

Readme file- Reproducible package

Readme File for Replication of the Paper: “Small Emerging Markets: A New Asset Class”

This readme file was created to guide the reproducibility package for the analysis of the paper.

Authors: Alvaro Pedraza and Dilip Ratha

Overview

This paper examines whether small low- and middle-income countries (LMICs) offer excess returns and diversification benefits to global investors. Using monthly equity price data from 2015–2025, we pool listed stocks from LMICs whose populations fall below a given threshold and construct value-weighted portfolios, sorted by population and land area. We then estimate a two-factor asset-pricing model to evaluate their risk-adjusted performance relative to developed- and emerging-market benchmarks. Portfolios restricted to countries with fewer than 20 million people –or less than 100,000 square kilometers– earn positive and economically meaningful alphas alongside persistently low covariances with global-market returns. Risk-adjusted performance, measured by an adjusted Treynor ratio, peaks in the 15–20 million population range. Once larger LMICs enter the portfolio, integration with the global markets rises sharply, alphas converge to zero, and diversification benefits vanish. No analogous pattern appears among high-income countries, where even very small markets are fully integrated. These results point to a “Small Emerging Market Asset Class”: small developing economies remain partially segmented, exposing investors to undiversifiable country-specific risks stemming from structural vulnerability (e.g., heightened climate and environmental risk), and for which they appear to be compensated. Small LMICs thus form a distinct frontier for international portfolio diversification.

Data Sources and Access

This project combines publicly available country-level data with proprietary financial market data.

Publicly Available Data

1. **World Development Indicators (WDI).** Country-level macroeconomic and geographic variables are obtained from the World Bank's World Development Indicators. These include surface area, GDP (constant 2015 US dollars), and GDP per capita (current US dollars). The dataset is stored as Country_gdp_area.xlsx. All variables are measured for **year 2014**. Data were downloaded on December 12, 2025. Source: <https://databank.worldbank.org/source/world-development-indicators>
2. **Climate and Environmental Vulnerability.** Country-level vulnerability measures are obtained from the Notre Dame Global Adaptation Initiative (ND-GAIN) Country Index. We use the vulnerability component of the index, which captures countries' exposure to environmental hazards, sensitivity of key systems, and adaptive capacity across multiple sectors. The data are sourced from the ND-GAIN bulk download archive, file path resources/vulnerability/vulnerability.csv, and correspond to the **2015 index values**. Data were downloaded on December 12, 2025. Source: <https://gain.nd.edu/our-work/country-index/download-data/>
3. **Financial Openness.** Capital account openness is measured using the Chinn–Ito (KAOPEN) index. We use country-level values of the KAOPEN variable for **year 2015**, reflecting de jure restrictions on cross-border financial transactions. Data were downloaded on December 12, 2025. Source: https://web.pdx.edu/~ito/Chinn-Ito_website.htm
4. **Export Concentration.** Economic concentration is measured using the export product concentration index compiled by UNCTAD and distributed through UNCTADstat. The index is based on a Herfindahl–Hirschman formulation of merchandise export shares. We use annual country-level values for **year 2015**. Data were downloaded on December 12, 2025. Source: <https://unctadstat.unctad.org/datacentre/> (UNCTADstat → Merchandise trade → Product concentration and diversification indices)

For variables described in items (2), (3), and (4), we construct a single country-level dataset, **Channels.dta**, by merging the three public data sources at the country level. Climate and environmental vulnerability is merged using the ISO 3-digit country code (ISO3) from the ND-GAIN data, and the 2015 vulnerability index is stored as variable vulnerability. Financial openness is merged using the 3-digit country code variable ccode from the Chinn–Ito dataset, and the 2015 KAOPEN value is stored as variable

kopen. Export concentration is merged using country names (Economy Label) from UNCTADstat matched to the variable country_name, and the 2015 export product concentration index is stored as variable concentration. All three sources contain public, country-level information. The resulting dataset **Channels.dta** includes the variables iso3, country, vulnerability, kopen, and concentration, and is fully public and can be included in the reproducibility package.

Proprietary Data

1. **Bloomberg.** Equity market data are obtained from Bloomberg and include monthly closing prices of major country equity indices, end-of-month exchange rates against the US dollar, market capitalization of index constituents at month-end, index tickers, and the US short-term interest rate. Processed datasets include Data.xlsx and msc_global_eme.dta, which contain country-level returns and developed- and emerging-market benchmark indices. Bloomberg data are proprietary and subject to copyright restrictions and therefore cannot be redistributed as part of the replication package. Researchers with Bloomberg access can reproduce these data using the documented tickers and variable definitions provided in the code.

Summary of Data Inputs

Data Input	Description	File name
Bloomberg	Equity market data: prices, exchange rates, market cap, interest rates.	Data/Raw/data.xlsx
Bloomberg	MSCI Developed and Emerging Market Indices	Data/Raw/msc_global_eme.dta
WDI	Country GDP, GDP per capita, land area and population	Data/Raw/country_gdp_area.xlsx
ND-GAIN, KAOPEN, UNCTADstat	Country vulnerability, openness and concentration index	Data/Raw/Channels.dta

Instructions for Replicators

Code

The analysis in this paper is executed through the master script “**Code/Main.do**”, which organizes all steps required to build the analytic dataset and reproduce every table and figure in the paper. All directory paths are defined at the top of the script, and users should update the global project path before running.

All code files are organized sequentially:

- **1_Data_preparation.do**
- **2_Motivation_figures.do**
- **3_Threshold_population.do**
- **4_GDP_percapita_area.do**
- **5_Time_series_correlation.do**
- **6_Channels.do**

Computational settings

The script was developed and tested using **Stata version MP 18.0** on a **64-bit Windows system** (MinGW environment) with **32 GB of RAM (32000 MB)**. While the script ran successfully in this environment, we expect it would also run with less RAM.

Under the described settings, the reproducibility package takes around **2 minutes** to run from beginning to end.

List of Exhibits

Exhibit name	Output file name	Script	Folder Path
Table 1. Summary statistics by country	Sum_stats.csv	3_Threshold_population.do Part 3	/Outputs/Tables
Table 2. Correlation between	Correlation_Matrix.xlsx	6_Channels.do	/Outputs/Tables

country
characteristics
and risk-
adjusted
returns

Table 3. Cross-country regressions of risk-adjusted returns	Findings.xlsx	6_Channels	/Outputs/Tables
Figure 1. Emerging markets in the last 30 years	Panel A: Correlation3.pdf Panel B: Alpha_eme.pdf	2_Motivation_figures.do	/Outputs/Figures
Figure 2. Covariance of P portoflios	Panel A: covariance_pop_lmic.pdf Panel B: beta_pop_lmic.pdf	3_Threshld_population Part 1	/Outputs/Figures
Figure 3. Risk-adjusted returns by population threshold	Alpha_pop_lmic.pdf	3_Threshld_population Part 1	/Outputs/Figures
Figure 4. Risk-adjusted returns per unit of risk	Covariance_pop_treynor.pdf	3_Threshld_population Part 1	/Outputs/Figures
Figure 5. Analysis for HIC	Panel A: covariance_pop_hic.pdf Panel B: beta_pop_hic.pdf	3_Threshld_population Part 2	/Outputs/Figures
Figure 6. Portfolios are sorted by GDP per capita, GDP, Land Area, and Vulnerabilty	Panel A: alpha_GDPcapita.pdf Panel C: alpha_GDP.pdf Panel E: alpha_area.pdf Panel G: alpha_vulnerability.pdf Panel B: covar_GDPcapita.pdf Panel D: covar_GDP.pdf Panel F: covar_area.pdf Panel H: covar_vulnerability.pdf	4_GDP_percapita_area.do	/Outputs/Figures
Figure 7. Time series correlation between sample countries and	time_correlation.pdf	5_Time_series_correlation	/Outputs/Figures

Folder structure

C:\Users\WB468173\OneDrive - WBG\Documents\Cross support\Small States\Replication Package

- Code
 - 1_Data_preparation.do
 - 2_Motivation_figures.do
 - 3_Threshold_population.do
 - 4_GDP_percapita_area.do
 - 5_Time_series_correlation.do
 - 6_Channels.do
 - Main.do
- Data
 - Data/Raw
 - Channels.dta
 - Country_gdp_area.xlsx
 - Data.xlsx
 - Msc_global_eme.dta
 - Data/Temp
 - country_code.dta
 - results.dta
 - Countries_HIC.dta
 - reg_results_by_pop.dta
 - Countries_LMIC.dta
 - n_dates.dta
 - country_gdp_area.dta
 - Ex_rate.dta
 - Full_data.dta
 - Income_class.dta
 - Index_country.dta
 - Market_ret.dta
 - Mkt_cap.dta
 - Index.dta

- Population.dta
 - eu_data.dta
 - eu_rates.dta
 - Countries.dta
- Readme
 - Readme.docx
- Outputs
 - Outputs/Figures
 - alpha_area.pdf
 - alpha_eme.pdf
 - alpha_GDP.pdf
 - alpha_GDPcapita.pdf
 - alpha_pop_lmic.pdf
 - alpha_vulnerability.pdf
 - beta_pop_hic.pdf
 - beta_pop_lmic.pdf
 - correlation3.pdf
 - covar_area.pdf
 - covar_GDP.pdf
 - covar_GDPcapita.pdf
 - covar_vulnerability.pdf
 - covariance_pop_hic.pdf
 - covariance_pop_lmic.pdf
 - covariance_pop_treynor.pdf
 - time_correlation.pdf
 - Outputs/Tables
 - Correlation_Matrix.xlsx
 - Findings.xlsx
 - Sum_stats.csv