Run EDGAR Run: SEC dissemination in a high-frequency world

Jonathan L. Rogers
University of Colorado

Douglas J. Skinner
University of Chicago

Sarah L. C. Zechman
University of Colorado

Compliance with Data Policy for the Journal of Accounting Research
April 2017

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

All data was obtained as described in #2 below and was handled and analyzed by Jonathan Rogers and Sarah Zechman. The details on data handling and analyses are described in #2 below.

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.

We have used the following data sources for our analyses:

a. Thomson-Reuters Insider Trading, TAQ, CRSP, and Compustat data were last downloaded from WRDS by Sarah Zechman and Jonathan Rogers on the following dates. Thomson-Reuters insider trade data was downloaded in September 2015. CRSP was downloaded in February 2015. Compustat data was downloaded in June 2014. TAQ was downloaded between April 2014 and May 2015.

b. Dow Jones archive dataset was obtained from RavenPack by Sarah Zechman. The data were downloaded in June 2014.

c. SEC header dataset was scraped from the SEC filings on the EDGAR website by an undergraduate research assistant in two phases. The first set of header data were scraped June 2014 with an update to the data in February 2015.

d. FTP date-times were scraped from the SEC's FTP site by an undergraduate research assistant in May 2015.

e. PDS feed times and SEC post times were obtained from a PDS subscriber. The data were received in March 2014.

All data were processed jointly by both Jonathan Rogers and Sarah Zechman.

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 agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).

The PDS feed times and SEC post times described in 2e above were provided by a PDS subscriber who requested anonymity and provided the data on a proprietary basis. We have communicated the requested details to the editors.

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 papers, we require information about subject eligibility and/or selection, as well as any exclusion criteria.

We describe our data in the paper and the detailed data steps.

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. Instead of the program, researchers can provide a detailed step-by-step description that enables other researchers to arrive at the same dataset used in the analysis. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the sample was formed, including the treatment of outliers, Winsorization, truncation, etc. This programming is in most circumstances not proprietary. However, we recognize that some parts of the data generation process may indeed be proprietary or otherwise cannot be made publicly available. In such cases, the authors should inform the editors upon submission, so that the editors can consider an exemption from this requirement.

We use a combination of SAS and Stata programs to convert the raw data and perform all analyses. We have provided a detailed description of the steps involved. In addition, we have provided details on our final sample.

6. Data and programs should be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

The data and programs will be maintained for at least six years as required.