

# README file for “Internationalizing Like China”

Christopher Clayton\* Amanda Dos Santos† Matteo Maggiori‡ Jesse Schreger§

November 2024

## 1 Overview

This README file describes the replication package associated with the paper “Internationalizing Like China” by Christopher Clayton, Amanda Dos Santos, Matteo Maggiori, and Jesse Schreger.

The replication package consists of code and, whenever possible, input datasets. These replicate all tables and figures contained in the paper and Online Appendix. The uppermost directory of the replication folder contains the following objects:

1. `README.pdf`: this file; details the replication package
2. `dataset_list.xls`: file listing all datasets
3. `input`: folder containing the files used as inputs for our analysis
4. `output`: folder containing the files, figures and tables that are produced by the codes
5. `rmb_replication`: folder containing the codes to produce files, figures and tables in the paper and paper appendix

## 2 Data Availability and Provenance Statements

This paper involves the analysis of external data coming from various sources. As the authors, we are secondary data users (i.e., we did not generate the data), and therefore the provenance and data availability statements coincide.

### Statement about Rights

We, the authors, certify that we have legitimate access to and permission to use all data in this manuscript. For data included in the replication package, we confirm the data is publicly available and may be freely accessed as described in the Section 2.1 below.

---

\*Yale School of Management, NBER; christopher.clayton@yale.edu.

†Columbia Business School; amanda.dossantos@gsb.columbia.edu.

‡Stanford University Graduate School of Business, NBER, and CEPR; maggiori@stanford.edu.

§Columbia Business School, NBER, and CEPR; jesse.schreger@columbia.edu.

## Summary of Data Availability

Some data **cannot be made** publicly available. Files listed in Section 2.1 are publicly available and, whenever possible, included in the accompanying replication package. Files in Section 2.2 are commercially available and must be purchased from the respective data providers; in these cases the list below describe the conditions under which we obtained the data, and any future users might get access to the data. More information on the datasets and variables is found in the spreadsheet that comes in the replication package, named `dataset_list.xls`.

## Details on each Data Source

The following is a list of the raw input files required to replicate the analysis, together with a description of the data, and the conditions under which we obtained the data and others can obtain it. All paths listed are relative to the root replication folder referred to as `cdms1_replication` in the code. The path structure outlined below is the one that is referenced in the code and described on Section 5.1.

### 2.1 Publicly Available Datasets

#### 2.1.1 China Central and Depository Clearing (CCDC) Statistical Reports

The paper uses data obtained from CCDC Statistical Reports ([CCDC, 2014-2021](#)).

**Data access description:** We downloaded Monthly Statistical Reports for the months at the end of each quarter. These are publicly available on CCDC’s website ([https://www.chinabond.com.cn/zsjs/zsjs\\_tjsj/tjsj\\_tjyb/](https://www.chinabond.com.cn/zsjs/zsjs_tjsj/tjsj_tjyb/)).

#### Datafiles:

- (i) `input/ccdc/*`

**Description:** These files contain CCDC bond depository balance (stock) by investor type since January 1997. The category of interest (“External Institutions”) only appears in the files after June 2014. Date accessed: 10/06/2021.

- (ii) `input/ccdc/composition/202112_T06.xlsx`

**Description:** This file contains CCDC bond holding structure for selected bond types in December 2021. Date accessed: 01/27/2022.

### 2.1.2 Shanghai Clearing House (SHCH) Fixed Income Products Statistical Reports

The paper uses data obtained from CCDC Statistical Reports ([SHCH, 2014-2021](#)).

**Data access description:** We downloaded Monthly Statistical Reports (in Chinese) for the months at the end of each quarter. These are publicly available on SHCS’s website (<https://www.shclearing.com.cn/sjtj/tjyb/>).

#### Datafiles:

- (i) `input/shch/raw_cn/*`

**Description:** These files contain SHCH bond depository balance (stock) by investor type and are only available in Mandarin. Date accessed: 10/06/2021.

- (ii) `input/shch/raw_en/*`

**Description:** These English version files are a translation of the original in Mandarin. An automatic translation software (DeepL API) was used. Please see the Section 5.3 for further details on implementation. These translations were then reviewed by a native Mandarin speaker.

### 2.1.3 Bond Connect

**Data access description:** We collected data from Bond Connect’s website (more details below).

#### Datafiles:

- (i) Foreign Holdings: `input/bond_connect/bc_holdings.xlsx`

**Description:** Total foreign holdings published by Bond Connect ([Bond Connect, 2017-2023b](#)). Data includes but is not limited to holdings through Bond Connect. Primary sources are CCDC and SHCH monthly statistics described above. Last Date accessed: 10/27/2023.

**Source:** “Foreign Holding Information” data is available online in the Market Data page of the website: [www.chinabondconnect.com/en/Resource/Market-Data.html](http://www.chinabondconnect.com/en/Resource/Market-Data.html).

- (ii) List of Approved Investors: `input/bond_connect/raw/*`

**Description:** List of overseas institutional investors which enter the China Interbank Bond Market through Bond Connect ([Bond Connect, 2017-2023a](#)). Last Date accessed: 10/27/2023.

**Source:** <https://www.chinabondconnect.com/en/Northbound/Onboarding/Approved-Investors.html>

#### 2.1.4 China Interbank Bond Market (CIBM) List of Overseas Participants

**Data access description:** We downloaded monthly lists of CIBM participants on the People's Bank of China's website (People's Bank of China, 2017-2021): <http://www.pbc.gov.cn/english/3251616/3251629/3251632/3254047/index.html>.

**Datafiles:**

- (i) input/cibm/raw/\*

**Description:** Monthly list of overseas institutional investors in China's Interbank Bond Market separated by (i) foreign central banks and similar institutions and (ii) commercial institutions. Last Date accessed: 9/15/2023.

#### 2.1.5 Qualified Foreign Institutional Investor (QFII) List

**Data access description:** We obtained the list of QFII participants with Investment Quotas Granted by the State Administration of Foreign Exchange (SAFE) on its website: <https://www.safe.gov.cn/en/2019/0719/1532.html>.

**Datafiles:**

- (i) input/qfii/QFII\_participants.xlsx

**Description:** List of Qualified Foreign Institutional Investors(QFIIs) with Investment Quotas Granted by the SAFE with approval date (SAFE, 2002-2021a). Last Date accessed: 10/16/2023.

#### 2.1.6 Renminbi Qualified Foreign Institutional Investor (RQFII) List

**Data access description:** We downloaded the list of RMB QFII participants with Investment Quotas Granted by the State Administration of Foreign Exchange (SAFE) on its website: <https://www.safe.gov.cn/en/2019/0719/1533.html>.

**Datafiles:**

- (i) input/rqfii/RQFII\_participants.xlsx

**Description:** List of RMB Qualified Foreign Institutional Investors(RQFIIs) with Investment Quotas Granted by the SAFE with approval date (SAFE, 2011-2021b). Last date accessed: 10/16/2023.

### 2.1.7 Currency Composition of Official Foreign Exchange Reserves (COFER)

**Data access description:** COFER is managed by the Statistics Department of the International Monetary Fund (IMF). Its website disseminates end-of-period quarterly data in aggregate format: [data.imf.org](https://data.imf.org).

#### Datfiles:

- (i) Aggregate Holdings by Currency: `input/reserves/cofer.xlsx`

**Description:** Aggregate holdings in RMB is available since the fourth quarter of 2016. Prior to that date, holdings of RMB were aggregated into “Other Currencies” (IMF, 1999-2022c). Last date accessed: 10/24/2023.

- (ii) List of Reporting Countries: `input/reserves/cofer_reporters_list.dta`

**Description:** At the time when we conducted the analysis, there were 149 countries disclosing the currency composition of their reserve holdings to the IMF and agreed to have their names released (IMF, 2022a). Last date accessed: 01/14/2022.

- (iii) Survey on the holdings of currencies in official foreign currency assets:

`input/miscellaneous/survey_rmb_total.dta`

**Description:** Aggregate holdings in RMB by Monetary Authorities from an ad-hoc survey conducted by the IMF in April-May 2015 Provides amount held in 2013 and 2014 (IMF, 2015). Available at <https://www.imf.org/en/~media/AF651C1D464A4E21AA94A971A6214818.ashx>. Last date accessed: 10/16/2023.

### 2.1.8 Special Data Dissemination Standard (SDDS) Plus

**Data access description:** For the adherent countries of SDDS Plus, we obtained their FX reserves position denominated in RMB. SDDS Plus is the highest tier of the IMF Data Standards Initiatives, and it is optional to SDDS subscribers. We obtained data from IMF’s data portal: [data.imf.org](https://data.imf.org).

**Datafile:** `input/reserves/SDDS_full.dta`

**Description:** Foreign reserve holding in RMB for SDDS Plus adherents (IMF, 1996-2022e). Last date accessed: 06/15/2022.

### 2.1.9 Central Bank Reports

**Data access description:** We separately hand-collected the currency breakdown directly from central banks' documents for a list of non-SDDS Plus reporting countries for which we were able to find extra documentation. The non-SDDS countries for which documentation was manually collected are: Czech Republic, Italy, Kazakhstan, Romania, South Africa, Spain, Tanzania and United Kingdom. Every document we used data from in the available spreadsheet below is saved at `input/reserves/official_documents`.

**Datafile:** `input/reserves/manual_reserves.xlsx`

**Description:** This file contains the amount of official reserves in RMB by country and by year, as well as precise information on where to find these numbers in the accompanying documents. Last date accessed: 05/10/2023.

### 2.1.10 IMF International Financial Statistics (IFS)

**Data access description:** We bulk download IFS from the IMF's data portal (<https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42&sid=1479329334655>). A login is necessary to perform the download, but researchers can sign-up for free and minimal requirements (name, email and user-created password).

**Datafile:**

(i) `input/reserves/IFS_full.csv`

**Description:** The International Financial Statistics is based on various IMF data collections and topics covered vary. IFS covers 194 countries and areas. Data available starts in 1948 for many IMF member countries (although availability of series and years vary by country) ([IMF, 1948-2022d](#)). Last date accessed: 11/15/2022.

(ii) `input/imf_ifs/ISO_currency.xls`

**Description:** This is a mapping file for ISO currency codes.

(iii) `input/imf_ifs/IMF_code1.xlsx`

**Description:** This file is a mapping from IMF codes to ISO country codes.

(iv) `input/imf_ifs/IFS_ERdata.csv`

**Description:** This file contains a panel of exchange rates.

### 2.1.11 IMF Coordinated Portfolio Investment Survey (CPIS)

**Data access description:** Similarly to IFS, we bulk download CPIS (IMF, 2001-2022b) from IMF's data portal (<https://data.imf.org/?sk=b981b4e3-4e58-467e-9b90-9de0c3367363>). A login is necessary to perform the download, but researchers can sign-up for free and minimal requirements (name, email and user-created password).

#### Datafile:

- (i) `input/imf_cpis/imf_cpis_assets.dta`

**Description:** This file contain CPIS reporters' investment in debt securities in China. Last date accessed: 06/08/2022.

- (ii) Table 2: Currency Breakdown of Portfolio Investment Assets:

`input/imf_cpis/imf_cpis_chn_debt.xlsx`

**Description:** This contains the currency breakdown of CPIS portfolio assets. Last date accessed: 06/08/2022.

### 2.1.12 BIS Debt Securities Statistics

**Data access description:** We downloaded the International debt securities data from BIS (BIS, 2010-2022) in a single file ([https://www.bis.org/statistics/full\\_debt\\_sec2\\_pub\\_csv.zip](https://www.bis.org/statistics/full_debt_sec2_pub_csv.zip)).

#### Datafile:

- (i) `input/bis/bis_debt_securities_flat.dta`

**Description:** IDS tracks outstanding amounts of international bonds, i.e. bonds issued outside the local market of the country where the borrower resides. Last date accessed: 01/31/2022.

### 2.1.13 Dataset for Covered Interest Rate Parity Deviations

**Data access description:** Obtained online from <https://sites.google.com/view/jschreger/CIP>. For more details, see Du et al. (2018).

#### Datafile:

- (i) `input/miscellaneous/dis_cip_all_adj_june2021.dta`

**Description:** Measure of the deviation from covered interest rate parity (CIP) between government bond yields in the United States and other countries. Last date accessed: 10/29/2021.

#### 2.1.14 Cboe Volatility Index (VIX Index)

The paper uses historical data for Cboe VIX Index (CBOE, 1990-2022).

**Data access description:** The full dataset and documentation can be downloaded from [https://www.cboe.com/tradable\\_products/vix/vix\\_historical\\_data/](https://www.cboe.com/tradable_products/vix/vix_historical_data/).

**Datafile:** input/miscellaneous/vix.dta

**Description:** Historical Price Data for VIX Index (daily, since 1990). Last date downloaded: 10/29/2021.

#### 2.1.15 Chinese Yuan Renminbi to U.S. Dollar Spot Exchange Rate

**Data access description:** Obtained online from <https://fred.stlouisfed.org/series/DEXCHUS>.

**Datafiles:**

(i) input/miscellaneous/fx\_rate\_eom.dta

**Description:** Monthly frequency (end-of-month). Last date accessed: 10/24/2023.

(ii) input/miscellaneous/fx\_rate\_daily.dta

**Description:** Daily frequency. Last date accessed: 10/24/2023.

#### 2.1.16 CEPII Gravity database

**Data access description:** Downloaded from [http://www.cepii.fr/CEPII/en/bdd\\_modele/bdd\\_modele\\_item.asp?id=8](http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8). More details available in Conte et al. (2022a).

**Datafiles:**

(i) input/gravity/Gravity\_V202202.dta

**Description:** For any pair of countries, from 1948 to 2020, Gravity provides all the information required to estimate gravity equations: trade flows, geographical distances, trade facilitation measures, macroeconomic indicators, among others (Conte et al., 2022b). Last date accessed: 12/27/2022.

#### 2.1.17 Treasury International Capital (TIC) Data

**Data access description:** Obtained online from <https://ticdata.treasury.gov/resource-center/data-chart-center/tic/Documents/shc2020-fullreport.pdf>.

**Datafiles:**

(i) input/miscellaneous/TIC\_table\_A6\_A7\_2020.dta



**Description:** U.S. holdings of foreign long (Table A6) and short-term (Table A7) debt securities in China in RMB ([Treasury Department, 2020](#)). Last Date accessed: 01/22/2022.

### 2.1.18 OpenFIGI API Data

**Data access description:** Raw data is obtained by querying the OpenFIGI API for each individual CUSIP and ISIN code found in the universe of the CGS data. As explained above, the API is free and open to the public, users should sign-up for an OpenFIGI account to obtain an API key which provides a higher rate-limit. More information is available at [openfigi.com](https://openfigi.com). Replication is provided in previous work by [Coppola, Maggiori, Neiman and Schreger \(2021\)](#). All details are available in the replication package at <https://globalcapitalallocation.s3.us-east-2.amazonaws.com/redrawing-the-map.zip>

**Datafiles:** `input/figi/figi_master_compact.dta`

**Description:** This file contains information from Bloomberg’s OpenFIGI API (including market sector, FIGI identifiers, and other security and issuer level information). Last Date accessed: 09/05/2024.

## 2.2 Commercially Available Datasets

### 2.2.1 CUSIP Global Services (CGS) Data

The paper uses data obtained from CUSIP Global Services ([CUSIP Global Services, 2014-2020](#)). We obtained security- and issuer-level files for global CUSIP-bearing securities. As described in detail in Section 5.1, [CUSIP Global Services \(2014-2020\)](#) is an input to the Morningstar data build.

**Data access description:** Obtained commercially from CUSIP Global Services (researchers can subscribe to the data). Some universities and research institutions have such a subscription. More info on data access can be found at <https://www.cusip.com/>.

**Datafiles:** The required files are the security- and issuer-level master files for global CUSIP-bearing securities below (named by the provider):

- (i) `input/cgs/[Vintage]/AIMASTER.PIP.zip`
- (ii) `input/cgs/[Vintage]/ALLCNPMMASTER_ISIN.PIP.zip`
- (iii) `input/cgs/[Vintage]/ALLCNPMMASTER_ISSUER.PIP.zip`

- (iv) `input/cgs/[Vintage]/CBRLEIMSTR.PIP`
- (v) `input/cgs/[Vintage]/FFAPlusMASTER.PIP.zip`
- (vi) `input/cgs/[Vintage]/INCMSTR.PIP.zip`

**Description:** In these paths `[Vintage]` refers to when we obtained the data. Our build code consolidates data obtained at three different points in time (Dec/2016, Nov/2018 and Nov/2022) to consider any potential data that may have been excluded from more recent issues.

### 2.2.2 Morningstar Data

The paper uses data obtained from Morningstar ([Morningstar, 2014-2020](#)). We obtained micro-data on portfolio investment from foreign investors via mutual funds and ETFs around the world. Our data include global mutual fund and exchange traded fund (ETF) holdings provided by Morningstar for each fund at the security level. We supplement it with information on the asset class, currency, market of issuance, nationality and residency of the issuer and its ultimate parent company, and other security characteristics. To do so we obtain (i) Morningstar holdings data, (ii) additional reference files from Morningstar, and (iii) auxiliary files from other sources.

#### Morningstar Holdings Data

**Data access description:** Obtained commercially from Morningstar (researchers can subscribe to the data). More info on data access can be found at <https://www.morningstar.com/>.

#### Datafiles:

- (i) `input/morningstar/historical`
- (ii) `input/morningstar/monthly_new`

**Description:** Morningstar data and code utilized in this paper align closely with the data and code utilized in earlier studies by [Maggiori et al. \(2020\)](#) and [Coppola et al. \(2021\)](#), with divergences in part due to adjustments in file names made by the data provider.

Raw data comes as a series of `.7z` and `.zip` archives containing 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 in the replication code.

The archives provided by Morningstar from 2010 to 2018 have holdings data for a given month, a particular universe of funds (FE: open-end mutual funds, FM: open-end money market funds, and FO: exchange-traded funds), a geographic region (US or non-US), and a given fund activity status

(active or inactive) and have names of the form:

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

From 2018 to 2020, the archives obtained were aggregated by the universe of funds and year of reporting, therefore named:

[FO/FM/FE]\_[Active/Inactive]\_[YYYY].[zip]

All these archives are collectively stored at `input/morningstar/historical`.

Starting in April 2020, we began to obtain monthly updates from Morningstar. Data from April 2020 to December 2020 collectively stored and `input/morningstar/monthly_new`.

### Morningstar Reference Data

**Data access description:** delivered directly by Morningstar (<https://www.morningstar.com/>), and can be purchased from the company.

#### **Datafiles:**

- (i) `input/morningstar/mapping/[year]/*`

**Description:** Static fund-level information obtained at [year].

### Auxiliary Files

Some files used in code to process and clean Morningstar come from other sources. We list them here to emphasize that these are required in constructing the Morningstar data we used in our analysis, despite those not being obtained through Morningstar.

1. **FactSet ETF Data:** `input/factset/etf_aum/latest/factset_etf_aum_updated.xlsx`

**Data access description:** Commercially available from Factset (researchers can subscribe to the data). Some universities and research institutions have such a subscription. More info on data access can be found at [www.factset.com](http://www.factset.com).

**Description:** This file provides the total AUM per year for a list of ETFs.

2. **Factset ISIN to Entity ID:** `input/cmns/gcap_isin_to_factset_entity_id.dta`

**Data access description:** Commercially available from Factset (researchers can subscribe to the data). Some universities and research institutions have such a subscription. More info on data access can be found at [www.factset.com](http://www.factset.com).

**Description:** This file is a list of security ISINs to Factset Entity IDs.

3. **Asset Class Mapping:**

`input/morningstar/morningstar_api_data/Categories_Asset_Class.dta`

**Data access description:** Internally created and code is included in this replication package

**Description:** This file is a mapping from the internal Morningstar security typecodes to our security classification. We provide this file in this replication package.

4. **Open FIGI Data:** Automated download

**Data access description:** Code downloads certain data from Bloomberg’s OpenFIGI API. API is free and open to the public, though unauthenticated traffic will be subject to a lower rate-limit. In order for the download to work, users should sign-up for an OpenFIGI account to obtain an API key which provides a higher rate-limit.

5. **Bloomberg ID auxiliary file:** `input/morningstar/externalid/donotsearch.csv`

**Description:** This is a list of external IDs known to not match a true security in the OpenFIGI universe and, therefore, a list of IDs we do not need to search for in FIGI (hence the file name “donotsearch”). It is optionally populated in order to speed up the build.

6. **Bloomberg ID auxiliary file:** `input/morningstar/externalid/externalid_keyfile.dta`

**Data access description:** This file is created within the build (saved in line 30 of the code file `rmb_replication/0_data_builds/morningstar/externalid/externalid_postbloomberg.do`). Since the data downloaded from Bloomberg and OpenFIGI is static, each security only needs to be downloaded once and therefore we use the version saved in the previous time we ran the build to speed up the process. Anyone who is running the process for the first time will not have this file. As long as donotsearch is not populated, they will instead obtain all the external IDs through the FigiAPI.

**Description:** This file contains identifiers to match to the Bloomberg download and provide information about the securities.

7. **Bloomberg Identifier Data:** `input/morningstar/externalid/bbg_figi.csv`

**Data access description:** This file should be manually constructed using a computer with the Bloomberg terminal Excel add-in. Bloomberg is commercially available and some universities and research institutions have a terminal and access through the Excel add-in. We list here the detailed steps to do so:

- (a) Open the file `output/morningstar/temp/externalid/externalid_keyfile.csv` con-

taining the FIGI codes

- (b) Construct the Bloomberg formulas with the following syntax: =BDP (“Security”, “Field”), where “Security” corresponds to the FIGI IDs in the file followed by the word “FIGI” and “Field” are: “id\_isin”, “id\_cusip”, “name”, “security\_name”, “entry\_issue\_iso”, “crncy”, “maturity”, “cpn”, “market\_sector\_des”, “gics\_sector\_name”, and “gics\_industry\_name”. Notice that data extracted through the Excel API is subject to monthly download limits.
- (c) Copy and paste as values the information obtained before closing the file.
- (d) Save this file as `input/morningstar/externalid/bbg_figi.csv`

**Description:** This file results from using the Bloomberg BDP function in Excel to download information from external IDs.

#### 8. GCAP Files: `input/cmns/cmns_aggregation.dta`

`input/cmns/gcap_factset_id_res_nat.dta`

`input/gcap/gcap_security_master*`

**Data access description:** These files are obtained from previous published work by [Coppola, Maggiori, Neiman and Schreger \(2021\)](#). Details are available in the replication package at <https://globalcapitalallocation.s3.us-east-2.amazonaws.com/redrawing-the-map.zip>.

**Description:** These are mappings used in the last step of the Morningstar build to ensure consistency between files.

#### 9. Morningstar List of Index Funds:

`input/morningstar/morningstar_api_data/ms_index_funds.dta`

**Data access description:** Obtained using Morningstar Direct, a platform available for subscribers. More information available at <https://www.morningstar.com/business/brands/data-analytics/products/direct>.

**Description:** A list of all funds in our analysis (by `MasterPortfolioId`, Morningstar’s unique fund identifier) and the variable `index_fund` that equals to “Yes” indicating a “pure index fund”.

### 2.2.3 FactSet Data Services

We use two products from FactSet, a commercial provider of financial data and analysis:

#### Factset API Search

**Data access description:** Obtained commercially from FactSet. Researchers can reach out to the

company and subscribe to this service. More information is available at <https://www.factset.com/marketplace/catalog/product/factset-id-lookup-api>.

**Datafile:** /input/auxiliar/factset\_api\_search.dta

**Description:** FactSet’s Entity API provides access to FactSet’s complete security and entity level symbology and comprehensive entity reference data to link an entity to a master entity identifier. In the particular case of this paper, we used the API in order to name match a list of investors entering the Chinese bond market (described above) to their country, sector and industry.

### FactSet Data Management Solutions

**Data access description:** Obtained commercially from FactSet (FactSet, 2003-2021). Researchers can reach out to the company and subscribe to this service. More information is available at <https://www.factset.com/marketplace/catalog/product/factset-data-management-solutions>.

#### **Datafiles:**

1. input/factset/ent\_entity\_coverage
2. input/factset/ent\_entity\_naics\_rank
3. input/factset/ent\_entity\_structure
4. input/factset/sym\_entity
5. input/factset/naics6\_map
6. input/factset/sic\_map
7. input/factset/factset\_industry\_map

**Description:** These files contain entity-level symbology and entity reference data that provides the connections needed to link disparate sources of information to a master identifier.

## 3 Dataset list

Please refer to the spreadsheet located in the replication package, named `dataset_list.xlsx`.

## 4 Computational Requirements

As we describe in Section 5.1, we split the analysis into empirical and theoretical parts. The computational requirements for each sections are described below.

## 4.1 Empirical Sections

We performed the empirical analysis using Stanford’s High Performance Computing (HPC) cluster (named Sherlock, more information is available at <https://www.sherlock.stanford.edu/>). Sherlock uses Slurm, an open-source resource manager and job scheduler used by many of the world’s supercomputers and computer clusters, which allows requesting resources and submitting jobs in a variety of ways. Our code is written with Slurm commands to schedule and submit jobs with the appropriate order and dependencies.

We parallelize particularly large jobs whenever possible. For each job, we define the (1) specific partition to run the job (usually contingent on size), (2) time allocated to run this job, (3) number of nodes we request, (4) amount of memory, (5) number of job arrays. Details on these parameters are available in the `Master_Build.sh` for most of the tasks and `Morningstar_Build_PreBloomberg.sh` and `Morningstar_Build_PostBloomberg.sh` for the tasks involving building the Morningstar data. The largest number of hours we allow a job to run is 168 hours and the highest amount of memory requested is 500G. The execution time of our code depends both on the available memory and number of processors. The computationally intensive parts of the code are what we called the “Data Builds” in Section 5.1, where we detail the code structure. These computationally intensive jobs are part of the Morningstar data build, the part of the code in which we process and prepare fund holdings data for analysis. The approximate time needed to reproduce the analyses on a standard 2024 desktop machine is more than 14 days. Most of the other jobs, in particular the ones directly related to the production of figures and tables in the paper and paper appendix are computationally smaller and could potentially be adapted to run in a standard PC. These other empirical blocks of the code (Foreign Holdings, Investors’ Entry, and Holdings Similarity – as per the block description in Section 5.1) take approximately 10-60 minutes to run.

### 4.1.1 Software Requirements

We used different software for different tasks through the code, which are all called in as modules on the Sherlock cluster. We list those and the required packages/dependencies (installed automatically in the code) below.

- **Stata/MP 17.0 for Unix** (primary software for empirical analysis). Packages used (and automatically installed in the code) are:
  - `reghdfe`
  - `unique`
  - `outreg2`

- ftools
- egenmore
- gtools
- labutil
- carryforward
- filelist
- matsave
- nmissing
- mmerge
- groups
- sxpose
- texdoc
- kountry

- **Python** (version 3.7.12). Packages used (and automatically installed in the code) are:

- regex
- openpyxl
- deepl
- xlsxwriter
- ipython
- numpy
- pandas

- **R** (version 4.2.0). Packages used (and automatically installed in the code) are:

- readstata13
- data.table
- httr
- jsonlite
- curl

- **p7zip** (command line version of 7-Zip for Unix)

- **SAS** (version 9.4)



## 4.2 Theory Section

This part performs the numerical solutions and graphical illustrations of the theory included in the paper. For this section, the required software is Matlab (we used version 9.13).

This section can be executed locally and independently from the rest of the code. Although the last job in `Master_Build.sh`, it does not require any dependencies, and therefore runs independently and much faster than the other blocks. This process does not require a HPC, as the empirical sections, and it is possible to execute this block in a standard PC equipped with Matlab. Approximate time needed to reproduce the analyses on a standard (CURRENT YEAR) desktop machine is less than 10 minutes.

## 5 Description of programs/code

### 5.1 Code Structure

All the codes needed for replicating the full paper (empirical and theoretical sections) are in the folder `rmb_replication`. We organize the execution of all files with a `Master_Build.sh` script, found in the uppermost code directory (`rmb_replication`), that calls controller files so that the full replication is automated and different parameters can be passed to the computing cluster scheduler for each part of the code. We separated the code in five large blocks:

1. **Data Builds:** details on cleaning and processing two large datasets required for the empirical analysis conducted in the paper
2. **Foreign Holdings:** figures and analysis directly related to the characterization of foreign investment in China's domestic bond market (Paper Section 2)
3. **Investors' Entry:** figures and analysis directly related to the selection of the foreign investor base (Paper Section 2.1)
4. **Holdings Similarity:** figures and analysis related to EM, DM, and Renminbi Bonds in Private Portfolios (Paper Section 2.2)
5. **Theory:** figures related to the model (Figure 6, Figure A.XIV, and Figure A.XV).

### 5.2 Data Builds

#### 5.2.1 CUSIP Global Services (CGS) Data Build

CGS Build takes a number of commercially purchased datasets described in detail above, combines, cleans, and manipulates them in order to prepare the data for analysis. The file structure for the CGS Build is the following:

```

rmb_replication/
├── 0_data_builds
│   └── cgs
│       ├── CGS_Build.do
│       ├── CGS_Build_Controller.sh
│       ├── CGS_Unzip.sh
│       └── cgs.do

```

**Code Execution Details:** The `Master_Build.sh` calls two scripts: `CGS_Unzip.sh` and `CGS_Build_Controller.sh`. `CGS_Unzip.sh` unzips the raw files that were downloaded for years 2016/2018/2022. After unzipping, `CGS_Build_Controller.sh` runs the `.do` file `CGS_Build.do`, that calls `cgs.do`, the script containing the processing and cleaning steps of the raw files.

The build outputs a series of files containing essential information about global CUSIP-bearing securities and their issuers, such as asset classification and residency information. Other than regular pre-processing of the raw files, the build creates compact versions of the data across different download years for us to be able to prioritize the recent versions over the older whenever possible. These intermediate files are later merged into the Morningstar build outputs.

### 5.2.2 Morningstar Build

The Morningstar build constructs the data of Morningstar portfolio holdings. The file structure is the following:

```

rmb_replication/
├── 0_data_builds
│   └── morningstar
│       ├── Create_Final_Files.do
│       ├── ER_Data_Build.do
│       ├── Final_Clean_for_Analysis.do
│       ├── HoldingDetail_Build_1.do
│       ├── HoldingDetail_Build_2.do
│       ├── HoldingDetail_Build_3.do
│       ├── Internal_Class.do
│       ├── Internal_Currency.do
│       ├── Manual_Corrections.do
│       ├── Morningstar_Build.do
│       ├── Morningstar_Build_Controller.sh
│       ├── Morningstar_Build_PostBloomberg.sh
│       └── Morningstar_Build_PreBloomberg.sh

```

```

├─ Morningstar_Mapping_Build.do
├─ PortfolioSummary_Build.do
├─ Refine_Cusip_Fill_Isin.do
├─ Refine_Cusip_Merge.do
├─ Refine_Extid_Merge.do
├─ Refine_Parse_Externalid.do
├─ externalid
│   ├── Figi_API.R
│   ├── R_Controller.sh
│   ├── externalid_collect_extid_master.do
│   ├── externalid_make_extid_csvtodta.do
│   ├── externalid_postbloomberg.do
│   ├── externalid_prebloomberg.do
│   └─ make_externalid_master.R
├─ read_xml
│   ├── convert_xml_to_dta.sas
│   ├── save_dta.sh
│   └─ unzip.sh

```

We split this process into Pre and Post Bloomberg stages with heavier data load and processing happening in the former. The code is split into these sections because there is a step involving access to a Bloomberg terminal in between them. We call these two build processes from the `Master_Build.sh` file, but these processes themselves can be thought of as independent build processes as the parameters provided are defined under the respective files and not under `Master_Build.sh`. As such, we'll describe the two processes in detail here.

**Pre-Bloomberg Build.** This part includes various build steps that are controlled by the `Morningstar_Build_PreBloomberg.sh` master file. This file includes all the build parameters described before (such as build time, number of nodes, or the partition) which then call on specific parts of the build to create the final holding detail files that are input for the figures we produce. Specifically, the Pre-Bloomberg build includes, in order of running, the following steps:

#### Part I: Unzip, save and prep jobs

1. `morningstar/read_xml/unzip.sh`

**Description:** Submits jobs to unzip the raw files.

2. `morningstar/read_xml/save_dta.sh`

**Description:** Submits jobs to convert the .xml files to .dta files.

3. `morningstar/Morningstar_Mapping_Build.do`

**Description:** Builds the accompanying metadata (fund-level information).

4. morningstar/ER\_Data\_Build.do

**Description:** Processes exchange rate data from the International Monetary Fund's International Financial Statistics (IFS).

5. morningstar/PortfolioSummary\_Build.do

**Description:** Generates a dataset with portfolio summary data.

## Part II: Building the holding detail files

1. morningstar/HoldingDetail\_Build\_1.do

**Description:** Generates raw monthly files for years monthly data is available.

2. morningstar/HoldingDetail\_Build\_2.do

**Description:** Combines monthly and historical (annual) files, cleans and appends them, and then merges with external ID and FX data.

3. morningstar/HoldingDetail\_Build\_3.do

**Description:** Appends monthly files to create annual files.

## Part III: Create and obtain external identifiers for query

This section creates a list of relevant identifiers then queries the OpenFIGI API for them. In this step we obtain information about the assets in the holdings data (type of the asset, coupon or maturity, or currency information, information on the issuer, etc).

1. morningstar/Refine\_Parse\_Externalid.do

**Description:** Cleans and parses the externalid field in the Morningstar holdings data which will be used in conjunction with the OpenFIGI API in order to identify securities for which we are otherwise lacking identifiers.

2. morningstar/externalid/externalid\_prebloomberg.do

**Description:** Consolidates the list of external IDs to be sent to OpenFIGI via API.

3. morningstar/externalid/FIGI\_API.R

**Description:** Runs OpenFIGI API using API key. We obtained the key through the free

account. Here are the steps to perform:

- (a) Navigate to OpenFIGI's website <https://www.openfigi.com/>
- (b) Sign up for a free account by clicking on the 'Sign Up' button located in the upper right corner.
- (c) Fill out the relevant personal information and once you have completed the form, your account should be created immediately (we've used a .edu email address).
- (d) Every time you log in, you will be prompted to enter a verification code sent to your email address.
- (e) Navigate to the 'API' page and scroll down to the 'API Key' section, locate or create (if it's your first time) the API key by clicking on the "Create API key" button.
- (f) Copy the API key (a series of characters)
- (g) Update the R Script:
  - i. Open the R script `/morningstar/externalid/Figi_API.R`
  - ii. Navigate to section SETUP - line 86: `apikey = '< INSERT KEY HERE >'`
  - iii. Replace the existing API key with the newly copied key

## Part IV: Manual Bloomberg data download

The build then stops here so that we use the relevant outputs of the previous stage along with the manual Bloomberg download to continue with the post-Bloomberg stage. The Bloomberg terminal data download is described in detail above. In sum, the process involves taking the file `output/morningstar/temp/externalid/externalid_keyfile.csv` containing the security IDs we need information from to a computer equipped with access to Bloomberg through the Excel add-in, a Bloomberg terminal, to obtain the asset characteristics.

**Post-Bloomberg Build.** This part includes various build steps that are controlled by the `Morningstar_Build_PostBloomberg.sh` master file. Specifically the following steps are executed sequentially:

## Part I: Securities characteristics cleaning

This step processes the external id file after the addition of manual Bloomberg downloads and improves the asset/security metadata in the Holding Detail files by using the outputs from CGS build and FIGI/Bloomberg downloads.

1. `morningstar/externalid/externalid_postbloomberg.do`

**Description:** Merges and produces a consolidated .dta file with information from OpenFIGI;

also creates a linking file to match external ID and ISIN/CUSIP.

2. `morningstar/Refine_Cusip_Fill_Isin.do`

**Description:** Performs a series of data cleaning steps that improve the quality of security metadata in the HoldingDetail files produced. First, it uses ISIN identifiers to merge in information from the CGS security master file. We use these to refine the security-level information such as asset class categorization. Second, we merge in data using the identifiers recovered via the OpenFIGI/Bloomberg data pull.

3. `morningstar/externalid/externalid_collect_extid_master.do`

`morningstar/externalid/make_externalid_master.R`

`morningstar/externalid/externalid_make_extid_csvtodta.do`

`morningstar/Refine_Extid_Merge.do`

**Description:** All these files are to create an internal flat-file which has all security-level details for each external ID in the Morningstar holdings data.

4. `morningstar/Internal_Currency.do`

**Description:** Constructs a dataset with the modal currency assignments for each security within the Morningstar holdings data. We look for modal currency assignments within funds and then across funds.

5. `morningstar/Refine_Cusip_Merge.do`

**Description:** Merges in security-level data from the CUSIP Global Services (CGS) master files into the holdings data.

6. `morningstar/Internal_Class.do`

**Description:** Finds the modal security type assigned to each fund in the Morningstar data.

7. `morningstar/Manual_Corrections.do`

**Description:** Manually corrects several incorrectly reported positions in the holding details data.

## Part II: Create final files

This steps produces the final “HD” (Holding Detail) files that are used in the analysis. They include all the corrections to security characteristics and guarantees consistency in the data.

1. `morningstar/Create_Final_Files.do`

**Description:** Produces the HD (Holding Detail) files.

2. `morningstar/Final_Clean_for_Analysis.do`

**Description:** Uses the final "HD" (Holding Detail) files to perform additional (project-specific) standardization.

### 5.3 Foreign Holdings

In this block we perform the analysis of foreign ownership of RMB bonds. This produces Paper Figures 1 and 8, Appendix Figures A.I, A.II, A.III, A.IV, and A.V. The following scripts are executed:

1. `rmb_replication/1_foreign_holdings/translate_shch.py`

**Description:** Uses the DeepL API to translate Shanghai Clearing House (SHCH) files from Mandarin Chinese to English. Before running this step, it is necessary to obtain an API key through the DeepL's website. We obtained the key through the free account. Here are the steps to perform:

- (a) Navigate to DeepL's website <https://www.deepl.com/>
- (b) Sign up for a free account
- (c) Click on your profile icon located in the upper right corner to access your personal account page
- (d) Under the 'API Keys' section, locate the API key. Ensure its status is marked as 'Active' before proceeding
- (e) Copy the API key (a series of characters)
- (f) Update the Python Script:
  - i. Open the python script `rmb_replication/1_foreign_holdings/translate_shch.py`
  - ii. Navigate to line 38: `translator = deepl.Translator(< INSERT KEY HERE >)`
  - iii. Replace the existing API key with the newly copied key

2. `rmb_replication/1_foreign_holdings/aggregate_data_series.do`

**Description:** Combines multiple data sources described in detail above (SHCH, CCDC and Bond Connect) for aggregate foreign investment in RMB and/or Chinese debt securities.

3. `rmb_replication/1_foreign_holdings/reserves_estimate.do`

**Description:** Constructs estimates of holdings of RMB bonds by central banks.

4. `rmb_replication/1_foreign_holdings/private_estimate.do`

**Description:** Constructs estimates of holdings of RMB bonds by private agents by country.

5. `rmb_replication/1_foreign_holdings/offshore_estimate.do`

**Description:** Constructs estimates of holdings of onshore and offshore RMB.

6. `rmb_replication/1_foreign_holdings/foreign_holdings_figures.do`

**Description:** Uses output files generated in (2), (3), (4) and (5) to produce Paper Figures 1 and 8, Appendix Figures A.I, A.II, A.III, A.IV, and A.V.

## 5.4 Investors' Entry

In this block we analyze the entry pattern of the investors. This produces Paper Figure 2 and both panels of Appendix Figure A.VI. The analysis is performed in the following order:

1. `rmb_replication/2_investor_entry/entry_clean_and_combine.do`

**Description:** Performs standardization to investors' names and combines information on entry from all programs (Bond Connect, CIBM, QFII and RQFII). Its main output is the file `output/investor_entry/accession_all.dta`, used in the next manual step.

2. Manual step: FactSet ID Lookup API

**Description:** Uses the FactSet ID Lookup API to obtain identifiers (FactSet Entity ID) for list of investor names. More information on how to obtain this data is available above and <https://www.factset.com/marketplace/catalog/product/factset-id-lookup-api>.

3. `rmb_replication/2_investor_entry/entry_main_analysis_part_1.do`

**Description:** Using FactSet Entity IDs, combines information from FactSet on country, sector and industry (SIC and NAICS codes) of investors.

4. Manual step: Google search

**Description:** Uses Google searches to obtain country, sector and industry for entities not matched on FactSet.

5. `rmb_replication/2_investor_entry/main_analysis_part_2.do`

**Description:** Combines additional information from FactSet, reclassifies investors into broader groups and computes the cumulative distribution function of investors' entry.



6. `rmb_replication/2_investor_entry/entry_figures.do`

**Description:** Produces Paper Figure 2 and both panels of Appendix Figure A.VI.

## 5.5 Holdings Similarity

This section produces Paper Figures 3 and 4, Appendix Figures A.VII to A.XIII and all appendix tables. We do so using the following code files:

1. `rmb_replication/3_holdings_similarity/bond_betas.do`

**Description:** Estimates bond betas via univariate country-specific regressions of quarterly bond returns.

2. `rmb_replication/3_holdings_similarity/correlations.do`

**Description:** Computes holdings correlation for baseline case from 2014 to 2020.

3. `rmb_replication/3_holdings_similarity/correlations_alt.do`

**Description:** Computes holdings correlation for the alternative cases in Appendix Figure A.VIII in 2020.

4. `rmb_replication/3_holdings_similarity/correlations_other_assets.do`

**Description:** Computes holdings correlation for US Bonds and equity by nationality of the issuer in 2020.

5. `rmb_replication/3_holdings_similarity/index_inclusion.do`

**Description:** Measures the aggregate value of RMB holdings for funds benchmarked to indices.

6. `rmb_replication/3_holdings_similarity/table_summary.do`

**Description:** Prepares data and produces Appendix Table A.I.

7. `rmb_replication/3_holdings_similarity/gravity.do`

**Description:** Estimates gravity regressions and outputs Appendix Table A.III.

8. `rmb_replication/3_holdings_similarity/figures.do`

**Description:** Produces Paper Figures 3 and 4, Appendix Figures A.VII, to A.XIII and Appendix Tables A.II and A.IV.

## 5.6 Theory

This section produces Paper Figure 6, Appendix Figures A.XIV and A.XV. We do so using the following code files:

1. `rmb_replication/4_theory/master_file.m`  
**Description:** Master File to Run all the required scripts.
2. `rmb_replication/4_theory/heterogeneous_figure.m`  
**Description:** Produces the four panels of Figure 6.
3. `rmb_replication/4_theory/homogeneous_figure.m`  
**Description:** Produces the four panels of Figure A.XIV.
4. `rmb_replication/4_theory/heterogeneous_demand_figure.m`  
**Description:** Produces the four panels of Figure A.XV.

We also require the following auxiliary files:

1. `rmb_replication/4_theory/parameters/params1.mat`  
**Description:** Define the parameters used in the numerical illustration.
2. `rmb_replication/4_theory/bisection_algorithm.m`  
**Description:** A bisection algorithm to find the zero of a decreasing function.
3. `rmb_replication/4_theory/piPathfcn.m`  
**Description:** Calculates the path of beliefs  $\pi_t$  for a path  $M_t$  and graduation date T.
4. `rmb_replication/4_theory/piTerminal.m`  
**Description:** Calculates the terminal belief  $\pi_t$  for a path  $M_t$  and graduation date T.

## 6 Step-by-Step Replication Procedure

A reader interested in replicating all the figures and tables in the paper and appendices should proceed as follows:

1. Download and save the replication folder
2. Obtain the commercially available datasets described above and store in the corresponding path
3. Obtain a free OpenFIGI as described on Section 5.2.2 and add it to the code
4. Obtain a free DeepL API as described on Section 5.3 and add it to the code
5. In `Project_globals.do`: change path to the where `cdms1_replication` was stored
6. In `master_shell_profile.sh`: add the corresponding user (username and email) and system variables (name and partitions), and adjust the main path `gcap_data` to the corresponding path `cdms1_replication` and the code path `rmb_replication` to the code directory
7. Execute `Master_Build.sh`: this will schedule and run all the required jobs to fully replicate the paper. To facilitate running only certain parts of the code, we named each section in the `Master_Build.sh`. A reader interested in running only certain parts of the code could do so by calling the name of the section as the first argument when executing `Master_Build.sh`. The name of the sections are: `cgs_build`, `morningstar_preBBG`, `morningstar_postBBG`, `foreign_holdings`, `investor_entry`, `holdings_similarity` and `theory`.

## 7 List of Figures and Tables

Figure/Table #	Program	Line Number
Figure 1	<code>foreign_holdings_figures.do</code>	52-129
Figure 2	<code>entry_figures.do</code>	41-100
Figure 3	<code>figures.do</code>	29-68
Figure 4	<code>figures.do</code>	108-140
Figure 6	<code>heterogeneous_figure.m</code>	196-272
Figure 8	<code>foreign_holdings_figures.do</code>	265-269
Figure A.I	<code>foreign_holdings_figures.do</code>	271-274
Figure A.II	<code>foreign_holdings_figures.do</code>	276-284
Figure A.III	<code>foreign_holdings_figures.do</code>	286-291
Figure A.IV	<code>foreign_holdings_figures.do</code>	293-301
Figure A.V	<code>foreign_holdings_figures.do</code>	303-324
Figure A.VI	<code>entry_figures.do</code>	102-299
Figure A.VII-A.XIII	<code>figures.do</code>	70-372
Figure A.XIV	<code>homogeneous_figure.m</code>	133-223
Table A.I	<code>table_summary.do</code>	201-228
Table A.II	<code>figures.do</code>	376-426
Table A.III	<code>gravity.do</code>	96-285

## References

- BIS**, “International Debt Securities Statistics,” 2010-2022. Downloaded: January 31, 2022.
- Bond Connect**, “Approved Investors List,” 2017-2023. Downloaded: October 27, 2023.
- Bond Connect**, “Foreign Holdings Information,” 2017-2023. Downloaded: October 27, 2023.
- CBOE**, “Historical Price Data for VIX Index,” 1990-2022. Downloaded: October 29, 2021.
- CCDC**, “Bond Market Statistical Reports: Depository Balance by Investor Type and Bond Holding Structure,” 2014-2021. Downloaded: October 06, 2021.
- Conte, Maddalena, Pierre Cotterlaz, and Thierry Mayer**, “The CEPII Gravity Database Highlights,” 2022.
- Conte, Matteo, Philip Cotterlaz, and Thierry Mayer**, “The CEPII Gravity Database,” Working Paper 2022-05, CEPII July 2022.
- 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,” *The Quarterly Journal of Economics*, 2021, 136 (3), 1499–1556.
- CUSIP Global Services**, “CUSIP Master Files,” 2014-2020.
- Du, Wenxin, Joanne Im, and Jesse Schreger**, “The us treasury premium,” *Journal of International Economics*, 2018, 112, 167–181.
- FactSet**, “FactSet Data Management Solutions,” 2003-2021.
- IMF**, “International Financial Statistics,” 1948-2022. Downloaded: November 15, 2022.
- IMF**, “Special Data Dissemination Standard Plus: Foreign Exchange Reserves,” 1996-2022. Downloaded: June 15, 2022.
- IMF**, “Currency Composition of Official Foreign Exchange Reserves,” 1999-2022. Downloaded: October 24, 2023.
- IMF**, “Coordinated Portfolio Investment Survey,” 2001-2022. Downloaded: June 8, 2022.
- IMF**, “Survey on the Holdings of Currency in Official Foreign Currency Assets,” 2015.
- IMF**, “COFER Reporting Countries List,” 2022. Downloaded: January 14, 2022.
- Maggiori, Matteo, Brent Neiman, and Jesse Schreger**, “International currencies and capital allocation,” *Journal of Political Economy*, 2020, 128 (6), 2019–2066.

**Morningstar**, “Portfolio Data: Holdings Data,” 2014-2020.

**People’s Bank of China**, “List of Overseas Institutional Investors in China’s Interbank Bond Market,” 2017-2021. Downloaded: September 15, 2023.

**SAFE**, “List of Qualified Foreign Institutional Investors (QFIIs) with Investment Quotas,” 2002-2021. Downloaded: October 16, 2023.

**SAFE**, “List of RMB Qualified Foreign Institutional Investors (RQFIIs) with Investment Quotas,” 2011-2021. Downloaded: October 16, 2023.

**SHCH**, “Fixed Income Products Statistical Reports: Depository Balance by Investor Type,” 2014-2021. Downloaded: October 06, 2021.

**Treasury Department**, “Report on U.S. Portfolio Holdings of Foreign Securities as of December 31, 2020,” 2020. Downloaded: January 22, 2022.