Data Description Sheet

Intangible investments, scaling, and the trend in the accrual-cash flow association

Jeremiah Green, Henock Louis, and Jalal Sani

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

Jalal Sani conducted data cleaning and analyses.

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.

Raw data used in the final sample are from Compustat (data downloaded on May 1, 2021), CRSP (data downloaded on May 28, 2021), and Fama-French Portfolios and Factors (data downloaded on May 28, 2021). These databases were accessed through Wharton Research Data Services (WRDS). Mispricing scores of Stambaugh, Yu, and Yuan [2015] were downloaded from Robert Stambaugh's website (http://finance.wharton.upenn.edu/~stambaug/) (data downloaded on May 28, 2021).

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 agreement, any restrictions imposed by the organization on the authors with respect to publishing certain results).

Not applicable to the paper. All data are publicly available.

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 papers, we require information about subject eligibility and/or selection, as well as any exclusion criteria.

Section 3.3 of the paper provides detailed sample construction procedures. The variable definitions appendix in the paper and the online appendix provide definitions of variables used in the analyses.

5. Prior to final acceptance of the paper, the computer program used to convert the raw data into the dataset used in the analysis plus a brief description that enables other researchers to use this program. Instead of the program, researchers can provide a detailed step-by-step description that enables other researchers to arrive at the same dataset used in the analysis. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the sample was formed, including the treatment of outliers, Winsorization, truncation, etc. This programming is in most circumstances not proprietary. However, we

recognize that some parts of the data generation process may indeed be proprietary or otherwise cannot be made publicly available. In such cases, the authors should inform the editors upon submission, so that the editors can consider an exemption from this 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 use a SAS program to convert the raw data into the final dataset used in the analyses and a STATA program to perform the analyses reported in the paper and the online appendix. We provide computer programs that replicate our analyses, a description of these programs, and the identifiers (GVKEYs) for all firms used in our study.

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.

We will maintain all data and programs for at least six years.