

README

Why Do Some Countries Build Safer? Economic Constraints, Disaster Learning, and the Two-Stage Housing Quality Ladder

Overview

This study constructs a cross-country dataset for 150 economies that includes all variables required for the empirical analysis.

1. Dependent Variable: Housing Robustness

Construction from GEM Building Taxonomy

The paper constructs housing robustness measures from the Global Exposure Model (GEM) developed by the Global Earthquake Model Foundation (GEM Foundation, 2023), accessed via its GitHub repository (https://github.com/gem/global_exposure_model). The GEM database provides comprehensive information on residential building stocks at the national level for 215 countries, including building taxonomy (structural system, materials, height, lateral load resistance), unit counts, and replacement costs. The 2023 release incorporates major methodological improvements implemented in 2022 using 2021 economic values and updated building inventories.

The paper assigns each GEM taxonomy to vulnerability categories (robust, medium, or fragile) using the GEM-to-vulnerability mapping developed by the World Bank Unbreakable study (Middelani et al., 2024). This mapping links structural characteristics to expected disaster performance under three hazard groups—earthquake, wind/storm, and water-related hazards (flood, storm surge, tsunami)—plus an overall combined vulnerability index. The structural characteristics include:

- Material and lateral load system
- Structural system
- Height information

The main specifications use the combined index, creating two categories to test the two-stage housing ladder framework:

1. **Robust + medium quality category** (eliminating fragile housing)
 2. **Robust-only category** (achieving resilient standards)
-

2. Independent Variables

Economic Development

GDP per capita, PPP-adjusted (constant 2021 USD) from World Bank World Development Indicators (WDI), using the most recent year before 2023 per country. The paper employs $\log(\text{GDP per capita})$.

Poverty

Poverty Gap at \$4.20/day (2021 PPP), from World Bank Poverty and Inequality Platform, using the most recent year before 2023 per country. The author uses the Poverty Gap instead of poverty headcount to avoid multicollinearity with GDP and to better understand the implications of poverty depth on housing robustness. \$4.20/day represents the median national poverty line among lower-middle-income countries (World Bank, 2022), thereby capturing a broader measure of deprivation than the international extreme poverty line (\$3.00/day). Housing affordability constraints bind not only for the extremely poor but for the broader lower-income population unable to access formal housing finance.

Urbanization

Proportion of population living in urban areas as defined by national statistical offices in each country from the World Bank data, using the most recent year before 2023 per country.

Institutional Quality

Government Effectiveness from the Worldwide Governance Indicators (WGI), 2024 update. Government Effectiveness captures “perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government’s commitment to such policies” (Kaufmann & Kraay, 2024). The paper selects Government Effectiveness as it most directly captures the state’s capacity to formulate, implement, and enforce the complex technical regulations, land-use policies, and public service provisions that underpin a resilient housing stock.

Building Quality Control

Building Quality Control Index (BQCI) from the World Bank Doing Business Archive (2020, last year published before indicator discontinuation). Scale: 0–15, with higher values indicating more stringent and well-enforced building regulations. Despite discontinuation, the paper includes this index because it remains the most recent comprehensive, cross-country measure of building-specific regulatory quality, and the year 2020 aligns with the timing of GEM taxonomy data. The index consists of:

1. Quality of building regulations
2. Quality control before construction
3. Quality control during construction
4. Quality control after construction
5. Liability and insurance regimes
6. Professional certification

Disaster Experience

The paper uses EM-DAT data to construct cumulative national disaster exposure for 2000–2020, aligning with the timing of the GEM data. EM-DAT records “significant” disasters satisfying at least one criterion: 10 deaths, or 100 people affected, or declaration of a state of emergency, or call for international assistance. This threshold excludes minor events while capturing disasters likely to trigger policy responses and societal learning. The paper focuses on three hazard types that directly affect housing structural integrity:

- Earthquakes
- Storms (tropical cyclones, hurricanes, typhoons)
- Floods

The primary exposure metric is the count of significant events in 2000–2020 by hazard type.

Data Availability Statement

Dataset Name	File Name(s)	Link/Access Instructions	Access Date	Data Access
Worldwide Governance Indicators	wgidataset.xlsx	World Bank WGI	2025-10-31	Open

Dataset Name	File Name(s)	Link/Access Instructions	Access Date	Data Access
Dealing with Construction Permits - World Bank Group	Dealing with Construction Permits.xlsx	World Bank Archive	2025-11-05	Open
World Bank Region Classification	CLASS.xlsx	World Bank	2025-10-23	
building_class_to_vulnerability_mapping	building_class_to_vulnerability_mapping.csv	GitHub Repository	2025-10-23	Open
country_vulnerability_classes	country_vulnerability_classes.csv	GitHub Repository	2025-10-23	Open
gem_taxonomy_fields	gem_taxonomy_fields.json	GitHub Repository	2025-10-23	Open
hazus-gem_mapping	hazus-gem_mapping.csv	GitHub Repository	2025-10-23	Open
gem-to-vulnerability_mapping_per_hazard	gem-to-vulnerability_mapping_per_hazard.xlsx	GitHub Repository	2025-10-23	Open
EM-DAT hazard data	public_emdat_custom_request_2025-11-07_f6f2cabe-ff99-4bc3-ba01-ab018fae62ec.xlsx	EM-DAT	2025-11-07	Accessible
MPI	MPI_national.xlsx	OPHI Global MPI	2025-11-07	Open

Statement About Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Instructions for Replicators

Six Jupyter notebooks (.ipynb) need to be run in the following order:

Step 1: Data Access and Cleaning

Run the following notebooks to access and clean each dataset:

1. **gather_gem_country.ipynb** – to retrieve GEM data
2. **WB_clean.ipynb** – to retrieve data from the World Bank
3. **Em-dat_clean_country_only.ipynb** – to process data from EM-DAT
4. **MPI_clean.ipynb** – to clean MPI data

Step 2: Data Merging

5. **Country_merge.ipynb** – to merge all datasets into one

Step 3: Analysis

6. **Country_regression-RegionalFE.ipynb** – to run all analysis from summary statistics to regression and robustness checks
-

Expected Outputs

Running the analysis notebook (**Country_regression-RegionalFE.ipynb**) will generate the following files in the **results/** folder:

Summary Statistics and Descriptive Analysis

- **summary_statistics_categorized.csv** – Summary statistics table (Table 1 in the main paper)
- **correlation_matrix_publication.csv** – Correlation matrix (Table A-1 in the appendix)
- **dropped_countries_analysis.csv** – List of countries excluded from the analysis with reasons (Table A-11 in the appendix)

Main Regression Results

The following files contain results from the main multi-regression models (Tables B, C, and Annex C in the paper) in different formats: - **regression_results_formatted.xlsx** – Excel format with formatted tables - **regression_results_pivot.xlsx** – Excel format in pivot table structure - **regression_results_report.html** – HTML format for web viewing

Robustness Checks

The following files contain robustness check results (Table 4 and Annex B in the main paper) in different formats: - **Table_Robustness_Checks.xlsx** – Excel format summary of robustness checks - **RobustnessChecks_Detailed_Comprehensive.csv** – CSV format with detailed comprehensive results

Computational Requirements and Runtime

The author ran the analysis in Google Colab environment with Python 3.

Software Requirements

- **Environment:** Google Colab
 - **Language:** Python 3
-

Folder Structure

The repository contains the following folders:

- **data_raw/** – Contains all raw data used for the paper. Some data was also accessed directly through APIs.
- **data_processed/** – Contains all processed data to be used for analysis (not uploaded to github).
- **results/** – Contains all results and outputs from the main analysis.
- **notebooks/** – Contains all six Jupyter notebooks for this analysis.