Data Guide for Coppola, Maggiori, Neiman, and Schreger: "Redrawing the Map of Global Capital Flows"

This Version: V1.0 (June 2019)

The following is a list of the raw input files used in the issuer-level ultimate parent aggregation procedure of Coppola et al. (2019), together with short descriptions of the data. The procedure associates the universe of traded equity and debt securities with their issuer's ultimate parent. All paths are relative to the root data folder referred to as *DATA_PATH* in the code. Please see the accompanying README for more details on how these data are deployed. The procedure can be run with a subset of these files, as outlined in the file `UP_Aggregation.py` in the code repository. However, we strongly recommend using all the aggregation sources if possible, since the quality of the final mapping will deteriorate as sources are removed.

The accompanying GitHub repository contains a sample directory structure corresponding to the one described in this guide. The files in the repository are sample versions that only include a limited number of rows — these files are included in order to demonstrate the structure of the data.

1. CUSIP Global Services (CGS) Data

- *Paths*:
  1. `raw/cgs_master/ALLMASTER_ISSUER.PIP.zip`
  2. `raw/cgs_master/INCMSTR.PIP.zip`
  3. `raw/cgs_master/ALLMASTER_ISIN.PIP.gz`
  4. `raw/cgs_master/AIMASTER.PIP.zip`
  5. `raw/cgs_master/CPMASTER_ATTRIBUTE.PIP.zip`
  6. `raw/cgs_master/CPMASTER_ISSUE.PIP.zip`
  7. `raw/cgs_master/CPMASTER_ISSUER.PIP.zip`
  8. `raw/cgs_master/FFAPlusMASTER.PIP.zip`
  9. `raw/cgs_master/TBA Master File [Vintage].zip`
  10. `raw/cgs_lei_plus/CBRLEIMSTR.PIP`
  11. `raw/cgs_master/master_[Vintage].GM`
  12. `raw/cgs_master/master_[Vintage].SB`
  13. `raw/cgs_master/master_[Vintage].FM`
  14. `raw/cgs_master/master_[Vintage].FD`
  15. `raw/cgs_master/master_[Vintage].IB`
  16. `raw/cgs_master/previous_versions/*`
- *Description:* The files listed here are the security- and issuer-level master files for global CUSIP-bearing securities that are provided by CUSIP Global Services (CGS). These files can be obtained commercially from CGS (cusip.com).

- File (1) is the main issuer-level master file from CGS and provides data from the CUSIP_db and CINS_db issuer master file products.
- Files (2) and (3) are the main security-level master files from CGS and provide data from the CUSIP_db, ISIN_db, and CINS_db issue master file products.
- File (4) is the CGS Associated Issuers master file, which provides information about relationships among various CUSIP6 issuer numbers – this is useful in establishing which CUSIP6 numbers belong to the same issuing entities.
- Files (5) through (7) contain data for commercial paper issuers and issues.
- File (8) contains security-level information on 144a issues.
- File (9) contains security-level information on TBA issues.
- File (10) constitutes the CGS LEI Plus mapping data product by CGS, which provides information mapping CUSIPs to Legal Entity Identifiers (LEIs).
- Files (11) through (15) contain security-level information on agency and sovranational issues: file (11) contains GNMA issues; file (12) contains SBA issues; file (13) contains FNMA issues; file (14) contains FHLMC issues; and file (15) contains World Bank issues.
- The files under path (16) are previous vintages of the same CGS master files that we received during the course of the project. The build code consolidates these old versions and the new versions in order to catch any potential data that may have been excluded from more recent issues. However, this is not essential and the code can be easily adapted to run with only the latest version of the CGS data.

2. Dealogic Data

- *Paths*:
  1. `raw/dealogic/stata/companylistings.dta`
  2. `raw/dealogic/stata/company.dta`
  3. `raw/dealogic/stata/dcmdealtranchesisins.dta`
  4. `raw/dealogic/stata/dcmdealtranches.dta`
  5. `raw/dealogic/stata/dcmdeal.dta`
  6. `raw/dealogic/stata/dcmdealtranchesvalue.dta`
- *Description:* All files correspond to tables in the Dealogic Debt Capital Markets (DCM) database, which can be purchased directly from Dealogic (dealogic.com). The raw Dealogic data come in the form of SAS files, and should be converted to the format above using software such as StatTransfer.

3. SDC Data

- *Paths*:
  1. `raw/SDC/bonds/[SubsamplePeriod].[xls/xlsx]`

2. `raw/SDC/loans/[SubsamplePeriod].[xls/xlsx]`
3. `raw/SDC/equities/Equities_[SubsamplePeriod].[xls/xlsx]`

- *Description:* These files constitute a bulk download of the Refinitiv SDC Platinum New Issues database (refinitiv.com), and were downloaded directly using the SDC Platinum software platform. The data were downloaded separately for the three sections of the New Issues database: "All Equity", "All Bonds", and "All Syndicated Loans". The data were downloaded in a series of files that cover separate time periods (for example, in half-year increments for the most recent bond data) — hence the notation [SubsamplePeriod] in the path structures highlighted above. The particular subsample partitioning is however not essential, and the data could be downloaded using any partitioning.

4. Factset Data

- *Paths*:
  1. `raw/factset/factset_ultimate_parents.xlsx`
  2. `raw/factset/HKG_Companies.xlsx`
  3. `raw/factset/LUX_Companies.xlsx`
- *Description:*
  - These files contain data from Factset's Debt Capital Structure database (factset.com) and were built using the Factset Excel add-in.
  - File (1) starts from the universe of CUSIPs observed in the CUSIP Global Services (CGS) data. This list of CUSIPs is included in the leftmost column of the file. The file then queries Factset for information about the ultimate parent of the issuer of each particular CUSIP, and ultimately link issuers to the CUSIPs of their ultimate parents, and the country of headquarters of their ultimate parents. The sample file included in the GitHub repository illustrates the formulas used to build the data.
  - Files (2) and (3) contain a list of public companies and corresponding equity CUSIP codes that are confirmed to be domiciled in Hong Kong and Luxembourg, respectively. We built these lists by using the Factset equity screening tool. Taking Hong Kong as an example, we search for companies for which both the firm's headquarters and the firm's ultimate parent's headquarters are located in Hong Kong, as reported by Factset. We sort firms by market capitalization, and select the 40 largest ones that meet our criteria for inclusion in file (2). The construction of the Luxembourg list in file (3) is analogous.

5. Capital IQ Data

- *Paths*:
  1. `raw/ciq/wrds_cusip.dta`
  2. `raw/ciq/ciq_ultimate_parents.csv`

- *Description:*
  - These files contain the data from Capital IQ (capitaliq.com) that is used in the procedure. We outline below how each of the files can be obtained.
  - File (1) is a list of all the CUSIP identifiers in Capital IQ (CIQ), mapped to the corresponding CIQ internal identifiers. This file can be downloaded directly from the WRDS (wrds-web.wharton.upenn.edu) Unix server. It comes in SAS format and should be converted to Stata format using software such as StatTransfer.
  - File (2) provides a link from each company in file (1) to its ultimate parent, inclusive of the parent's primary CUSIP, CIQID, and country of incorporation. A single-row sample version of the file with mock data is included in the replication packet. Ultimate parents and their CIQ IDs were downloaded for all the CUSIP-linked CIQIDs via CIQ's excel plug-in. Specifically in the CIQ excel plug-in, the formula to retrieve/download the ultimate parent's name is =CIQ($A2, "IQ_ULT_PARENT") wither $A2 refers to the cell where the entity's CIQID lives. The formula for the ultimate parent's CIQID is=CIQ($A2, "IQ_ULT_PARENT_CIQID").

6. Bureau van Dijk Data

- *Paths*:
  1. `raw/orbis/header/BvDIDChange.dta`
  2. `raw/orbis/header/ISIN_BvDID.dta`
  3. `raw/orbis/header/LEI_details.dta`
  4. `raw/orbis/ownership_data/[CountryCode]/SHARE_[CountryCode]_Link_allyrs.dta`
- *Description:*
  - These files contain corporate ownership data from the Orbis database provided by Bureau van Dijk (Bvd). The Orbis data can be sourced directly from BvD (bvdinfo.com).
  - Files (1) through (3) contain reference header data. File (1) contains a history of changes in the internal identifiers used by BvD. File (2) contains a mapping from BvD ID to ISIN, and file (3) contains a mapping from BvD ID to Legal Entity Identifier (LEI).
  - Item (4) shows the path structure for the Orbis ownership files. Under the path `raw/orbis/ownership_data`, there should be one folder for each of the countries included in Orbis (the placeholder [CountryCode] above refers to ISO2 codes). The SHARE_* files in these folders contain all historical subsidiary-shareholder links for which the subsidiary is located in the country designated by [CountryCode].

7. Morningstar Holdings Data

- *Paths*:
  - `raw/morningstar_holdings/[US/NonUS]_[YYYY]_m_step3.dta`

- In order to build a country assignment for each issuer from the Morningstar fund reports, we use the Morningstar holdings data for mutual funds and ETFs described in Maggiori, Neiman, and Schreger (2019). Please see that paper for details on construction of the Morningstar holdings sample. In particular, the raw files that are used in the procedure are referred to as *"step3"* files in the build code of MNS. The raw holdings data can be purchased from Morningstar (morningstar.com).

8. OpenFIGI API

- *Description:* If necessary, the code automatically downloads certain data from Bloomberg's OpenFIGI API. In order for this download to work, users should register with OpenFIGI and request an API key, which should be entered in the file `figi_download/figi_download.py`.

9. Utility files

- *Paths*:

  1. `raw/macro/Concordances/iso2_iso3.dta`

- *Description:*

  – The file number (1) is a simple mapping from ISO2 country codes to ISO3 country codes. The version of the file available in the GitHub repository contains the full mapping used in the procedure.

References

Coppola, Antonio, Matteo Maggiori, Brent Neiman, and Jesse Schreger, "Redrawing the Map of Global Capital Flows: The Role of Cross-Border Financing and Tax Havens," *Working Paper*, 2019.