Data Description Sheet for

Does Private Country-by-Country Reporting Deter Tax Avoidance and Income Shifting? Evidence from BEPS Action Item 13

Preetika Joshi
Desautels Faculty of Management
McGill University
Preetika.joshi@mcgill.ca

April 12th, 2020

In compliance with the *JAR* Data Policy, I provide the following information regarding the empirical data used in the JAR publication entitled "Does Private Country by Country Reporting Deter Tax Avoidance & Income Shifting? Evidence from BEPS Action Item 13".

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

   This is a single-authored paper, and I collected the raw data, collated the data and conducted all empirical 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.

   I have obtained the data from multiple sources and I can vouch for the sources of the raw data.

   a) *The primary source of the raw data is Orbis. Orbis is Bureau van Dijk's flagship database of private and listed companies.*
   b) *Information on cash taxes paid was collected from the Compustat database. To match observations between Orbis and Compustat, I hand-matched firms using the ISIN number as well by matching total assets and pre-tax income.*
   c) *Statutory tax rates and tax spending information was collected from OECD's tax database (*https://www.oecd.org/tax/tax-policy/tax-database/).
   d) *The public pressure index was constructed using google hits which were obtained from the google trends database and this database is available via the google search engine (*https://trends.google.com/trends/?geo=US).
   e) *To collect information on the Factiva hits, I used the Factiva database. The Factiva database is a research tool maintained by Dow Jones and this database aggregates content from both licensed and free sources and provides organizations with search, alerting, dissemination, and other information management capabilities.*

   The data were downloaded on the following specific dates:

   2018: May 25th, May 27th, June 7th, June 13th, September 25th, October 9th, November 10th

   2019: February 12th, March 3rd, June 12th, June 19th, December 9th

   2020: January 10th

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 study and all the data came from publicly available sources and, aside from the Google Trends Data, were purchased from the vendors referenced above.

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.

   I use STATA to manipulate the data and merge datasets. An overview of the steps necessary to collect and process the data has been included in sections 3.0 and 3.1 as well as in Table 1 in the paper. I provide further details under point (5) below.

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.

   The STATA code used to convert the dataset is attached. For data collected or merged manually, I include notes in the STATA.do file describing how I obtained the data. The ORBIS identifier for my final sample (which is not limited to EU headquartered MNCs) is reported in "Identifier_BVD ID Number.txt".

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.

   I will maintain all the data and programs used in this study until at least 2026.