

Data Log for “Leveraging big data to study information dissemination of material firm events”

Bin Li
Mohan Venkatachalam

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

Bin Li handled the data and performed the analyses. Mohan Venkatachalam dealt with the data on 8-K filings.

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 obtained proprietary refinery foot traffic data from Orbital Insight, Inc. on April 29, 2019 and August 7, 2019.

We obtained REFO data from Bloomberg terminal on January 17, 2020 and from Orbital Insight on April 29, 2019 and August 7, 2019.

We obtained media coverage data from Factiva between March 26, 2021 and April 17, 2021.

We obtained 8-K filing data from DirectEdgar on September 6, 2019.

We obtained the PADD-level acquisition cost and gas inventory data from the EIA website on March 3, 2020 (<https://www.eia.gov/petroleum/data.php#summary>).

Our research assistant, Mark Yue Ma (University of Oklahoma), helped us scrape data on city-level gasoline prices from Gasbuddy website on March 5, 2020 and March 31, 2020 (<http://fuelinsights.gasbuddy.com/Charts>).

We obtained all financial statement data from Compustat for the period 2016–2019 on September 21, 2020 and May 4, 2021.

We obtained all stock return data from CRSP for the period July 2015–December 2020 on September 20, 2020 and July 12, 2021.

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).

We obtained proprietary data from Orbital Insight under a Non-Disclosure and Confidentiality Agreement signed by Bin Li. Note that this data was obtained when Bin Li was a faculty at the University of Oklahoma. We have privately provided the editors with a copy of the Agreement and the contact information for a company representative from Orbital Insight. All other data are publicly available from the sources listed above.

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.

All relevant steps are described in *Section 3 “Data and Sample”* and *Section 4 “Research Design and Empirical Analyses”* in the manuscript and outlined in the computer programs described below.

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.

The SAS program “*big data sas codes - Final*” details the steps from converting the raw data into the main samples used in this paper. We do not provide identifiers because of the Non-Disclosure and Confidentiality Agreement with Orbital Insight.

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.

The authors agree to maintain the data and programs used in this paper for the six-year period suggested by the National Science Foundation.