Linguistic Complexity in Firm Disclosures: Obfuscation or Information?

Brian J. Bushee
The Wharton School
University of Pennsylvania

Ian D. Gow
University of Melbourne

Daniel J. Taylor
The Wharton School
University of Pennsylvania

Voluntary Compliance with Data Policy for the Journal of Accounting Research

Our paper was submitted in January of 2014, prior to JAR's adoption of the mandatory data policy. Nevertheless, we provide the following information regarding the empirical data used in our paper.

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

Gow and Taylor handled and analyzed the data.

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 use data from a variety of sources in our primary analysis. All data is available with subscription through either Wharton Research Data Services (hereafter WRDS) or Thomson StreetEvents (hereafter StreetEvents). Gow is responsible for the StreetEvents data that was obtained through his previous institutional affiliation with Harvard, and Taylor is responsible for the remainder of the data.

StreetEvents. Harvard maintains a local copy of all conference call transcripts from StreetEvents. We initially processed the conference call transcripts in August 2013, with a follow-up in November 2014.

IBES Files. We downloaded the IBES Unadjusted Summary File (WRDS: STATSUMU_EPSUS), Unadjusted Actual File (WRDS: ACTU_EPSUS), Unadjusted Pricing File (WRDS: ACTPSUMU_EPSUS), and CRSP-IBES link file (WRDS: ICLINK) from WRDS in December 2013. Subsequent analysis required the company issued guidance file (WRDS: DET_GUIDANCE) which was downloaded April 2015.

Compustat Files. We downloaded quarterly and annual Compustat files (WRDS: FUNDQ and WRDS: FUNDA) in May 2012 with a follow-up in May 2013. Subsequent analysis required the Compustat segments file (WRDS: WRDS_SEGMERGED) which was downloaded October 2016

CRSP files. We downloaded the CRSP daily stock returns file (WRDS: DSF) in February 2013.

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).

All of the source data for this project are available with subscription to WRDS and Thomson StreetEvents.

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.

We describe our data in Section 3 of the paper. For further details see item #2 above and item #5 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.

Step #1: Construct conference call data. Use natural language processing to process the respective portions of the StreetEvents data. We use ticker and date provided in StreetEvents to fetch CRSP identifier (PERMNO) and Compustat identifier (GVKEY).

Step #2: Clean and merge IBES. We merge the IBES Unadjusted Summary File, IBES Unadjusted Actual File and IBES Unadjusted Pricing File together. We retain the last consensus forecast prior to the earnings announcement, and retain observations where the earnings announcement occurred between zero and three months after the fiscal period end. We use the CRSP-IBES link file to fetch PERMNO, and merge IBES data to the conference call data based on PERMNO, earnings announcement date, and fiscal period end.

Step #3: Merge with Compustat Quarterly and Annual files based on GVKEY, earnings announcement date, and fiscal period end date to compute financial statement-based control variables.

Step #4: Merge with CRSP daily file on PERMNO, earnings announcement date, and fiscal period end date to compute market-based control variables and measures of information asymmetry.

Step #5: Add supplementary data on conference call transcripts (e.g., jargon, tone, etc.) and clean the dataset.

Step #6: Begin analysis using final dataset.

The SAS code used to generate the sample in the paper (Steps 1-5) are included with this document. Please cite the paper is you use any of the code or data.

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.

Gow and Taylor will retain all original data and programs for the required six years.

Gow and Taylor will periodically update the code and data for ongoing changes in database composition subsequent to publication. The most recent version of the code, data, and documentation is available at https://github.com/iangow/bgt. An updated SAS file with firm identifiers, earnings announcement dates, measures of linguistic complexity, and measures of additional characteristics of language on the call (number of words, jargon words, tone, forward-looking statements, etc.) is also available upon request (email dtayl@wharton.upenn.edu).