: Evidence on the Effects of Mandatory Disclosure Rules”**

Colleen Honigsberg  
October 2019

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

*Colleen Honigsberg handled the data and conducted the analyses.*

2. A detailed description of how the raw data were obtained or generated, including data sources, the 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.

*As discussed in Section 3.1, the sample consists primarily of two raw data sources.*

- *Form ADV was obtained by filing Freedom of Information Act (“FOIA”) requests with the SEC and sixteen state securities agencies. In particular, in addition to the SEC, I filed FOIA requests for historical Form ADV filings at the following state securities agencies: CA, CO, CT, FL, GA, IL, MA, MD, MN, NJ, NY, OH, PA, TX, VA, and WA. I began this process in December 2013, and I received data from the various regulators over the period from March 2014 through May 2015. The disclosures that I received were piecemeal (frequently multiple data dumps from the same regulator), so they needed significant cleaning.*

*I ultimately received Part 1 Form ADV filings for all SEC registered investment advisers (and exempt reporting advisers) from 2001-2014 and some Form ADV data for state registered investment advisers from 2006-2014. Although the states were very responsive to my request and I received some data from all the state regulators that I contacted, many states were unable to provide time-series historical data. In most cases, I received a snapshot of the investment advisers that were registered with the state at the beginning of each year from 2006 to 2014, but I did not receive full Form ADV filings.*

*When I had questions regarding the FOIA data that I received, I reviewed the historical data posted on the SEC’s website at <https://www.sec.gov/foia/iareports/inva-archive.htm> and the Investment Adviser Public Disclosure website at <https://www.adviserinfo.sec.gov/>. Researchers can use the data posted on the first link above to replicate my study. Although sufficient information to perform this study was not available through the website when I began this study, the SEC has now made far more data publicly available without a FOIA request. Further, unlike the data that I received, the data available online is now already merged and machine readable.*

- *The Lipper Hedge Fund database (TASS) was obtained through WRDS. I downloaded the entire database from WRDS in early 2014. Please note that funds are allowed to update their information in the Lipper Hedge Fund database. Additionally, funds enter/exit the database. Thus, if you were to download the data today, it likely would not match the data that I downloaded originally. Not only would certain variables have changed (e.g., monthly returns), but the funds in the database would not match. Because the data may differ, I would be happy to provide other researchers with the data I originally downloaded, provided that those researchers already have access to the Lipper Hedge Fund database through a legitimate source.*
- *In addition to these two primary data sources, I relied on the following information received in 2014:*
  - *Monthly Sadka permanent liquidity variable used in Sadka (2006). This variable was received from Prof. Sadka and used to generate the Sadka control variable reflecting the fund's sensitivity to liquidity (shown in the \_Controls do files).*
  - *Hedge Fund Factors used by Fung and Hsieh (2004). Prof. Hsieh provides detailed information on the hedge fund factors on his website: <https://faculty.fuqua.duke.edu/~dah7/HFRFData.htm>. The "trend risk factors" are available for download on the website and updated monthly, the bond market and credit spread factors are available for download from the Federal Reserve, and the equity oriented risk factors are available for download from Datastream (Prof. Hsieh provides specific instructions on his website). These hedge fund factors are merged into the main dta files in the \_Cookie Jar Flag do files.*

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 were not obtained on a proprietary basis.*

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.

*The steps necessary to collect and process the data are provided in Section 3 of the paper. For further detail, see item #2 above, item #5 below, and Exhibits 1-3 in the Online Appendix.*

5. Prior to final acceptance of the paper, the computer program used to convert the raw data into the 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.

*I have provided three folders with the following information:*

***Sample Funds Folder.*** *Three Stata dta files containing the TASS identifiers for the funds used in each of the three time-periods.*

***Stata Code to Generate Primary Datasets Folder.*** *This folder contains the Stata code used to generate the dta files used in the main analysis. The process to create these files is as follows:*

1. *Merge the Form ADV data with the TASS data to generate "raw" dta files.* *The code for this merge is not included because it contains some fund names, and I do not believe TASS allows the funds in its database to be made public. As such, please contact me directly if you are interested in this code.*

*The code merges on adviser name (a field provided in both Form ADV and TASS) after (1) removing all spaces and punctuation, and (2) converting all letters to lowercase. To obtain the best merge, I had to adjust some of the names by hand (e.g., ltd versus limited). These manual adjustments are included in the code.*

*After merging Form ADV and TASS, I created three "raw" dta files (one for each of the three settings, where each file includes only the 60 months analyzed in that setting). The "raw" dta files contain the monthly TASS data along with the following variables derived from Form ADV:*

- *a monthly "registered" variable reflecting whether the fund was registered in that month (as described in the paper, this is derived based on the date the fund first registered with the SEC);*
- *a "NEW" variable reflecting whether the fund is a newly registered treatment fund (you may see this named "reg2006" in the Hedge Fund setting);*
- *a "control" variable reflecting whether the fund is a control fund;*
- *a "deregister" variable indicating whether the fund withdrew following Goldstein (only included for the Goldstein setting);*
- *a "remain" variable indicating whether the fund remained registered following Goldstein (only included for the Goldstein setting); and*

- an “ERA” variable indicating whether the fund was an exempt reporting adviser (only included for the Dodd-Frank setting).

The “raw” dta files also contain a “Post” variable set to 1 in the 30-months after the event, and to 0 in the 30-months prior to the event.

2. Generate the control variables using “raw” dta files. I used the raw dta files to generate the controls. Please see the code provided in the Generate Control Variables folder.
3. Generate the measures of misreporting using “raw” dta files. I used the raw dta files to generate the flags for misreporting. Please see the code provided in the Generate Measures of Misreporting folder.
4. Generate the six dta files for the main analysis. To create the primary dta files used in the analysis, I merged the two measures of misreporting to create “FLAGS” dta files (the FLAGS dta files contain both misreporting flags). This merge is included in the “Cookie Jar Flag” do files. Upon completion, I had two primary dta files for each of the three settings: (1) a dta file with the control variables, and (2) a dta file with the measures of misreporting (i.e., six dta files in total). The six dta files are named as follows: (1) HF\_FLAGS.dta, (2) HF\_Controls.dta, (3) Gold\_FLAGS.dta, (4) Gold\_Controls.dta, (5) DF\_FLAGS.dta, and (6) DF\_Controls.dta.

**Log File of Analysis Folder.** The Stata log file presents the Stata code and corresponding output for the analysis in the paper. Each table is labeled. The log file relies primarily on the six dta files generated above. However, you will notice the additional dta files listed below. These dta files are supplementary and used in various robustness tests.

*Extended data for the 3-period model used in columns (5) and (6) of Table 3*

- 3Period\_Controls.dta. This dta file includes control variables for the 3-period analysis. Please see the code included in the Generate Control Variables folder.
- 3Period\_FLAGS.dta. This dta file includes measures of misreporting for the 3-period analysis. Please see the code included in the Generate Measures of Misreporting folder.

*Datasets containing subsets of matched funds*

- HF\_PSM.dta/Gold\_PSM.dta/DF\_PSM.dta. These dta files include the matched subsets of funds used in Tables 2 and 3. Please see the code in the Matching folder within the Stata Code to Generate Primary Datasets folder to see how these matched subsets were generated.

*Additional TASS data (no code; only adjustments from raw data were to add “modate” variable in some files reflecting month-year date and/or to reshape the data):*

- FullAuditStatus.dta. TASS data reflecting historical audit dates.
- FullProviderStatus.dta. TASS data reflecting historical auditor.
- HF\_FundDeath.dta/DF\_FundDeath.dta. TASS data reflecting full monthly variables for all funds included in the primary analysis.

- *HF\_Append.dta/DF\_Append.dta. TASS data reflecting monthly variables for funds available in TASS but not included in the primary analysis (i.e., funds lacking the relevant sixty months of data required for the balanced panel).*

*Merged Form ADV and TASS data (these all contain merged TASS and Form ADV data, but they are separate dta files from one another and from the “raw” dta files above due to the nature of the FOIA data I received; however, there’s no need for this piecemeal approach if you use the Form ADV now posted on the SEC’s website; no code provided publicly due to aforementioned confidentiality issue, so please contact me directly):*

- *HF\_Custody.dta/DF\_Custody.dta. dta files containing an indicator for whether the fund had custody of client assets in each period (custody variable from Form ADV); one observation per fund in each period.*
- *HF\_RedFlags.dta/DF\_RedFlags.dta. dta files containing the “red flag” indicators for each fund in each period (indicators from Form ADV); one observation per fund in each period.*
- *DF\_ADV.dta. dta file containing each fund’s reported assets under management and main office country in the Dodd-Frank period (variables from Form ADV); one observation per fund in each period.*

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.

*Colleen will retain original data and programs for the required time span.*