

Data Description

“Truncating Optimism”

By Zachary Kaplan, Xiumin Martin and Yifang Xie

June 2021

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

The base code was developed by Zachary Kaplan and Yifang Xie. Yifang Xie conducted all the analyses. Zachary Kaplan reviewed Yifang’s code prior to submission.

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 raw data are obtained from various vendors that make their data available on Wharton Research Data Services (WRDS) or from other sources as noted below. All data were downloaded at various times between June 2016 and September 2020.

- Individual forecasts data were downloaded in July 2016 from the I/B/E/S Unadjusted Detail—History file on WRDS.
- Earnings announcement information were downloaded in July 2016 from the I/B/E/S Unadjusted Detail—Actuals on WRDS.
- Analyst consensus data were downloaded in July 2016 from the I/B/E/S Unadjusted Summary History—Summary Statistics on WRDS.
- Forecast removal information were downloaded in June 2017 from the I/B/E/S Unadjusted Detail—Excluded Estimates on WRDS.
- Stopped estimates information were downloaded in June 2017 from the I/B/E/S Unadjusted Detail—Stopped Estimates on WRDS.
- Management guidance data were downloaded in August 2017 from the I/B/E/S Guidance on WRDS.
- Stock price data and stock split information were download in August 2017 from the CRSP Daily Stock file on WRDS.
- Information on firm characteristics were downloaded in August 2017 from COMPUSTAT Fundamentals Quarterly on WRDS.
- Investor relation officers data are from conference calls and were shared by Kimball Chapman

in November 2017.

- Information on S&P 500 index were downloaded in March 2018 from COMPUSTAT Index Constituents on WRDS.
- Executive compensation data were downloaded in September 2020 from COMPUSTAT EXECUCOMP—Annual Compensation on WRDS.
- Insider trading data were downloaded in September 2020 from Thomson Insiders Data on WRDS.

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).

No data were obtained from any organizations on a proprietary basis.

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 raw data used for this study were obtained from sources described in Item #2. Sections 3 and 4 describes the procedures to construct the samples and conduct the final analyses.

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.

We constructed the datasets for analyses using SAS. All continuous variables are winsorized in STATA using WINSOR2. All the regression analyses are conducted in STATA using REGHDFE. Section 3, Section 4, and Appendix B of the submitted manuscript describes the procedures of sample construction and variable definitions. We also provide the following code to fulfill JAR's Data and Code Sharing Policy:

- “KMX_code.sas”: This SAS program produces the samples for our empirical analyses.
- “Additional_Input.sas”: This SAS program produces two additional datasets. One dataset serves as an additional input for the construction of the variable *Optimism*. The other is used for the construction of the variable *Compensation Delta*.
- “identifier.dta”: This STATA dataset contains I/B/E/S ticker as the identifier code for our 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.

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