

# Guide to Replication Data for Maggiori, Neiman, and Schreger: “International Currencies and Capital Allocation”

June 20, 2019

The following is a list of the raw input files required to run the MNS build and analysis, together with short descriptions of the data. All paths are relative to the master data folder referred to as *mns\_data*. Please see the accompanying build and analysis README files for more details on how these data are deployed.

Files marked with [\*] are publicly available and are included in the accompanying replication packet. Files not marked with [\*] must be purchased from the respective data providers; in these cases the list below provides information as to how the data can be sourced. For some of these files, in the cases in which the structure of the data might be ambiguous, the replication packet includes sample version of the files that only include one row with mock data — these files are included in order to demonstrate the structure of the data. We indicate below the cases in which we include such mock files. The path structure outlined below is the one that is referenced in the build and analysis code.

## 1. Morningstar Holdings Data

- *Paths:* The raw Morningstar holdings data come as a series of .7z archives which are collectively stored at the path `raw/morningstar_ftp_master`. Each archive has holdings data for a given month, a particular universe of funds, a geographic region, and a given fund activity status. There are two geographic regions that are used to split the raw data: US and non-US (NonUS). There are three universes of funds: open-end mutual funds (with universe code FE), open-end money market funds (FM), and exchange-traded funds (FE). Lastly, there are two activity status codes used by Morningstar: active, and inactive. The .7z archives should have names of the form

[US/NonUS]\_[FO/FM/FE]\_[Active/Inactive]\_[YYYY]-[MM].7z.

So, for example, the archive `US_FO_Active_2000-01.7z` would contain data for active US open-end mutual funds for January 2000. Each of these archives should contain several XML files, one per portfolio. The XML files use the naming convention `[MasterPortfolioId].xml`, where `MasterPortfolioId` is the unique alphanumeric portfolio identifier used internally by Morningstar. These files are processed and converted to Stata format by the code under `build/xmlimport`.

- *Description:* These files are the raw mutual fund and ETF holdings data from Morningstar. The data can be purchased from Morningstar ([morningstar.com](http://morningstar.com)).

## 2. Morningstar Reference Data

- *Paths:*
  1. `raw/morningstar_direct_supplementary/Mapping_Booth_20181220.xlsx`
  2. `raw/morningstar_direct_supplementary/secid_all_monthly_returns.csv`

3. `raw/morningstar_direct_supplementary/secid_all_monthly_flows.csv`
4. `raw/morningstar_direct_supplementary/all_indices_20180820.xlsx`
5. `raw/Datafeed_Delivery_Documentation/Categories_Asset_Class.xls` [\*]

- *Description:*

- These files provide metadata for the Morningstar holdings data: file (1) provides static fund-level information; file (2) provides a time series of monthly fund returns; file (3) provides a time series of monthly fund flows; file (4) provides a time series of index returns; file (5) provides a mapping from the internal Morningstar security typecodes to the security classification used in the present build and analysis.
- The files (1) and (5) were delivered directly by Morningstar, and can be sourced from the company ([morningstar.com](http://morningstar.com)). The files (2) and (3) were downloaded using the open-end mutual fund section of the Morningstar Direct software platform, while the file (4) was downloaded using the return indices section of the Morningstar Direct software platform. We include single-row mock files for all of these in order to demonstrate the structure of the data, as well as a full version of file (5).

### 3. CPIS Data [\*]

- *Paths:* All files included in the replication packet at `raw/CPIS`.
- *Description:* This dataset is a bulk download of the CPIS time series data for the external assets and external liabilities of many IMF member countries. The data was obtained from the IMF's website.

### 4. TIC Data [\*]

- *Paths:* All files included in the replication packet at `raw/TIC`.
- *Description:* These files are copies of the statistical reports provided by the U.S. Treasury's Treasury International Capital (TIC) system. Some of the files included in the replication packet have been digitized from PDF copies of TIC's reports. These data provide a statistical account of U.S. external assets and external liabilities. The files include various annual-frequency vintages of the TIC tables used in the paper.

### 5. ICI Data [\*]

- *Paths:* All files included in the replication packet at `raw/ICI`.
- *Description:* These files are digitized versions of key tables provided by the Investment Company Institute (ICI), which give details on the size and structure of worldwide asset management markets.

### 6. OECD Data [\*]

- *Paths:* All files included in the replication packet at `raw/OECD`.
- *Description:* These files provide selected indicators from the OECD's statistical warehouse, and were downloaded directly from the OECD's website.

## 7. Dealogic Data

- *Paths:*

1. raw/dealogic/stata/companylistings.dta
2. raw/dealogic/stata/company.dta
3. raw/dealogic/stata/dcmdealtranchesins.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](http://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.

## 8. 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](http://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. Sample single-row files with mock data are included in order to demonstrate the structure of the data, separately for bonds, equities, and loans.

## 9. Factset Data

- *Paths:*

1. raw/factset/factset\_ultimate\_parents\_MS.xlsx
2. raw/factset/factset\_ultimate\_parents\_CGS.xlsx
3. raw/factset/factset\_etf\_aum.xlsx
4. raw/factset/Factset\_FX\_EOY\_2017.xlsx

- *Description:*

- These files contain data from Factset ([factset.com](http://factset.com)) and were built using the Factset Excel add-in. Sample single-row files with mock data are included in the replication packet — these contain Excel formulas that demonstrate how to build these files. Note that file (3) has a horizontal layout and hence we provide a single-column version rather than a single-row version.
- Files (1) and (2) start from the universe of CUSIPs observed in the Morningstar and in the CUSIP Global Services (CGS) data, respectively. These lists of CUSIPs are included in the leftmost column of the files. The files then query 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.
- File (3) provides a time series of assets under management for a sample of ETFs, while file (4) provides a sample of exchange rate data.

## 10. Capital IQ Data

- *Paths:*

1. raw/ciq/wrds\_cusip.dta
2. raw/ciq/wrds\_gvkey.dta
3. raw/ciq/cusip6\_ultimateparents.csv
4. raw/ciq/ciqcountry.dta
5. raw/ciq/ciqcompanytype.dta
6. raw/ciq/ciqcompanysic.dta
7. raw/ciq/gics\_structure.xls [\*]
8. raw/ciq/ciq\_fundamentals\_for\_probits.xlsx

- *Description:*

- These files contain the data from Capital IQ ([capitaliq.com](http://capitaliq.com)) that is used in the paper. We outline below how each of the files can be obtained.
- Files (1) and (2) are lists of all the CUSIP and GVKEY identifiers in Capital IQ (CIQ), mapped to the corresponding CIQ internal identifiers. These files can be downloaded directly from the WRDS ([wrds-web.wharton.upenn.edu](http://wrds-web.wharton.upenn.edu)) Unix server. They come in SAS format and should be converted to Stata format using software such as StatTransfer.
- File (3) 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 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”).

- Files (4) and (5) are mapping from CIQ internal codes to country names and company types. They can be downloaded directly from the WRDS Unix server and should be converted using StatTransfer.
- File (6) provides industry classifications (SIC codes) for each company in CIQ. It can be downloaded directly from the WRDS Unix server and should be converted using StatTransfer.
- File (7) is a spreadsheet detailing the structure of the GICS classification codes. It is effective as of September 2016. The file is included in the replication packet.
- File (8) contains fundamentals data downloaded from Capital IQ for a subset of companies. This file should be assembled using the Capital IQ Excel plug-in, and a sample single-row version of the file with mock data is provided to illustrate the structure of the data. The four leftmost columns, with CUSIP and CIQ identifiers, should be populated using the data produced by the code at the path `temp/probits/ciq_for_retrieval.xlsx`. The columns containing fundamental data can then be populated via a query on CIQ ID using the CIQ Excel plug-in.

## 11. BIS Data [\*]

- *Paths:* All files included in the replication packet at `raw/BIS`.
- *Description:* This dataset was obtained directly from the BIS website and contains a bulk download of the BIS International Debt Statistics.

## 12. 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](http://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.

### 13. Datastream Data

- *Paths:*
  1. `raw/macro/Datastream/WMRExRatetoUSDList.xlsx [*]`
  2. `raw/macro/Datastream/WMReuters_Exchange_Rates_to_USD[Vintage].xlsm`
- *Description:* These files contain exchange rate time series obtained from Refinitiv's Datastream ([refinitiv.com](http://refinitiv.com)). File (1) is included in the replication packet and provides a list of the series identifiers for the data downloaded from Datastream, while file (2) contains mock data showing the structure of the output. The code consolidates various vintages of the downloaded data, but can easily be adapted to only use one version of the data.

## 14. 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_all yrs.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](http://bvdinfo.com)). Sample single-row versions of these files with mock data are provided in order to illustrate the structure of the data.
  - 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]`.

## 15. Labeled Training Set for Fuzzy Merge Procedure [\*]

- *Paths:* All files included in the replication packet at `raw/fuzzy_merge_training_data`.
- *Description:* These sets of manually-labeled fuzzy merge examples are used to train the probabilistic record linkage model used in the build. These data were produced using the script at `build/fuzzy/Fuzzy_Merge_Build_Training_Set.py` (which is not run as part of the build since it requires human input, but is included for illustrative purposes).

## 16. Bloomberg Identifier Data

- *Paths:*
  1. `raw/externalid/bbg_fig1.csv`
  2. `raw/externalid/known_unmatched_identifiers.csv`
- *Description:*
  - The file number (1) should be manually constructed using a computer with the Bloomberg terminal Excel add-in. This file starts from the list of FIGIs that are obtained by searching for externalids in OpenFIGI (see the file `build/externalid_matching/fig1_api.R`). This list is populated to `raw/externalid/externalid_keyfile.csv`: this list should

be copied to the leftmost column of the Excel spreadsheet, and the rest of the columns should be populated using the BDP function in Excel. The resulting file should be then placed at `raw/externalid/bbg_figi.csv`.

- The file number (2) is an optional list of externalid identifiers (one per row) that are known to not match to a true security in the OpenFIGI universe. An example file with one row is included in the replication packet; this can be optionally populated in order to speed up the build.
- Sample single-row files with mock data are included in the replication packet in order to demonstrate the structure of the data.

## 17. API Key for OpenFIGI Data

- *Paths:* None.
- *Description:* The build 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 `build/externalid_matching/figi_api.R`.

## 18. Compustat Data

- *Paths:*
  1. `raw/wrds/Compustat/cs_global_feb19.dta`
  2. `raw/wrds/Compustat/cs_northam_feb19.dta`
  3. `raw/wrds/Compustat/funda.dta`
  4. `raw/wrds/Compustat/bank_funda.dta`
  5. `raw/wrds/Compustat/g_funda.dta`
- *Description:* These files constitute a bulk download of Compustat data ([compustat.com](http://compustat.com)) obtained from WRDS. The files correspond respectively to: (1) the full Global Compustat yearly database; (2) the full North American Compustat yearly database; (3) non-financial company fundamentals (North America); (4) bank fundamentals; and (5) non-financial company fundamentals. Sample single-row versions of these files with mock data are included in the replication packet in order to illustrate the structure of the data.

## 19. Worldscope Data

- *Paths:*
  1. `raw/wrds/Worldscope/worldscope_feb19.dta`
  2. `raw/segment/wrds_ws_segment.dta`
- *Description:* These files constitute a bulk download of Worldscope data ([refinitiv.com](http://refinitiv.com)) obtained from WRDS. The files correspond respectively to: (1) the full Worldscope fundamentals database; (2) geographic segment sales data. Sample single-row versions of these files with mock data are included in the replication packet in order to illustrate the structure of the data.



## 20. IMF Data [\*]

- *Paths:* All data included in the replication packet at `raw/macro/Concordances` and `raw/macro/IFS`.
- *Description:* This data was sourced from the IMF's website and provides selected indicators from the IMF's International Financial Statistics (IFS) data, as well as basic country- and currency-level symbology. The file `raw/macro/IFS/IFS_Flatfile.csv` contains a single row because the actual file is very large — this sample illustrates the structure of the data, and users are encouraged to download the full IFS data directly from the IMF ([data.imf.org](http://data.imf.org)).