Data description form for manuscript "Non-GAAP Earnings Disclosure in Loss Firms" by Edith Leung and David Veenman

Following the *JAR* data policies, this document contains detailed information on the data collection and handling for this manuscript. The (STATA) programming code and Perl code used for our text search are available at:
https://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements.

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

The data used in this study were either hand-collected from 8-K filings or retrieved from WRDS (Compustat, CRSP, I/B/E/S). The non-GAAP data were hand-collected from 8-K filings by both co-authors (David Veenman and Edith Leung) and two trained research assistants (Faissal El Allaoui and Gijs van Vugt). Data from WRDS were retrieved by David Veenman. For tests comparing non-GAAP reporting frequencies based on our text-search with those from Bentley, Christensen, Gee, and Whipple (2018) (BCGW), we obtained their non-GAAP data in February 2018 from: https://sites.google.com/view/kurthgee/data. In April 2018, we used these data to conduct an additional test suggested by the reviewer, to corroborate the results of one of our tests based on a random sample of GAAP profit firms (see point 2b below). The BCGW data were downloaded by David Veenman. The analyses were mostly conducted by David Veenman. All data and code are stored and have been reviewed by both co-authors.

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.

a. Hand-collected non-GAAP data

We hand-collected non-GAAP earnings data from quarterly earnings releases in 8-K filings (item 2.02), for firms with a quarterly GAAP loss in the period 2006-2014. Our sample selection procedures are described in Table 1 of the paper. We obtained all 8-K earnings releases for this sample from SEC EDGAR and used a Perl keyword search to identify non-GAAP earnings disclosures. Appendix A of the manuscript contains a detailed description of our manual data collection procedures. As described in Appendix A, we collected the data in two rounds, with the help of two trained research assistants in round two. We provide more details on which co-author and/or research assistant handled the data below.

Round one (August 2015): In the first round of data collection, we coded 3,629 filings with a high likelihood of being "loss converters" (see Appendix A for details). We first ensure consistent coding across co-authors by both independently coding the same set of 100 press releases. We

discussed any differences in our coding and agreed on a common approach. We equally split the remaining number of press releases and proceeded to collect data, consulting each other on ambiguous cases. After both authors had finished collecting and checking the data for their own sample, we exchanged the datasets to verify each other's data for potential errors. Specifically, we checked observations for which there were indications of errors: (1) when there was a deviation between the GAAP earnings from Compustat (*IB*) and GAAP earnings retrieved from the filing, (2) when the sign of collected GAAP earnings was not negative, (3) when we did not find non-GAAP data in the press release. We also randomly inspected 10 percent of the collected data for accuracy.

Round two (March – May 2016): In the second round, two research assistants (Faissal El Allaoui and Gijs van Vugt) aided in collecting data for the remaining 2,779 press releases identified as containing non-GAAP earnings (see Appendix A). We first held an instruction session to instruct our research assistants on how to collect the data and coded ten press releases together with them. We then gave them a training sample of 79 randomly selected press releases that we had also coded, so we could assess their ability to correctly identify and collect the necessary data. After providing them with detailed feedback on any mistakes they made, we gave each research assistant 1,350 press releases to code. The research assistants used the same procedures explained above to verify the accuracy of their coding after they were finished. Finally, both co-authors also checked the datasets for errors.

Data for profitable firms (July – August 2016): For the additional analyses of non-GAAP reporting in profitable firms, added in the first revision of the manuscript, we randomly selected a sample of 4,000 firm-quarter observations for which income before extraordinary items and operating earnings were positive, and for which our text-search indicated a high likelihood of non-GAAP reporting in the 8-K filing (i.e., *Non-GAAP trigger* equal to 1). Our research assistant Faissal El Allaoui then read and coded the press releases, identifying 3,589 press releases containing non-GAAP earnings. He applied the same procedures explained above to verify the accuracy of his coding and then both co-authors checked the datasets for errors.

b. BCGW non-GAAP data for profit firms (April 2018)

We replicated the results for our randomly selected sample of profit firm-quarter observations (see above) using the BCGW data. We used the non-GAAP earnings from their dataset for all firm-quarter observations for which income before extraordinary items and operating earnings are positive.

c. WRDS data

The following table explains how we obtain the data needed for the steps outlined in Table 1 and the additional data needed for the analyses (all data were retrieved by David Veenman), alphabetical order:

File name	**Description**	**Source**	**Date**
bcgw_2018_ng_data_2003to 2016.dta	Dataset from Bentley et al. (JAR2018)	https://sites.google.com/view /kurthgee/data	2/1/2018
compustat_age_data.dta	Financial statement data from Compustat used to determine firm age	WRDS → Compustat-Capital IQ → North America → Fundamentals Annual	4/8/2016
crspm.dta and crspm_2015_2017.dta	File containing end-of-month stock price and return data from CRSP	WRDS → CRSP → Monthly Stock File	originally retrieved on 8/6/2015, updated on 7/3/2017
crspm_vwretd.dta	File containing the value-weighted return including distributions (VWRETD)	WRDS → CRSP → Monthly Stock File	10/26/2015
erdport1.dta	Daily size-adjusted returns based on CRSP's NYSE/AMEX/NASDAQ breakpoints	Obtained from WRDS server through SAS/SSH ('/wrds/crsp/sasdata/a_indexes')	4/14/2016
extra_data_2016.dta	Additional quarterly financial statement data from Compustat Quarterly files, retrieved later to ensure complete data availability for future performance tests	WRDS → Compustat-Capital IQ → North America → Fundamentals Quarterly	4/6/2016
extra_data_2017.dta	Additional quarterly financial statement data from CRSP/Compustat Merged files, retrieved to calculate 12-quarter ahead cash flow from operations for extra tests	WRDS → Compustat-Capital IQ → North America → Fundamentals Quarterly	6/8/2017
ff_factors_4.dta	Data for Fama-French four-factor model regressions (Table 7)	WRDS → Fama-French Portfolios and Factors → Factors - Monthly Frequency	7/3/2017
fundq.dta	Quarterly financial statement data from CRSP/Compustat Merged files	WRDS → CRSP → CRSP/Compustat Merged → Fundamentals Quarterly	8/6/2015

ibes_actu.dta	File containing actual value of quarterly EPS from I/B/E/S	WRDS → I/B/E/S → Unadjusted Actual (measure="EPS", pdicity="QTR")	8/6/2015
ibes_actu_gaap.dta	File containing actual value of quarterly GAAP EPS from I/B/E/S	WRDS → I/B/E/S → Unadjusted Actual (measure="GPS", pdicity="QTR")	8/6/2015
ibes_statsumu.dta	File containing number of estimates (numest) from I/B/E/S, used to determine analyst coverage	WRDS → I/B/E/S → Unadjusted Summary Statistics (measure="EPS", FPI="6")	8/6/2015
ibes_statsumu_eps.dta	File containing number of estimates, median and mean forecast value of EPS from I/B/E/S, used for the ERC tests	WRDS → I/B/E/S → Unadjusted Summary Statistics (measure="EPS", FPI="6")	4/11/2016
ibes_statsumu_gps.dta	File containing number of estimates, median and mean forecast value of GAAP EPS from I/B/E/S, used for the ERC tests	WRDS → I/B/E/S → Unadjusted Summary Statistics (measure="GPS", FPI="6")	4/11/2016
items8k.dta	List of all 8-K filings between 2006 and 2014, used to identify filings of earnings press releases (item 2.02)	WRDS → SEC Analytics Suite → SEC Filings Index	8/6/2015
xtravar.dta	Additional quarterly financial statement data from Compustat Quarterly files, used to calculate historical earnings volatility and sales growth	WRDS → Compustat-Capital IQ → North America → Fundamentals Quarterly	10/20/2015

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.

Not applicable.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper.

We describe the details of the data collection and processing as carefully as possible in the body of the manuscript, Table 1, and Appendix A, and in point 2 above.

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. … Whenever feasible, authors should also provide the identifiers (e.g., CIK, CUSIP) for their final sample.

We provide the full STATA programming code used to convert the raw data into the final dataset used in the analysis ("lv_jar_stata_code.do"), as well as the Perl code used for our text search of 8-K earnings press releases. We share the final sample of 59821 firm-quarters (see Table 1) with the hand-collected non-GAAP variables attached: "lv_jar_dataset.dta". We cannot share variables in the final dataset that are based on data purchased from data providers such as Compustat or I/B/E/S.

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.

Both authors agree to store and maintain the data and programs for at least six years following publication of the paper.