

Journal of Accounting Research Data Description Sheet for “Public Information, Relative Overconfidence, and Capital Flows”

Balakrishnan, Bernard, Rennekamp, and Steenhoven

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

Darren Bernard handled the archival data and conducted the analyses.

Kristi Rennekamp and Blake Steenhoven handled the experimental data and conducted the 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.

Karthik Balakrishnan and Darren Bernard purchased and received Trackmaster (an Equibase subsidiary) Race and Start files for 1991 - 1995 racing results on June 8, 2017. Both sets of data are freely available for purchase (i.e., neither is proprietary). Primary contact at Trackmaster was David Siegel.

Darren Bernard retrieved microfiche of 1991-1995 racing results (with some exceptions) on July 13, 2017 from Keeneland Library. Primary contact at Keeneland Library was Becky Ryder (now retired) and Roda Ferraro. Karthik Balakrishnan and Darren Bernard used outside services to convert first batch of microfiche to PDF in September and October 2017 and retrieve WPS betting handle information from PDFs of first batch of microfiche (completed January 1, 2018). Karthik Balakrishnan and Darren Bernard used outside services to convert second batch of microfiche to PDF in December 2018 and retrieve WPS betting handle information from PDFs of second batch of microfiche (completed February 19, 2019).

Darren Bernard retrieved Beyer Speed Figure information from paper copies of the Daily Racing Form held by a Seattle, WA-area collector in March, April, and May of 2015.

Trackmaster and converted microfiche data are used in Tables 1-7. Beyer Speed Figure data are used in Table 6.

We collected all experimental data via Qualtrics surveys. Participants were recruited through the Cornell Business Simulation Lab, and completed the study between November 29th, 2022 and December 7th, 2022.

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 proprietary data are used in this paper.

4. A complete description of the steps necessary to download, obtain or collect as well as 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.

Section 3.3 of the manuscript describes the steps necessary to download, obtain, collect, and process the data used in the final analyses reported in the paper.

Section 4 of the manuscript provides detailed description of the data collection processes for the experiment. Our experiment was conducted in Qualtrics using the instrument provided in the online appendix.

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.

All data manipulations for the archival data are conducted via computer programs, and the relevant code is submitted (see response to #6).

For the experimental data, we downloaded the raw data file from Qualtrics and made four manual adjustments based on communication from the BSL lab manager. In total, Qualtrics recorded 127 responses to our study. Of these, two responses (Participant #222 and #381844) were removed because the “responses” were actually from research assistants clicking through to reset the study for subsequent participants. One additional “response” (Participant

#922019) was removed because a single participant was given two different participant numbers by a research assistant, and so Participant #922019 was a duplicate of Participant #922919 (where we retained the latter in the data). This resulted in the sample of N=124 reported in the paper. Our last manual adjustment related to two different participants who were inadvertently given the same participant number (#914555) to enter in Qualtrics for payment purposes. As these represented two distinct participants, we manually re-numbered one of these participants to #754128 (the next random number on our generated list of participant IDs), so that we could tell them apart. Finally, as Qualtrics output lists each participant as a separate line, we convert the dataset into stacked data with each participant-round as a separate line to allow JMP to calculate participant random effects.

6. The computer programs (i.e., code) used to (1) convert the raw data into the final dataset used in the analysis, (2) to execute the statistical or econometric analysis, and (3) to generate the tables or to produce the output used in constructing tables of the manuscript. A brief description that enables other researchers to understand and run the code should be provided. 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, and which commands were used in the analysis, 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 disclosing the proprietary portion 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 results that the authors obtained and presented in their manuscript. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption allowing the step-by-step description. Whenever feasible, authors are required to 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 codes used to generate the final dataset are: “_JAR Stata code start data.do”, “_JAR Stata code main data.do”, and “_JAR Stata code Beyer data.do”. The codes used to execute analysis and generate the tables in the manuscript are “_JAR Stata code main data.do” and “_JAR Stata code Beyer data.do”. The code includes extensive comments to explain each step and help researchers understand and run the code.

Data is automatically generated via Qualtrics based on participants' decisions during the experiment and exported to Excel. Following the adjustments noted in

Response #5, data are imported to JMP. Code for analyses conducted are included in the JMP log file "Experiment_DataAnalysisLog.jsl".

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.

Log files showing the execution of the Stata code have been provided.

A log file showing the execution of the JMP code for analysis of the experimental data has been provided.

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.

We will ensure that the data and programs are maintained for the six-year period.