Data Description Sheet[1] for "Transparency and Tax Evasion: Evidence from the Foreign Account Tax Compliance Act (FATCA)"

Lisa De Simone
Stanford University

Rebecca Lester
Stanford University

Kevin Markle
University of Iowa

November 13, 2019

[1] In line with the Data and Code Sharing Policy for the Journal of Accounting Research as of December 2014 (updated November 2016): https://research.chicagobooth.edu/-/media/research/arc/docs/journal/updated-data-policy-for-jar.pdf?la=en&hash=CD1D669E679E9649CEA9F89948F616E84989DEE4

1. Authors

All three authors were aware of the data collection process and involved in the analysis of the final data set. Markle was in charge of the initial data collection and of the merging of data from various sources for the main tests with foreign portfolio investment (FPI) as the dependent variable. De Simone was in charge of the data collection and analysis of the expatriations and real estate data. Lester was in charge of data collection and analysis of the data on art and UCITS. De Simone and Lester also oversaw data analyses for the Hong Kong pension funds and cryptocurrency, respectively; while we do not include the results in the paper and only discuss in one footnote, we include the data references here for completeness.

2. Data Sources

Most of the primary data collection for TIC and CPIS data was done in Spring 2016. The calculations following Zucman [2013] were done in Summer 2017. Real estate, art, UCITS, Hong Kong pension funds, and bitcoin data were collected between October 2017 and January 2019.
We used various data sources for our analysis:

1) **TIC DATA (for inbound FPI tests)**
   - Source: United States Treasury website (https://www.treasury.gov/resource-center/data-chart-center/tic/Pages/ticsec2.aspx)
   - We downloaded .csv files of all available data for our sample period. We then converted all files to SAS format for merging into one dataset. There was no manual entry at any point in the process.

2) **CPIS DATA (for worldwide FPI tests)**
   - Source: International Monetary Fund website (http://data.imf.org/?sk=B981B4E3-4E58-467E-9B90-9DE0C3367363&sId=1424963554286)
   - We downloaded Table 11 (Geographic Breakdown of Total Portfolio Investment Assets) for all available years. We then converted all files to SAS format for merging into one dataset. There was no manual entry at any point in the process.

3) **ADJUSTED CPIS AND ZUCMAN GAP DATA:**
   We used the downloaded CPIS data to calculate these measures. To do so, we followed the code posted by Gabriel Zucman (http://gabriel-zucman.eu/) and updated the code accordingly for our sample period.
   Sources:
   - Netherlands SFI data: https://www.dnb.nl/en/statistics/statistics-dnb/balance-of-payments-and-international-investment-position/index.jsp
   - Exchange rates from www.oanda.com
   - IMF International Investment Position (IIP) data: http://data.imf.org/?sk=7A51304B-6426-40C0-83DD-CA473CA1FD52&sId=1409773422141

4) **TAX RATE DATA:**
   - We obtain the top statutory individual tax rate in each country-year. As no single source had complete data, we compile our dataset from the University of Michigan, the OECD, KPMG, the Tax Foundation, and Trading Economics.

5) **WORLD BANK DATA:**
   - We use the following data items from the World Bank to construct control variables:
     1. Government effectiveness – from Worldwide Governance Indicators Project
     2. Population
     3. Gross Domestic Product
     4. Number of mobile phone subscriptions per 100 people

6) **MATCHING:**
   - All datasets are merged using country ISO codes and year.

7) **EXPATRIATIONS DATA:**
   - Source: Quarterly Publication of Individuals, Who Have Chosen to Expatriate, as Required by Section 6039G from the Federal Register (https://www.federalregister.gov/agencies/internal-revenue-service)

8) **REAL ESTATE DATA:**
   We compiled several sources for the country-year values of the global house price index (real and nominal), housing rents, and disposable income per person. Sources:
   - Economist (2017), "Global House Prices" (https://www.economist.com/blogs/graphicdetail/2017/03/daily-chart-6) accessed on April 18, 2018
   - OECD (2018), "Prices: Analytical house price indicators", Main Economic Indicators database (http://dx.doi.org/10.1787/cbcc2905-en) accessed on April 18, 2018
   - Bank for International Settlements, Residential Property Prices from FRED, Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/QHUN628BIS) accessed on April 26, 2018.

9) **ART MARKET DATA:**
   - Source: Swiss Federal Customs Administration FCA website, the Swiss-Impex. https://www.gate.ezv.admin.ch/swissimpex/. Cite originally accessed in August 2017 and again accessed in April, May, and June 2018.
   - Source: Geneva Freeport Financial Statements and Powerpoint Presentation from June 2016. http://geneva-freeports.ch/en/ Cite originally accessed in January 2019.

10) **UCITS DATA:**
   - Source: European Fund and Asset Management Association, publications of European Quarterly Statistical Releases. This page includes reports from 2011 to 2017. http://www.efama.org/statistics/SitePages/EuropeanQuarterlyStatisticalRelease.aspx
   - Reports for the period preceding 2011 were found through internet searches.

11) **HONG KONG PENSION FUND DATA (results not tabulated in paper):**
   - Source: Mandatory Provident Fund Schemes Authority public registers (http://www.mpfa.org.hk/eng/public_registers/orso_schemes/orsearch.do)

12) **CRYPTOCURRENCY DATA (results not tabulated in paper):**
   - Results not tabulated primary due to inability to verify quality and accuracy of the country-level data.

- Data sources considered include Coin Dance, accessed here: https://coin.dance/volume/localbitcoins

3. Proprietary Data

All of the data above are publicly available, with the exception of the Swiss customs data. These data are available through the Swiss Impex site. This site provides a free basic subscription, but we paid for the "expert subscription" over a two year period so as to access additional details and the data used in this paper. Our understanding is that these data are available for any individual that purchases the expert subscription.

No survey instrument was used, nor was any experiment conducted.

4. Collection and processing of data

Raw data were obtained from the sources listed in Item #2 above and imported into SAS or Stata to be included in the final dataset. Our unit of observation is country-year, and we list all countries included in our sample in Table 1 of the paper. There is no separate file of identifiers.

5. Code to convert raw data into the final dataset

We describe the processes used to create the main dataset in Section 4.2 of the paper. The SAS program that produces the dataset, DLM Data Code.sas, is being provided at the same time as this data description.

6. Maintenance of data and programs

We agree to maintain data and programs used in this paper for the six-year time period suggested by the National Science Foundation.

References

ZUCMAN, G. 'The Missing Wealth of Nations: Are Europe and The U.S. Net Debtors or Net Creditors? Data Appendix.' Paris School of Economics.