"Corporate Scandals and Regulation"
Compliance with Data Policy for JAR

Luzi Hail
The Wharton School, University of Pennsylvania

Ahmed Tahoun
London Business School

Clare Wang
Tippie College of Business, University of Iowa

January 2018

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

   All authors were involved in the collection as well as the analysis of the data. All data used for analysis is stored centrally so that all authors were able to access the data at will. Data handling and analyses were performed as described in more detail below.

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.*

   We have used the following data for our analysis:

   a. Scandals and regulation data (individual incidents and numbers of media mentions): the authors and local country research assistants hand-collected all the data between October 2015 and July 2017 as described in Section 2.2 of the paper. We also outline the detailed data collection procedures and sources by individual country in the Internet appendix available at either https://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements or http://ssrn.com/abstract=2960069.

   b. Annual *GDP per Capita* (in 1990 international dollars): downloaded from the Maddison-Project (http://www.ggdc.net/maddison/maddison-project/home.htm, 2013 version) on April 5, 2016. To fill in the years 2011 to 2015 (for which the above source does not provide data), we used data from the Conference Board Total Economy Database (http://www.conference-board.org/data/economydatabase, November 2016 version) downloaded on January 29, 2017. We adjusted these recent years (130 observations) to the 1990 international dollars base used in the Maddison-Project data.

   c. Annual *Inflation* data: downloaded from the online data set accompanying Reinhart and Rogoff [2011] (https://www.aeaweb.org/articles?id=10.1257/aer.101.5.1676) on March 8, 2016. To fill in the years 2009 to 2015 (for which the above source does not

provide data), we used data from the IMF (http://www.imf.org/en/Data) downloaded on January 29, 2017. For Israel (which is not covered in Reinhart and Rogoff [2011]), we used data from the World Bank (https://data.worldbank.org/data-catalog/world-development-indicators) downloaded on January 29, 2017.

d. Annual *Financial Crisis* data: downloaded from the online data set accompanying Reinhart and Rogoff [2011] (https://www.aeaweb.org/articles?id=10.1257/aer.101.5.1676) on March 8, 2016. We did not update this data set for the final years of our sample (i.e., 2009 to 2015), but assume no crises instead. Reinhart and Rogoff [2011] do not cover Israel in their analysis and, hence, the country is excluded from specifications requiring *Financial Crisis* data.

As described in Footnote 26 of the paper, data for *GDP per Capita* and *Inflation* are missing for some of the historical country-years. We linearly extrapolate missing years from adjacent observations to fill gaps in the time series. If the gaps are too long (i.e., > 10 years) or precede the available data, we backfill the missing data by using the annual percentage changes from a peer country with high observed correlations in subsequent years. For example, we backfill Germany's *GDP per Capita* data before 1850 using the U.K.'s growth rates in *GDP per capita* for the same period. We chose the U.K. because the correlation between the two countries is 96 percent over the 1850 to 1880 period.

In total, we filled-in 539 country-years (including the 130 observations mentioned in b above) with *GDP per Capita* (13.2 percent of total observations) and 145 country-years with *Inflation* data (3.6 percent). For replication purposes, we provide the Stata dataset "inflation_gdp_est.dta" containing the manually filled-in *GDP per Capita* and *Inflation* observations for our sample.

All authors vouch for the stated sources of the raw data.

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).*

   All of the source data for this project came from publicly available sources.

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.*

   We describe our data steps in section 2.2 of the paper. We also outline the detailed data collection and sources by individual country in the Internet appendix.

   The scandals and regulation data used in the analysis is available here (subject to an initial data embargo until May 2019): https://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements.

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.*

   We use Stata to convert the raw data and perform all the analyses. The Stata do-file "HTW_analyses.do" (plus two separate macro do-files containing the code of the main analyses) are provided in the folder "STATA code." The analyses do-file prepares the data for analysis and outputs the descriptive statistics (i.e., Tables 3 and 4 in the paper) as well as the main analyses (i.e., Tables 5 and 6 in the paper).

6. *Data and programs should be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.*

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