Data Description Sheet

The Dog that Didn't Bark:

Limited Price Efficiency and Strategic Nondisclosure

Frank Zhou and Yuqing Zhou

January 2020

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

   Frank Zhou and Yuqing Zhou have both handled all data and jointly conducted all the empirical 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.*

   The final sample is constructed by Frank Zhou and Yuqing Zhou. We combine three sources of raw data: IBES Guidance, downloaded from WRDS on 03/10/2018, Compustat-CRSP Merged, downloaded from WRDS on 02/27/2018, CRSP, downloaded from WRDS on 10/16/2018, RavenPack, downloaded from WRDS on 12/10/2018, Markit, downloaded from WRDS on 11/10/2018.

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 data can be obtained from WRDS.

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 sample periods and steps to clean the raw data are described below. Please refer to the README file under the Code folder for more details about the codes for each step.

(a) IBES Guidance

Below is the blueprint for the sample construction of IBES Guidance. The most important part is to assign management guidance to the correct fiscal quarter.

i. Import IBES Guidance 2003-2017.

ii. Drop non-US firms.

iii. Obtain PERMNO using IBES-CRSP linking file. We need to use PERMNOs to merged IBES with Compustat-CRSP merged data (CCM), so we drop firms that do not have PERMNO.

(b) Compustat-CRSP Merged (CCM)

i. Import CCM annual 2003-2017.

ii. Keep firms that have valid tickers using IBES-CRSP linking file.

iii. Merge with the file from the previous step.

To minimize the presence of irregular firms (for example, firms that restated earnings), we delete firms that have cases of larger than a 120-day gap between adjacent fiscal-quarter end dates or firms that have instances of larger than a 180-day gap or smaller than a 28-day gap between adjacent earnings announcement dates. We require firms to have IBES tickers and nonmissing earnings announcement dates throughout the sample period.

(c) CRSP

i. Import CRSP 2003-2017 and keep stock return information.

ii. Merge with IBES Guidance and CCM by the date of the stock returns.

(d) RavenPack

   i. Import Dow Jones Newswires articles data from RavenPack 2003-2017.

   ii. Merge with IBES Guidance, CCM and CRSP.

(e) Markit

   i. Import short-selling cost data from Markit 2003-2017.

   ii. Merge with IBES Guidance, CCM and CRSP.

(f) Construct the dependent and independent variables

   i. Generate cumulative returns from 5 days before the earnings announcement date to 2 days after the earnings announcement date.

   ii. Generate other independent variables (see the codes).

(g) Data cleaning

   i. Keep earnings related forecasts

   ii. Keep the earliest earnings related forecasts

   iii. Drop pre-announcements, that is, forecasts issued after the fiscal quarter end date, because these forecasts are typically pre-announced earnings instead of forward-looking guidance.

   iv. If a quarter only has pre-announcements, the quarter will be dropped and will not appear in our sample.

   v. Drop missing control variables and return outliers

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.

Our code folder contains eight folders, which should be run sequentially to produce the final sample. The folder also contains a README file that describes the input files to be used and the output files to be generated. A list of the identifiers (PERMNO and DATADATE) for the final sample can be found in the file Identifiers.dta.

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 time period suggested by the National Science Foundation.