

## **Data Description Sheet for “Anonymous Equity Research”**

Travis Dyer  
Cornell University

Eunjee Kim  
Cornell University

February, 2021

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

Both Travis Dyer and Eunjee Kim were aware of the data collection process and involved in the empirical analysis of the final dataset.

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.*

We have obtained data from multiple sources.

- a. We obtain firm-level financials from Compustat, stock returns from CRSP, institutional ownership and insider trading from Thomson Reuters, analyst forecasts from I/B/E/S, and media coverage from RavenPack.
- b. We obtain Seeking Alpha articles written from January 2006 to October 2018 from Seeking Alpha (SA). This dataset includes the date of publication, author names, tagged stock tickers, and the text content of articles.
- c. We collected detailed information about author biographical profiles (e.g., author names, company names, Twitter accounts, LinkedIn accounts, biographical description) from the SA website.

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).*

Travis Dyer signed a Non-Disclosure Agreement with Seeking Alpha and obtained Seeking Alpha data from Itai Blitzer ([itai@seekingalpha.com](mailto:itai@seekingalpha.com)). Travis is not allowed to share the confidential data that he received from Seeking Alpha with others. There are no restrictions to the publication of research that uses the confidential data.

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.*

Raw data were obtained from the sources listed in Item #2 above and imported into SAS or Stata to be included in the final dataset. We describe the processes used to create the main dataset in Section 3.1 of the paper.

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. The 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.*

The SAS, PERL, and STATA programs that produce the final dataset and perform analyses are provided along with this datasheet. The firm and quarter identifiers for the observations included in our final sample are reported in "DK\_identifiers.xlsx."

6. *Maintenance of data and programs*

The authors will maintain all data and programs for at least six years.