

Data Description Sheet

Consensus? An Examination of Differences in Earnings Information Across Forecast Data Providers

Stephannie Larocque
University of Notre Dame

Jessica Watkins
University of Notre Dame

Eric Weisbrod
University of Kansas

In compliance with the JAR Data Policy, we provide the following information regarding the empirical data used in the paper.

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

Data collection and analyses were primarily handled by Jessica Watkins and Eric Weisbrod. Stephannie Larocque managed the data collection from Bloomberg and the Nexis Uni media-citation counts with support from Notre Dame graduate student research assistants. Notre Dame data analysts Brandon Greenawalt and James Ng provided research assistance with running and validating several of the data collection scripts (FactSet, Capital IQ).

2. A detailed description of how the raw data were obtained or generated, including data sources, the specific 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.

In the paper, we examine differences in the earnings data produced by Bloomberg, I/B/E/S, S&P Capital IQ, FactSet, and Zacks. All of our data are obtained from commercially available sources. Bloomberg, Capital IQ, and FactSet data were obtained directly from each respective data provider. I/B/E/S and Zacks data were obtained via WRDS. Additional WRDS data needed to construct our variables were obtained from Compustat, CRSP, RavenPack (RPA 1.0), TAQMSEC, TAQ IID, and Thomson Reuters CDA/Spectrum Institutional Holdings (s34). Media-citation counts used in Figure 2 were hand-collected by Stephannie Larocque from Nexis Uni. Table 1 in the paper summarizes

our sample construction procedure and Appendix C provides the details of each variable definition.

Bloomberg data was collected during 2020-2022 (final pull: bql_market_e_and_a 2022-05-17, bql_f 2022-05-10). Nexis Uni citation counts were collected during 2024. The remaining data download dates are as follows:

Source	Files	Date
WRDS Compustat	fundq	2025-07-17
WRDS Compustat	funda	2025-07-24
WRDS Compustat	company	2025-12-09
WRDS CRSP	ccmxpf_linktable, stocknames, dsf	2025-07-17
WRDS CRSP	dsix	2025-07-24
WRDS I/B/E/S	surpsum, surpsumu, statsum_epsus, statsumu_epsus, actu_epsus, ibes_adj, iclink	2025-07-18
WRDS I/B/E/S guidance	det_guidance	2025-06-24
WRDS Zacks	eps_surp, company_info	2025-07-18
WRDS Capital IQ link	wrds_ciqsymbol	2025-07-19
Capital IQ Xpressfeed	ciqEstimateConsensus, ciqEstimatePeriod, ciqEstimatePrimaryEarnings, ciqTradingItem, ciqSecurity, ciqEstimateAnalysisData, ciqEstimateNumericData_YEAR	2025-06-16
FactSet Snowflake	quarterly_consensus_basic_YEAR (1997-2024)	2025-08-28
WRDS TAQMSEC	wct_YYYYMMDD (daily, 2004-2025)	2025-07-30 through 2026-01-14
WRDS TAQMSEC	taqmclink	2025-07-24
WRDS TAQ IID	iid_msec_YEAR (2003-2025)	2024-04-05
WRDS RavenPack	rpa_full_equities_YEAR (2002-2024)	2025-08-22
WRDS Thomson Reuters	WRDS_IO_TimeSeries (Inst. Ownership)	2025-12-13

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).

Not applicable. All data are obtained from commercially available sources accessible through standard academic subscriptions (WRDS, FactSet, Bloomberg, Capital IQ Xpressfeed) or hand-collected from public sources (Nexis Uni).

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 sample in Section 3 of the paper and define all variables in Appendix A. Table 1 summarizes the sample selection procedure. The complete set of steps necessary to process the raw data into the final dataset is documented in the computer programs referred to in part 5, with per-script Inputs/Outputs sections and the README.md / README.txt files in the code package.

5. After downloading or obtaining the raw data, all manipulations of the data should be done via computer programs. The code for these manipulations should be included in the code submitted upon acceptance (see below). No manipulations of raw data can take place manually or outside the computer code provided. If compliance with this requirement is not feasible, the authors need to explain and disclose any manipulations of the raw data (e.g., manually created variables or file conversions). When feasible, we also encourage the authors to share the code that downloads the data.

The replication package (consensus.zip) contains all R, SAS, and Stata code used to convert the raw data into the final dataset, execute the regression analyses, and produce all tables and figures reported in the paper. No manual manipulation of raw data was performed.

6. 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.

The replication package (consensus.zip) contains all R, SAS, and Stata code used to convert the raw data into the final dataset, execute the regression analyses, and produce all tables and figures reported in the paper. src/run-all.R orchestrates the full sixteen-step pipeline (001- through 016-); a per-script log is written to log/. Identifiers for the firm-quarters in the final regression sample are provided in output/sample-identifiers.csv (94,030 rows; gvkey, permno, cusip, FDP-specific identifiers, and announcement timestamps). The package's README.md / README.txt documents prerequisites, the .env configuration, and the pipeline structure. The code repository is also publicly hosted at <https://github.com/eweisbrod/consensus>.

7. A comprehensive log file that shows the execution of the entire code. This log file should cover all the steps that convert the raw data into a final dataset and the execution of all statistical and econometric analyses presented in the tables of the manuscript. The portion of the log file that shows proprietary code or data may be masked. In this case, the reader should be referred to the step-by-step description provided as per the requirements in Item 6.

Comprehensive log files have been provided in the replication package (consensus.zip) in the log subfolder.

8. 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 at least six years, consistent with National Science Foundation guidelines.