THE DOMESTIC POLITICAL ECONOMY OF CHINA'S FOREIGN AID*

Joris Mueller

README Documentation – July 2025

Contents

A Introduction **2**

B Computational Requirements **2**

C Description of Programs **2**
C.1 Folder Structure . . . 2
C.2 Programs . . . 2
C.3 Replication Instructions . . . 4
C.4 Packages Required . . . 4
C.5 Output List . . . 4

D Data Availability and Provenance Statement **5**
D.1 Chinese Aid Projects . . . 5
D.2 Unrest Data . . . 5
D.3 Prefecture-level Administrative Data . . . 6
D.4 Chinese Trade Data . . . 6
D.5 Firm-level Administrative Data . . . 6
D.6 List of Licensed Chinese Firms . . . 6
D.7 Firm Registry Data . . . 6
D.8 Firm Annual Reports . . . 7
D.9 Procurement Data . . . 7
D.10 Weather Data . . . 7
D.11 Aid Recipient Country Trade Data . . . 7
D.12 Other Country-level Variables . . . 7
D.13 Crosswalks of Firm Names, Chinese Prefectures, and Country Names . . . 8
D.14 Geospatial Data . . . 8
D.14.1 Chinese Prefecture Boundaries . . . 8
D.14.2 World Map . . . 8
D.15 Datasets List . . . 9

*I am extremely grateful to Chuyue Tian for exceptional research assistance and curating this replication package.

A Introduction

This README file describes the data and code required to replicate "The Domestic Political Economy of China's Foreign Aid," authored by Joris Mueller. The package includes several datasets from primary and secondary sources, alongside code files that produce the results presented in the paper. The replication package is structured into several folders that contain raw data, cleaned datasets, and programs in Stata and Python, ensuring replicability of the analysis. Please also consult the Data Appendix included in the paper's Online Appendix and contact the author at `jorismueller@nus.edu.sg` if you have questions.

The README file is divided into multiple sections:

- Section B outlines the computational requirements.
- Section C describes the structure of the replication folder and details the programs used.
- Section D provides a data availability statement, explaining the datasets included in the package and their sources.

B Computational Requirements

Software and Hardware Requirements

- **Stata:** all programs were executed using Stata version 18.0.
- **Python:** all programs were executed using Python 3.12.5.

All programs were last executed in July 2025 on a laptop equipped with a 13th Gen Intel Core processor, 16 GB of RAM, and running Windows 11.

C Description of Programs

C.1 Folder Structure

The replication package consists of the following main folders:

- `Data/Raw Data/`: this folder includes all raw Data Files described in Section D, each in a separate sub-folder clearly named after the data it includes.
- `Data/Temp Data/`: this folder stores all auxiliary outputs produced by programs in the folder `Do` needed as intermediate processing steps.
- `Data/Clean Data/`: this folder stores the cleaned datasets used for analysis.
- `Do`: this folder contains all Stata and Python codes that cleaned the data from the `Data/Raw Data/` folder and produced the results. All these programs are described below.
- `Output`: this folder includes all outputs produced by the code files in `Do` and is divided in the following sub-folders, organized by type of output: (i) output figures are saved in `figures`, (ii) output tables are saved in `tables`, and (iii) output statistics and back-of-envelope calculations are stored in `macros.tex`.

C.2 Programs

This section describes the programs that clean the data and produce all the results reported in the paper and that are stored in the folder `Do`. All the programs in Stata and Python are executed, in the correct order, by the do-file `Do/0_MASTERdo-file.do`. This do-file also executes the programs needed to read and clean the raw data files, which are contained in the folder `Data/Raw Data/`, described above. The programs in Stata and Python stored in `Do/auxiliary` are not directly executed by `Do/0_MASTERdo-file.do`. Instead, they are executed by the respective Stata programs stored in `Do/`.

The programs are:

- `Do/0_MASTERdo-file.do`: this is the master do-file, it sets up paths, graph options, and runs all the steps in sequence.
- `Do/1_MOFCOM_cleaning.do`: this file prepares the list of firms to be linked to other datasets.

- `Do/2_tianyancha_cleaning.do`: this file cleans basic firm information and firm ownership data.
- `Do/3_tyc_location_matching.do`: this file links firm locations to prefecture identifiers.
- `Do/4_taxcode_cleaning.do`: this file cleans firm-level administrative data, including employment.
- `Do/5_customsdata_cleaning.do`: this file cleans yearly Chinese prefecture export data.
- `Do/6_weather_IV.do`: this file cleans daily Chinese prefecture weather data.
- `Do/7_yearbook_cleaning.do`: this file cleans prefecture-year level control variables and outcomes.
- `Do/8_laborstrikes_cleaning.do`: this file cleans labor unrest data.
- `Do/9_aiddata_2_cleaning.do`: this file cleans AidData 2.0 aid project data.
- `Do/10_annual_reports_cleaning.do`: this file cleans data on the number of keywords mentioned in annual reports of firms listed on mainland Chinese stock exchanges.
- `Do/11_country_variables_cleaning.do`: this file cleans yearly data of aid recipient countries.
- `Do/12_merging_data.do`: this file merges the datasets cleaned by the previous steps and creates two main panel datasets, one at firm-year level and the other at aid recipient country-year level, and prepares the data for analysis.
- `Do/13_make_procurement_data.do`: this file creates a firm-quarter panel dataset for the number and value of domestic procurement contracts received by firms.
- `Do/14_analysis.do`: this file produces most statistical analysis reported in the paper, including all tables (except Table A.1, which was manually created by the author), figures (except Figure A.1, which is unrelated to any data and was manually created by the author), statistics and back-of-the-envelope calculations.
- `Do/auxiliary/cleaning_historical_shareholding.do`: this file identifies the list of historical shareholders for firms in the sample.
- `Do/auxiliary/clean_procurement_contract.py`: this file cleans the raw procurement dataset, identifies and counts the procurement contract(s) received by each firm in the sample, and uses the OpenAI API to read the unstructured text and extract contract value.
- `Do/auxiliary/firm_names_cleaning.do`: this file identifies all historical shareholders of each firm stored in the raw change of shareholder record from the firm registry data.
- `Do/auxiliary/globals for analysis part 1.do`: this file stores the names of variables used for firm and prefecture level analysis as Stata globals.
- `Do/auxiliary/globals for analysis part 2.do`: this file stores the names of variables used for country level analysis as Stata globals.
- `Do/auxiliary/harmonise city ids.do`: this file updates the prefecture IDs and Chinese names in a given dataset, so as to standardize the prefecture IDs and names in accordance with China's prefecture borders in 2003.
- `Do/auxiliary/label vars.do`: this file labels variables in the cleaned datasets used for analysis.
- `Do/auxiliary/make_maps.py`: this file produces Figures A.2, A.3, A.4, A.5, and A.6.
- `Do/auxiliary/make_numbers.do`: this file produces statistics stored in `Output/macros.tex`.
- `Do/auxiliary/setup.do`: this file sets the Stata version, random seed, and system directory to `Do/packages/`, in which the additional Stata packages required for the replication package are stored. The file also sets the scheme and font of plots generated by Stata.
- `Do/auxiliary/table formatting.do`: this file counts the number of observations and computes R-squared. It is called for regressions run in `Do/14_analysis.do`.

The program `Do/0_MASTERdo-file.do` was last run top to bottom in July 2025.

C.3 Replication Instructions

Replicators need to adjust the default paths in the do-file `Do/0_MASTERdo-file.do` and in the Python program `Do/auxiliary/make_maps.py`. Since most directories, packages and global variables are set in `Do/0_MASTERdo-file.do`, there is no guarantee that running each individual do. file separately will yield the correct results. By default, the author assumes that the replicator has no access to the proprietary data, so the global `$alldata` is set to 0. The code requiring proprietary data is then skipped in the execution. If the replicator has access to all data, set `$alldata` to 1 to run the complete code.

C.4 Packages Required

Stata: researchers must install the packages `asgen`, `binscatter`, `blindschemes`, `carryforward`, `coefplot`, `csdid`, `estout`, `ftools`, `gtools`, `pdslasso`, `reghdfe`, `rlasso`, `texsave`, and `winsor2`. To use the same versions of the packages used to produce the results of this paper, researchers should use the `ado` files saved in `Do/packages` (recommended). Otherwise, all these packages can be downloaded and installed via `ssc install`.

Python: researchers must install the following libraries: `concurrent.futures`, `cartopy`, `getpass`, `geopandas`, `logging`, `matplotlib`, `numpy`, `openai`, `os`, `pandas`, `PyShp`, `pyproj`, `re`, `time`, `sys`, `shapely` (all as of 2025-6-10).

C.5 Output List

Table/Figure	Program	Output File	Note
Table 1	`Do/14_analysis.do`	`output/tables/Table_1.tex`	
Table 2	`Do/14_analysis.do`	`output/tables/Table_2.tex`	
Table 3	`Do/14_analysis.do`	`output/tables/Table_3.tex`	Note: requires proprietary data.
Table 4	`Do/14_analysis.do`	`output/tables/Table_4.tex`	
Figure 1	`Do/14_analysis.do`	`output/figures/Figure_1.png`	
Figure 2	`Do/14_analysis.do`	`output/figures/Figure_2.png`	
Table A.1	n.a.	n.a.	Source: manually compiled by author based on *China Strikes Crowdmap* and *China Labour Bulletin* data.
Table A.2	`Do/14_analysis.do`	`output/tables/Table_A2.tex`	
Table A.3	`Do/14_analysis.do`	`output/tables/Table_A3.tex`	
Table A.4	`Do/14_analysis.do`	`output/tables/Table_A4.tex`	Note: requires proprietary data.
Table A.5	`Do/14_analysis.do`	`output/tables/Table_A5.tex`	
Table A.6	`Do/14_analysis.do`	`output/tables/Table_A6.tex`	
Table A.7	`Do/14_analysis.do`	`output/tables/Table_A7.tex`	
Table A.8	`Do/14_analysis.do`	`output/tables/Table_A8.tex`	
Table A.9	`Do/14_analysis.do`	`output/tables/Table_A9.tex`	
Table A.10	`Do/14_analysis.do`	`output/tables/Table_A10.tex`	
Table A.11	`Do/14_analysis.do`	`output/tables/Table_A11.tex`	
Table A.12	`Do/14_analysis.do`	`output/tables/Table_A12.tex`	
Table A.13	`Do/14_analysis.do`	`output/tables/Table_A13.tex`	
Table A.14	`Do/14_analysis.do`	`output/tables/Table_A14.tex`	
Table A.15	`Do/14_analysis.do`	`output/tables/Table_A15.tex`	
Table A.16	`Do/14_analysis.do`	`output/tables/Table_A16.tex`	
Table A.17	`Do/14_analysis.do`	`output/tables/Table_A17.tex`	
Table A.18	`Do/14_analysis.do`	`output/tables/Table_A18.tex`	
Table A.19	`Do/14_analysis.do`	`output/tables/Table_A19.tex`	
Table A.20	`Do/14_analysis.do`	`output/tables/Table_A20.tex`	
Table A.21	`Do/14_analysis.do`	`output/tables/Table_A21.tex`	
Table A.22	`Do/14_analysis.do`	`output/tables/Table_A22.tex`	
Table A.23	`Do/14_analysis.do`	`output/tables/Table_A23.tex`	Note: requires proprietary data.

Table/Figure	Program	Output File	Note
Figure A.1	n.a.	output/figures/Figure_A1.png	Source: author's illustration based on Bräutigam (2011), Zhang and Smith (2017) and Horn et al. (2021)
Figure A.2	Do/14_analysis.do	output/figures/Figure_A2.png	
Figure A.3	Do/14_analysis.do	output/figures/Figure_A3.png	
Figure A.4	Do/14_analysis.do	output/figures/Figure_A4.png	
Figure A.5	Do/14_analysis.do	output/figures/Figure_A5.png	
Figure A.6	Do/14_analysis.do	output/figures/Figure_A6.png	
Figure A.7	Do/14_analysis.do	output/figures/Figure_A7.png	
Figure A.8	Do/14_analysis.do	output/figures/Figure_A8.png	
Figure A.9	Do/14_analysis.do	output/figures/Figure_A9.png	
Figure A.10	Do/14_analysis.do	output/figures/Figure_A10.png	
Figure A.11	Do/14_analysis.do	output/figures/Figure_A11.png	
Figure A.12	Do/14_analysis.do	output/figures/Figure_A12.png	

D Data Availability and Provenance Statement

This section lists all the raw data sources used in the paper and instructions on how to access them. The other datasets included in the replication package are produced by the provided code. The datasets used for the analysis are included in the Data/CleanData folder; the variables therein are clearly labelled and should be self-explanatory. Note that some datasets are proprietary and not shared in the replication package. However, the replication package is structured such that the master .do file can be run without those data. See Section C.3 for instructions if the replicator has obtained the proprietary data.

D.1 Chinese Aid Projects

The main data source for information on Chinese aid projects is AidData's Global Chinese Development Finance Dataset, Version 2.0 (Custer et al., 2021; Dreher et al., 2022). The original data are available in the replication package, located in the file Data/RawData/AidData/AidDatasGlobalChineseDevelopmentFinanceDataset_v2.0.xlsx. The author manually links the aid projects to the firms implementing the projects following the procedure described in the Data Appendix and the files included in this replication package. Some intermediary matching results are in the file Data/RawData/AidData/aiddata2_handmatched_projects.dta. See also 9_aiddata_2_cleaning.do for details on the procedure and sources used. Data/RawData/AidData/handmatched_results/ListofVariables.docx and Data/RawData/AidData/handmatched_results/README.docx provide further details on the variables and matching procedure.

Data Files:

- Data/RawData/AidData/AidDatasGlobalChineseDevelopmentFinanceDataset_v2.0.xlsx
- Data/RawData/AidData/aiddata2_handmatched_projects.dta

D.2 Unrest Data

The data on labor unrest events in Chinese prefectures are from two sources. Data up to 2011 was obtained from Elfstrom (2017). Data after 2011 was obtained from China Labour Bulletin (2019). The author linked the geo-ID of each unrest event in the original data to the corresponding Chinese prefecture for data up to 2011 (Data/RawData/Strikes/China_strikes_prefectures.dta). For data after 2011, in addition to the provided geo-data, the author identifies prefectures in which the events took place based on their event descriptions, as described in the Data Appendix (Data/RawData/Strikes/CLB_additional_citynames.dta). The data described are stored in the folder Data/RawData/Strikes/.

Data Files:

- Data/RawData/Strikes/China_Strikes_2003_2012_elfstrom.csv
- Data/RawData/Strikes/CLB_StrikeMap.2020-10-12.xlsx
- Data/RawData/Strikes/China_strikes_prefectures.dta
- Data/RawData/Strikes/CLB_additional_citynames.dta

D.3 Prefecture-level Administrative Data

Data on prefecture-year level characteristics, including population, GDP, public income and spending, average wage, and prefectures' English names, are from the China City Statistical Yearbooks by the National Bureau of Statistics of China (China Statistics Press, 2020). The data are provided in the replication package, stored in the folder `Data/RawData/China_City_Statistical_Yearbooks/`.

Data Files:

- `Data/RawData/China_City_Statistical_Yearbooks/China_city_directory.dta`
- `Data/RawData/China_City_Statistical_Yearbooks/china_city_statistical_yearbook.dta`

D.4 Chinese Trade Data

Chinese customs data are from the China Import & Export Trade Data database by China Customs. The data are subject to a redistribution restriction, but researchers can access the data through data portals available at major Chinese universities and through data vendors for a fee.

Data Files:

- `Data/RawData/Chinese_Customs_Data/data_city_year/xxxx_city_origin.dta`, xxxx equals to 2000 to 2015 (not shared)

D.5 Firm-level Administrative Data

Firm-level data on employment for the list of firms matched and requested for the paper (`Data/RawData/TaxSurvey/sample_firm_list.dta`) are from the National Tax Survey Database by the State Administration of Taxation of China and the Ministry of Finance of China (SAT-MOF). The data are subject to a redistribution restriction, but researchers can access the data through data portals available at major Chinese universities and through data vendors for a fee. The file `Data/RawData/TaxSurvey/sample_firm_list.dta` includes firm identifiers the replicator may use to merge in the tax survey data if they obtain the data. Replicators who obtain the data can contact the author to request this list for replication.

Data Files:

- `Data/RawData/TaxSurvey/sample_firm_list.dta` (not shared)
- `Data/RawData/TaxSurvey/tax_data_sample/data_xxxx.dta`, xxxx equals 2007 to 2015 (not shared)

D.6 List of Licensed Chinese Firms

The list of Chinese firms licensed to conduct overseas construction projects that served as an auxiliary dataset to help construct the firm-level panel, which is described in the paper, is from the Chinese Ministry of Commerce (MOFCOM) and was constructed by the author based on data from the publicly accessible website `http://xzsx.mofcom.gov.cn:80/xzsp/advSearch.jhtml` in June 2020. The data is provided in the replication package, which is stored as `Data/RawData/MOFCOM/mofcom_license_list_2018.csv`.

Data File:

- `Data/RawData/MOFCOM/mofcom_license_list_2018.csv`

D.7 Firm Registry Data

The firm registry data described in the paper, which served as an auxiliary dataset to help construct the firm-level panel, is provided with permission by Liu et al. (2022), who compiled the data from Tianyancha, a publicly accessible database on Chinese firm registrations, licensed by China's central bank. The shared data include firms' basic registry information and ultimate ownership, which was determined by Liu et al. (2022). The raw data on current and historical shareholding records for individual firms are not provided, but can be obtained by subscribing to Tianyancha. The data described in this paragraph are stored in the folder `Data/RawData/Tianyancha/`.

Data Files:

- `Data/RawData/Tianyancha/firm_matched_tianyancha.dta`
- `Data/RawData/Tianyancha/data/ultimate_owner.dta`

- Data/RawData/Tianyancha/data/shareholders_records.dta (not shared)
- Data/RawData/Tianyancha/data/shareholding_xxx.dta, xxx equals pre and post (not shared)

D.8 Firm Annual Reports

The keyword counts extracted from firms' annual reports are based on data from Mueller et al. (2023). The underlying annual reports are available from providers such as Wind Financial Terminal and are subject to a redistribution restriction, but can be obtained from these providers for a fee. The keyword counts from firms' annual reports are stored in the folder Data/RawData/AnnualReports/.

Data Files:

- Data/RawData/AnnualReports/keyword_ar_basic.xlsx
- Data/RawData/AnnualReports/keyword_phrases.xlsx
- Data/RawData/AnnualReports/listed_firms.dta

D.9 Procurement Data

Data on Chinese government procurement bids were collected by the author from the publicly accessible Chinese Government Procurement website https://www.ccgp.gov.cn/ in January 2024. Since the raw data contains a very large amount of unstructured text, only the processed data that include bids assigned to firms in the sample are provided. The financial value of each contract is extracted by feeding the unstructured text into ChatGPT model gpt-4o-mini. The program used to process the data is provided in Do/auxiliary/clean_procurement_contract.py. The log file of ChatGPT's response is stored in Data/RawData/Procurement/openai_log_procurement.txt.

Data File:

- Data/RawData/Procurement/matched_contracts_value.xlsx

D.10 Weather Data

Data from Chinese weather stations are from the National Oceanic and Atmospheric Administration (NOAA) and were originally collected by the World Meteorological Organization (National Centers for Environmental Information, 2020). The raw data can be freely accessed at https://www.ncei.noaa.gov/cdo-web/. The processed data assigning daily weather to Chinese prefectures are provided and stored in the folder Data/RawData/WMONOAAData/.

Data File:

- Data/RawData/WMONOAAData/noaa_daily_weather.dta

D.11 Aid Recipient Country Trade Data

Data on countries' trade with China are from Gaulier and Zignago (2010). The data were downloaded by the author in February 2024 (version 202401) from https://www.cepii.fr/DATA_DOWNLOAD/baci/legacy/202401/BACI_HS96_V202401.zip. The data files are stored in the folder Data/RawData/BACI_HS96_V202401/.

Data File:

- Data/RawData/BACI_HS96_V202401/BACI2000-2022.dta

D.12 Other Country-level Variables

The paper uses several further country-level variables, including exchange rates, GDP deflators, population, net inflow of FDI, and net bilateral aid flows from DAC donors, from the World Bank (World Bank, 2022). The data were downloaded by the author from https://databank.worldbank.org/source/world-development-indicators in February 2024. The data files containing the selected variables are stored in the folder Data/RawData/WDI_csv/.

Data Files:

- Data/RawData/WDI_csv/chinaexchangerate.xlsx
- Data/RawData/WDI_csv/deflatortable.xls
- Data/RawData/WDI_csv/WDI{xxxx-xxxx}.csv, xxxx-xxxx equals 2000-2004, 2005-2010, 2011-2016, 2017-2023
- Data/RawData/WDI_csv/wdi_indicators.dta

D.13 Crosswalks of Firm Names, Chinese Prefectures, and Country Names

Firm Crosswalk. The author manually created a crosswalk between firm names in the MOFCOM list of firms and firm registry data. This crosswalk is stored as Data/RawData/MOFCOM/aid_mofcom_firm_name_manual_cross_walk.dta. The information sources used to construct the crosswalk are included in the file.

Prefecture Crosswalk. The author manually created a crosswalk to merge Chinese prefectures across different datasets using names, GB codes and coordinates according to Chinese prefecture borders in 2003. This crosswalk is stored as Data/RawData/ChinaShapefile/cityiddirectory.dta.

Country Crosswalks. The author manually linked Chinese country names in the China Customs Database with their ISO 3166-1 alpha-3 country codes. The crosswalk between country names and country codes for the country-level trade data is provided by Gaulier and Zignago (2010). The author also created a list of countries and classification of high-income status based on information provided by World Bank (2022). These files are stored as Data/RawData/Chinese_Customs_Data/add_countrycodes_names.dta, Data/RawData/BACI_HS96_V202401/country_codes_V202401.csv, and Data/RawData/WDI_csv/recipient_country_comp_isocode_cons_correspondence.dta, respectively.

Data Files:

- Data/RawData/MOFCOM/aid_mofcom_firm_name_manual_cross_walk.dta
- Data/RawData/ChinaShapefile/cityiddirectory.dta
- Data/RawData/Chinese_Customs_Data/add_countrycodes_names.dta
- Data/RawData/WDI_csv/recipient_country_comp_isocode_cons_correspondence.dta

D.14 Geospatial Data

D.14.1 Chinese Prefecture Boundaries

The base map of Chinese prefecture borders was downloaded from the Spatial Data Lab (Yuen, 2020) at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GLKQME. The files are stored in the folder Data/RawData/ChinaShapefile/administrative_boundaries/CityMap_0010_census2000-2010/.

The neighbors of each prefecture were determined by the author using the Polygon Neighbors tool in the ArcGIS Desktop 10.8.2 graphical user interface. The dataset is stored as Data/RawData/ChinaShapefile/neighbors.dbf.

Data Files:

- Data/RawData/ChinaShapefile/administrative_boundaries/CityMap_0010_census2000-2010/Map/City0010.XXX, XXX equal to cpg, dbf, prj, sbn, sbx, shp, shx
- Data/RawData/ChinaShapefile/neighbors.dbf

D.14.2 World Map

The map for world administrative boundaries was downloaded from opendatasoft at https://public.opendatasoft.com/explore/dataset/world-administrative-boundaries/export/ (opendatasoft, 2019). The files are stored in the folder Data/RawData/Countries_WGS84/.

Data Files:

- Data/RawData/Countries_WGS84/world-administrative-boundaries.xxx, xxx equals cpg, dbf, prj, shp, shx

D.15 Datasets List

Table 2: Datasets, Sources and Availability

Data file	Source	Notes	Provided
Data/RawData/AidData/AidDatasGlobalChineseDevelopmentFinanceDataset_v2.0.xlsx	AidData	AidData's Global Chinese Development Finance Dataset, Version 2.0	Yes
Data/RawData/AidData/handmatched_results/aiddata2_handmatched_projects.dta	Author	AidData projects matched with their contractors	Yes
Data/RawData/AnnualReports/keyword_ar_basic.xlsx	Mueller et al. (2023)	Word counts in firm annual reports	Yes
Data/RawData/AnnualReports/keyword_phrases.xlsx	Author	Keywords selected	Yes
Data/RawData/AnnualReports/listed_firms.dta	Mueller et al. (2023)	List of listed firms, their security tickers, and prefecture locations	Yes
Data/RawData/BACI_HS96_V202401/BACI2000-2022.dta	CEPII BACI Database	Bilateral trade data between countries	Yes
Data/RawData/ChinaShapefile/cityiddirectory.dta	Author	Crosswalk between Chinese prefecture code used across different data sources	Yes
Data/RawData/ChinaShapefile/neighbors.dbf	Author	List of neighboring prefectures for each Chinese prefecture	Yes
Data/RawData/ChinaShapefile/administrative_boundaries/City0010.XXX, XXX equal to cpg, dbf, prj, sbn, sbx, shp, shx	Spatial Data Lab	China prefecture map	Yes
Data/RawData/China_City_Statistical_Yearbooks/China_city_directory.dta	National Bureau of Statistics of China	China prefecture boundaries and English names	Yes
Data/RawData/China_City_Statistical_Yearbooks/china_city_statistical_yearbook.dta	National Bureau of Statistics of China	China prefecture-level administrative data	Yes
Data/RawData/Chinese_Customs_Data/data_city_year/xxxx_city_origin.dta, xxxx equals to 2000 to 2015	China customs	Data on prefecture-level yearly export	No
Data/RawData/Chinese_Customs_Data/add_countrycodes_names.dta	Author	Crosswalk between Chinese country name and country code	Yes
Data/RawData/Countries_WGS84/world-administrative-boundaries.xxx, xxx equals cpg, dbf, prj, shp, shx	opendatasoft	World country map	Yes
Data/RawData/MOFCOM/mofcom_license_list_2018.csv	China Ministry of Commerce	List of Chinese firms licensed by MOFCOM	Yes
Data/RawData/MOFCOM/aid_mofcom_firm_name_manual_cross_walk.dta	Author	Crosswalk between MOFCOM firm names and aid contractors in AidData data	Yes
Data/RawData/Procurement/matched_contracts_value.xlsx	Chinese Government Procurement website	Record of procurement contracts assigned to MOFCOM firms	Yes

Continued on next page

Table 2 – Continued from previous page

Data file	Source	Notes	Provided
Data/RawData/Strikes/China_Strikes_2003_2012_elfstrom.csv	Manfred Elfstrom	Data on Chinese unrest up to 2011	Yes
Data/RawData/Strikes/CLB_StrikeMap.2020-10-12.xlsx	China Labour Bulletin	Data on Chinese unrest after 2011	Yes
Data/RawData/Strikes/China_strikes_prefectures.dta	Author	Crosswalk between the geo-ID of unrest event and the corresponding prefecture	Yes
Data/RawData/Strikes/CLB_additional_citynames.dta	Author	Crosswalk between unrest event description and corresponding prefecture	Yes
Data/RawData/TaxSurvey/sample_firm_list.dta	Author	List of MOFCOM firms matched to tax survey data	Yes
Data/RawData/TaxSurvey/tax_data_sample/data_xxxx.dta, xxxx equals 2007 to 2015	SAT-MOF	Tax survey data	No
Data/RawData/Tianyancha/firm_matched_tianyancha.dta	Liu et al. (2022) (2022)	Firm registry data	Yes
Data/RawData/Tianyancha/data/shareholders_records.dta	Liu et al. (2022) (2022)	Current shareholding data	No
Data/RawData/Tianyancha/data/shareholding_xxx.dta, xxx equals pre and post	Liu et al. (2022)	Firms' historical shareholding data before and after share transactions	No
Data/RawData/Tianyancha/data/ultimate_owner.dta	Liu et al. (2022)	Firms' ownership data	Yes
Data/RawData/WDI_csv/chinaexchangerate.xlsx	World Bank	CNY to USD exchange rate	Yes
Data/RawData/WDI_csv/deflatortable.xls	World Bank	World GDP deflator	Yes
Data/RawData/WDI_csv/WDI{xxxx-xxxx}.csv, xxxx-xxxx equals 2000-2004, 2005-2010, 2011-2016, 2017-2023	World Bank	World Development Indicators	Yes
Data/RawData/WDI_csv/wdi_indicators.dta	World Bank	World Development Indicators codebook	Yes
Data/RawData/WDI_csv/recipient_country_comp_isocode_cons_correspondence.dta	World Bank	Classification of country region, income, and OECD membership	Yes
Data/RawData/WMONOAAData/noaa_daily_weather.dta	NOAA	China daily weather data	Yes

References

Bräutigam, D. (2011). *The Dragon's Gift: The Real Story of China in Africa*. Oxford University Press.

China Labour Bulletin (2019). China labour bulletin strike map. *Computer file. Available at: https://https://maps.clb.org.hk*.

China Statistics Press (2020). *China City Statistical Yearbooks 2003–2019*. Beijing, China.

Custer, S., Dreher, A., Elston, T., Fuchs, A., Ghose, S., Lin, J., Malik, A., Parks, B., Russell, B., Solomon, K., Strange, A., Tierney, M., Walsh, K., Zaleski, L., and Zhang, S. (2021). Tracking chinese development finance: An application of aiddata's tuff 2.0 methodology. *Williamsburg, VA: AidData*.

Dreher, A., Fuchs, A., Parks, B., Strange, A., and Tierney, M. J. (2022). *Banking on Beijing*. Cambridge University Press.

Elfstrom, M. (2017). China strikes. *Computer file. Available at: https://chinastrikes.crowdmap.com*.

Gaulier, G. and Zignago, S. (2010). BACI: International Trade Database at the Product-Level. The 1994–2007 Version. CEPII Working Paper 2010-23, CEPII.

Horn, S., Reinhart, C. M., and Trebesch, C. (2021). China's overseas lending. *Journal of International Economics*, 133:103539.

Liu, E., Lu, Y., Peng, W., and Wang, S. (2022). Judicial independence, local protectionism, and economic integration: Evidence from china. *National Bureau of Economic Research*.

Mueller, J., Wen, J., and Wu, C. (2023). The party and the firm. *HBS Working Paper*.

National Centers for Environmental Information (2020). Wmo climate normals. `https://www.ncei.noaa.gov/data/oceans/archive/arc0216/0253808/6.6/data/0-data/Region-2-WMO-Normals-9120/`. online; accessed 9-18-2022.

opendatasoft (2019). `https://geonode.wfp.org/layers/geonode%3Awld_bnd_adm0_wfp`. online; accessed 9-18-2022.

World Bank (2022). *World Development Indicators*. The World Bank, Washington D.C.

Yuen, S. (2020). China Prefecture Map with 2000-2010 Population Census Data. `https://doi.org/10.7910/DVN/GLKQME`. online; accessed 9-18-2022.

Zhang, D. and Smith, G. (2017). China's foreign aid system: Structure, agencies, and identities. *Third World Quarterly*, 38(10):2330–2346.