

README: Venture Capital Around the World Reproducibility Package

Introduction

This reproducibility package accompanies the paper Venture Capital Around the World: How It Differs in Developing Countries. The package is organized so that the full analysis can be launched from a single Stata master file, with code, data, temporary files, and outputs separated into dedicated folders.

The recommended entry point is Code\Main_dofile.do. This file defines the project paths, installs any required user-written Stata packages, creates output folders when needed, applies a common graph style, and runs the packaged analysis file Code\01_analysis.do.

Main figures are written to Output\Main\Figures and main tables are written to Output\Main\Tables. Temporary intermediate files are written to Data\TempData during execution.

Requirements

- **Software:** Use Stata 18.5 or a compatible version that supports the commands used in the scripts.
 - Main_dofile.do installs the required user-written packages, including coefplot, winsor2, outreg2, estout, spmap, sencode, grstyle, palettes, and colrspace.

Folder Structure

- Code
 - Main_dofile.do
 - 01_analysis.do
- Data
 - RawData
 - TempData
 - AuxData
- Output
 - Main
 - Tables
 - Figures
- ReadMe
- Paper

Code

Filename 1: Main_dofile.do

This is the master do-file for the package. It defines the global paths used throughout the project, installs required user-written Stata packages, creates output and temporary folders, opens a log file, applies a common graph style, and launches the main analysis.

Filename 2: 01_analysis.do

This is the packaged main analysis script. It reads the harmonized venture capital data, constructs temporary analysis datasets, merges auxiliary country-level data, and produces the tables and figures used in the main paper.

Data

Filename 1: Harmonized_Feb2026.dta

- **Source:** Calculations from PitchBook, a Morningstar company, and Crunchbase.
- **URL to request access:**
 - [PitchBook](#)
 - [Crunchbase](#)
- **Path in folder:** This auxiliary dataset should be stored in Data\RawData.
- **Access year:** Pitchbook (2025) and Crunchbase (2023)
- **Description:** It contains the harmonized venture capital deal-level data used throughout the analysis, including deal sizes, company identifiers, years, sectors, and country classifications.
- **PitchBook Disclaimer:** The cited data has not been reviewed by PitchBook analysts and may be inconsistent with PitchBook methodology.
- **PitchBook Access Details:** PitchBook data require access through a paid PitchBook subscription or an institutional license. The data used in this paper were accessed through licensed PitchBook services for research purposes, covering private market activity, including venture capital and private equity. This subscription is managed by members of DEC Private Markets, World Bank Group.
- **Crunchbase Access Details:** Crunchbase data require access through a paid Crunchbase subscription, API license, or institutional data agreement. The data used in this paper were accessed through licensed Crunchbase services, covering venture capital activity. This subscription is managed by members of DEC Private Markets, World Bank Group.

Filename 2: EquityBonds_CountrySectorYear_2010.dta

- **Source:** Calculations from the Securities Data Platinum Company of the London Stock Exchange Group (LSEG).
- **URL to request access:** <https://www.lseg.com/en/data-analytics/products/sdc-platinum-financial-securities>
- **Path in folder:** This auxiliary dataset should be stored in Data\AuxData.
- **Access year:** December 2025
- **Description:** It provides country-sector-year information on equity market activity and is used in analyses related to equity issuance, sector composition, and comparisons between venture capital activity and broader equity market outcomes.
- **Access Details:** LSEG issuance and deals data require access through LSEG Workspace and/or the SDC Platinum add-on, depending on the institution's subscription. SDC Platinum is a proprietary LSEG product used to monitor deal flow and analyze transaction activity across regions, asset classes, and industries. The subscription is managed by members of DEC Private Markets, World Bank Group.

Filename 3: WDI_agr.dta

- **Source:** World Development Indicators (WDI) of the World Bank Group
- **URL:** [Website](#)
- **Path in folder:** This auxiliary dataset should be stored in Data\AuxData.
- **Access year:** December 2025
- **Description:** It contains cleaned country-level aggregates for working-age population, GDP, and GDP per capita across the analysis periods. The analysis repeatedly merges this file to apply population thresholds and construct country-level normalization variables.
- **Disclaimer:** All calculations, harmonization procedures, and resulting measures are undertaken by the authors and do not represent official estimates of the underlying data providers.

Filename 4: CFLosses.dta

- **Source:** PricewaterhouseCoopers, Worldwide Tax Summaries
- **Date Accessed:** December 2025
- **URL:** <https://taxsummaries.pwc.com/>
- **Path in folder:** This auxiliary dataset should be stored in Data\AuxData.
- **Description:** Country-level tax loss carryforward data, used to construct the variable `noL_cf_years`, which measures the number of years firms can carry forward tax losses. In the analysis, this is used as the "Tax loss carryforward (years)" variable in the business environment regressions.

Filename 5: FD Database (STATA).dta

- **Source:** International Monetary Fund, Financial Development Index Database
- **Date Accessed:** December 2025
- **URL:** <https://data.imf.org/en/datasets/IMF.MCM:FDI>
- **Path in folder:** This auxiliary dataset should be stored in Data\AuxData.
- **Description:** Country-level Financial Development Index data from the IMF. The paper uses this dataset to compare VC/GDP with overall financial development in Figure 5. The IMF database includes indices summarizing financial institutions and financial markets by depth, access, and efficiency.

Filename 6: LMrestrictions.dta

- **Source:** World Bank World Development Indicators
- **Date Accessed:** December 2025
- **URL:** <https://databank.worldbank.org/source/world-development-indicators>
- **Path in folder:** This auxiliary dataset should be stored in Data\AuxData.
- **Description:** Country-level labor market restriction data.

Filename 7: MinorityInvestors_CountryYear.dta

- **Source:** World Bank Doing Business
- **Date Accessed:** December 2025
- **URL:** <https://databank.worldbank.org/source/doing-business>
- **Description:** Country-year data on minority investor protection.

Filename 8: PrivateEquity.dta

- **Source:** Calculations from the Securities Data Platinum Company of the London Stock Exchange Group (LSEG)
- **Date Accessed:** December 2025
- **URL:** <https://www.lseg.com/en/data-analytics/products/sdc-platinum-financial-securities>
- **Description:** Country-year private equity deal data used to calculate private equity growth by period. In the code, the file is collapsed to country-period PE growth and merged with VC and equity issuance data to construct the VC share of total equity markets.

Filename 9: PwC.dta

- **Source:** PricewaterhouseCoopers, Worldwide Tax Summaries
- **Date Accessed:** December 2025
- **URL:** <https://taxsummaries.pwc.com/>
- **Description:** Country-level corporate tax data. In the code, this file provides eff_corp_cgt and cit_rate, labeled as “Effective capital gains tax rate (pct)” and “Corporate income tax rate (pct).”

Filename 10: WB_GAD_ADM0_complete_shp.dta

- **Source:** Database of Global Administrative Areas (GADM)
- **Date Accessed:** December 2025
- **URL:** <https://gadm.org/index.html>
- **Description:** Stata-format shapefile used for world map figures in the paper. In the code, this file is used with spmap to draw Figure 4 and Figure 7.

Filename 11: WB_GAD_ADM0_complete_data.dta

- **Source:** Database of Global Administrative Areas (GADM)
- **Date Accessed:** December 2025
- **URL:** <https://gadm.org/index.html>
- **Description:** Stata-format country attribute file paired with WB_GAD_ADM0_complete_shp.dta. In the code, this file is merged with country-level analysis variables before using spmap.

Filename 12: WGI2021.dta

- **Source:** World Bank, Worldwide Governance Indicators
- **Date Accessed:** December 2025
- **URL:** <https://databank.worldbank.org/source/worldwide-governance-indicators>
- **Description:** Country-level governance data from the Worldwide Governance Indicators.

Notes

- **Data Disclaimer:** All calculations, harmonization procedures, and resulting measures are undertaken by the authors and do not represent official estimates of the underlying data providers.
- **Data:** The main analysis window is 2012 to 2023, countries below the configured population threshold are excluded, and several temporary datasets are generated automatically in Data\TempData during the run.
- **Rights statement:** We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.