Global MNC Panel Dataset (2010–2024): Firm–Country Data on Strategy, Institutions, and Contexts

Integrated Panel Data on Multinational Corporations and Country-Level Environments

DATA USER GUIDE

Version 1.0 – August 2025

Sergey Morgulis-Yakushev (Principal Investigator)
Uppsala University / Mälardalen University

Olivia Kang (Project Lead)
Mälardalen University

Emre Yildiz, Aswo Safari, Sara Melén Hånell (Advisory Group)
Mälardalen University

sergey.morgulis-yakushev@fek.uu.se

https://mjserg.github.io/mnc-panel-dataset/

Mälardalen University, Sweden

1. Overview and Purpose

The *MNC Firm–Country Panel Dataset (2010–2024)* is a multi-level, panel-structured empirical dataset designed to support high-quality academic research in the fields of international business, comparative political economy, strategic management, and institutional theory.

It integrates two analytical levels:

- **Firm-level data** for 560 publicly listed multinational manufacturing firms across 40+ countries, including financial performance, corporate governance, employment, and internationalization metrics;
- **Country-level data** capturing institutional, political, economic, cultural, and environmental indicators for up to 85 countries over a 15-year period.

This dual-level configuration enables researchers to examine how firm strategies and outcomes are shaped by — and interact with — diverse national contexts. It supports both within-firm longitudinal analysis and cross-country institutional comparisons.

Intended Audience

This dataset is intended for:

- **Academic researchers** conducting empirical studies in business, strategy, or political economy;
- **Doctoral students and postgraduates** preparing dissertations or coursework in international business and economics;
- **Instructors and lecturers** developing teaching cases or empirical assignments based on real-world multinational data.

What Makes This Dataset Unique?

Unlike many existing firm datasets, which are either country-specific or limited to cross-sectional snapshots, this dataset offers:

- **Panel depth**: longitudinal data covering 15 consecutive years (2010–2024);
- **Cross-national breadth**: firm coverage spanning all major world regions and a diverse set of institutional environments;
- **Multi-level architecture**: a clean, linkable structure that merges internal firm metrics with external country-level variables;
- **Research-ready design**: standardized formats, reproducible documentation, and transparent metadata.

It provides a rare empirical foundation for studying internationalization, corporate governance, institutional interactions, and global strategy in a way that is scalable, comparative, and methodologically robust.

2. Dataset Contents and Structure

The dataset consists of two interlinked components — **firm-level data** and **country-level data** — harmonized and structured to support cross-sectional, longitudinal, and multi-level empirical analysis. Both components are delivered as .csv files in UTF-8 encoding to ensure compatibility with major statistical software (R, Stata, Python, SPSS).

2.1 Firm-Level Dataset

The firm-level panel includes **560 publicly listed multinational corporations (MNCs)** operating in the manufacturing sector across five continents. Observations span the years **2010 to 2024**, resulting in up to 15 time points per firm.

Key variable domains:

- **Financial performance**: revenue, net income, total assets, market capitalization
- **Profitability ratios**: return on assets (ROA), return on equity (ROE)
- **Employment**: total number of employees
- **Internationalization**: share of foreign sales and foreign assets (as % of total)
- **Corporate governance**: board size, gender diversity, average tenure, CEO identity
- **Company metadata**: firm name, sector classification (NAICS/SIC), year of incorporation, country of registration

Each observation includes a firm identifier and a year, enabling firm-level time-series and panel models.

2.2 Country-Level Dataset

The country-level dataset provides macro- and institutional indicators for up to **85 countries** over the same **15-year period (2010–2024)**. These data offer the national-level context in which MNCs operate.

Key variable domains:

- **Institutional quality**: rule of law, corruption control, regulatory quality
- **Political regime characteristics**: democracy level, civil liberties, governance capacity
- **Cultural dimensions**: Hofstede's six-factor model, values from World Values Survey
- **Macroeconomics**: GDP per capita, GDP growth, inflation, Gini index, FDI
- **Development indicators**: urbanization, education spending, internet penetration

- **Environmental and health metrics**: CO_2 emissions per capita, renewable energy share, environmental performance, health security
- **Innovation and competitiveness**: patent applications, R&D spending, ease of doing business

Each observation is uniquely identified by a country code and year.

2.3 Panel Structure and Integration

Both datasets are structured in **long (tidy) format**, suitable for panel data methods.

Dataset	Key identifiers	Panel Type
Firm-level	firm_id, year	Balanced (firm × year)
Country-level	iso3c, year	Balanced (country × year)

Linkage mechanism:

- **ISO3 country codes** (iso3c) and calendar **year** serve as the linking keys between the two levels.
- This enables **multi-level modeling**, e.g., regressing firm-level outcomes on country-level variables while accounting for hierarchical nesting.

2.4 File Overview (v1.0, August 2025)

File Name	Description
MNC_Panel_Firm_Level_Data_2010_2024.csv	Firm-year panel dataset including 560 publicly listed manufacturing MNCs. Covers financials, governance, employment, and internationalization metrics from 2010 to 2024.
MNC_Panel_Country_Level_Data_2010_2024.csv	Country-year panel with 75+ annual macro-level variables across 80+ countries. Includes institutional, political, cultural, environmental, and economic indicators.

File Name	Description
MNC_Panel_Variable_Dictionary_2010_2024.csv	Unified variable dictionary for all firm-level and country-level variables. Includes names, definitions, categories, units, sources, and time coverage.

All variables follow **snake_case naming** conventions, and **missing values are encoded as NA**.

3. Getting Started: Accessing and Opening the Data

This section provides guidance on how to locate, open, and begin working with the dataset across different statistical environments. It also outlines the structure and use of the accompanying documentation.

3.1 File Access and Structure

The dataset is distributed as part of the project archive, which includes the following key components:

File Name	Description
MNC_Panel_Firm_Level_Data_2010_2024.csv	Firm-year panel dataset including 560 publicly listed manufacturing MNCs. Covers financials, governance, employment, and internationalization metrics from 2010 to 2024.
MNC_Panel_Country_Level_Data_2010_2024.csv	Country-year panel with 75+ annual macro-level variables across 80+ countries. Includes institutional, political, cultural, environmental, and economic indicators.
MNC_Panel_Variable_Dictionary_2010_2024.csv	Unified variable dictionary for all firm-level and country-level variables. Includes names, definitions, categories, units, sources, and time coverage.
MNC_Panel_Technical_Appendix_v1.0_2025.pdf	Full technical documentation outlining sampling, data construction, variable definitions, source integration, and panel design.

File Name	Description
MNC_Panel_User_Guide_2025.pdf	Practical manual for data users. Includes file formats, variable conventions, coding logic, and instructions for importing and analyzing the dataset.
MNC_Panel_Changelog_and_Version_History_v1.0_2025.pdf	Version history and changelog. Documents key development milestones, structural changes, and planned future updates.
MNC_Panel_Final_Report_v1.0_2025.pdf	Executive summary and research overview. Designed for publication, citation, or public dissemination. Includes visuals, highlights, and dataset contributions.

All files are provided in **.csv format with UTF-8 encoding** to ensure compatibility across platforms (Windows, macOS, Linux).

3.2 Opening the Files in Common Statistical Software

In R

```r
# Load necessary libraries
library(readr)
library(dplyr)

# Read the firm-level data
firm_data <- read_csv("Comp_Data.csv")

# Read the country-level data
country_data <- read_csv("Country_Data.csv")
```

In Python

```python
import pandas as pd

# Read the datasets
firm_data = pd.read_csv("Comp_Data.csv", encoding='utf-8')
country_data = pd.read_csv("Country_Data.csv", encoding='utf-8')
```

In Stata

From Stata 15 onward, UTF-8 encoded CSVs can be opened directly:

```
import delimited "Comp_Data.csv", varnames(1) encoding(utf8)
clear
import delimited "Country_Data.csv", varnames(1) encoding(utf8)
clear
```

3.3 Using Variable Dictionaries

Each dataset is accompanied by a detailed variable dictionary in CSV format. These files include:

- variable_name – exact column name in the dataset (in snake_case),
- label – human-readable label or description,
- unit – unit of measurement (e.g., %, millions USD),
- source – data origin (e.g., Eikon, World Bank),
- notes – definitions, coverage notes, special codes (if applicable).

We recommend consulting the dictionaries before beginning any analysis to ensure correct interpretation of variables.

3.4 Initial Validation and Summary Checks

Before launching into modeling or visualization, users are advised to perform basic data checks:

In R

```
# View structure and summary
glimpse(firm_data)
summary(firm_data)

# Check missingness
colSums(is.na(firm_data))
```

In Python

```
# Quick overview
firm_data.info()
firm_data.describe()

# Missing values
firm_data.isnull().sum()
```

In Stata

```
describe
summarize
misstable summarize
```

These checks help identify:

- columns with large amounts of missing data,
- unusual value distributions or outliers,
- potential mismatches in data types (e.g., year as string instead of integer).

4. Variable Naming and Metadata

This section explains the logic behind the naming conventions used in the dataset, how to interpret variable names, and how to use the accompanying data dictionaries to navigate the structure of the dataset effectively.

4.1 Naming Convention

All variables follow a **standardized snake_case** naming format to ensure consistency and readability across statistical software (R, Python, Stata).

The naming logic encodes four elements:

```
[content]_[year]_[unit]_[source]
```

Component	Description	Example
content	What the variable measures	board_size, revenue, gdp_per_capita
year	Year of measurement (if time-specific)	y2023, y2015, y2020
unit	Unit of measurement	percent, millions_usd, count
source	Original data source or collection platform	eikon, forbes, wdi, vdem

Examples:

- board_size_y2024_count_eikon

→ Board size in 2024, as a count, sourced from Eikon

- revenue_y2020_millions_usd_eikon

→ Firm revenue in 2020 in millions of USD, from Eikon

4.2 Variable Types

Variables generally fall into the following categories:

Category	Examples
Financial metrics	net_income, assets, market_value
Governance characteristics	board_size, avg_tenure, gender_ratio
Macroeconomic indicators	gdp_per_capita, fdi_inflow, tax_revenue
Institutional metrics	rule_of_law, voice_accountability
Cultural values	pdi, idv, uai (Hofstede dimensions)
Environmental indicators	co2_emissions, renewable_share

4.3 Using the Variable Dictionaries

The dataset is accompanied by a single integrated variable dictionary:

File Name	Purpose
MNC_Panel_Variable_Dictionary_2010_2024.csv	Full documentation of all firm-level and country-level variables, including metadata, units, sources, and definitions.

This dictionary provides a comprehensive reference for users working with the data. It includes the following columns:

Column	Description
variable_name	Exact variable name as used in the dataset (in snake_case format)
label	Human-readable description of the variable
unit	Unit of measurement (e.g., %, index score, USD millions, count)
source	Data origin (e.g., Refinitiv Eikon, V-Dem, World Bank, Forbes, WVS)
notes	Extended definitions, special cases, coding remarks, and caveats if applicable

This file ensures transparency and reproducibility for users seeking to understand, subset, or model the dataset's structure across both levels of analysis.

4.4 Where to Look for Metadata

In addition to the dictionaries, each variable is described in the **Technical Appendix** and documented in relevant sections of the **Final Report**.

For new users, we recommend:

- Searching variable names by keyword (ctrl+F) in the dictionaries.
- Referring to the "label" and "notes" fields for clarification.
- Using summary statistics to explore distributions and coverage.

5. Dealing with Missing Data

This section provides guidance on how missing data are represented in the dataset, what level of missingness users can expect, and recommended strategies for addressing missing values in empirical analysis.

5.1 Representation of Missing Data

All missing values in both the firm-level and country-level datasets are encoded as:

```
NA
```

This is consistent with conventions used in R and Python and is recognized by major statistical packages (e.g., . in Stata, np.nan in pandas, NA in R). No missing values are coded as zero, blank strings, or custom codes (e.g., -99).

5.2 Levels of Missingness

Firm-Level Dataset:

- **Most financial variables** (e.g., revenue, assets, net income) have near-complete coverage across the 2010–2024 panel.
- **Board-related governance variables** (e.g., board_member_compensation, avg_tenure) exhibit moderate missingness, particularly for firms in jurisdictions with limited disclosure requirements.
- **Internationalization metrics** are broadly available but not always reported uniformly for all years.

Country-Level Dataset:

- **Macroeconomic indicators** (e.g., GDP, FDI, tax revenue) have strong coverage across countries and years.

- **Institutional and political indicators** (e.g., V-Dem, Freedom House) show gaps, especially in early years or low-income countries.
- **Cultural variables** (e.g., Hofstede dimensions, WVS indicators) are time-invariant or sparsely updated and therefore not available annually.

5.3 Suggested Analytical Strategies

There is no one-size-fits-all solution for handling missing data. The appropriate method depends on your research question, design, and model assumptions. Common approaches include:

Strategy	Description
Listwise Deletion	Drop observations with any missing values (default in many regressions). Suitable for large-N models with minimal missingness.
Mean/Median Imputation	For descriptive purposes only. Not recommended for inferential modeling.
Multiple Imputation (MI)	Recommended when missingness is non-random but data are MAR (missing at random). Packages: mice in R, mi in Stata.
Robustness Checks	Run alternative models with and without variables prone to missingness. Document differences in findings.

5.4 Variables with Systematic Missingness

Below is a non-exhaustive list of variables with known patterns of missing data:

Variable Name	Description	Common Reason
board_member_compensation	Total board compensation (USD)	Missing in countries with opaque disclosure
avg_board_tenure	Average tenure of board members	Often not reported in Asia/EMEA
employees_y2010_count_eikon	Number of employees (2010)	Early years often unreported
religiosity_wvs	Self-reported religiosity (WVS)	Available only in certain survey waves
vdem_deliberative_index	Deliberative democracy score	Gaps in 2010–2012 for some countries

5.5 Summary Recommendation

We strongly encourage users to:

- Examine missingness patterns **early in the workflow** using exploratory summaries (summary(), is.na(), naniar in R, or missingno in Python).
- Justify any imputation techniques and document their effect on findings.
- Consult the **Technical Appendix** for notes on data cleaning and known caveats.

Proper handling of missing data is essential for generating robust, interpretable, and reproducible results.

6. Use Cases and Sample Analyses

This section presents illustrative research applications of the dataset, including example hypotheses, modeling strategies, and recommended tools for implementation. It is intended to guide users—especially researchers, doctoral students, and instructors—in effectively leveraging the dataset for academic work in international business, comparative politics, and strategic management.

6.1 Example Research Questions and Models

The dataset supports a wide range of empirical investigations that examine how institutional environments shape firm behavior, performance, and internationalization strategies.

Example 1: Democracy and Firm Performance

Research Question:
Does democratic governance in a firm's home country affect its financial performance?

Model Specification:

$$ROA_{it} = \beta_0 + \beta_1 * democracy\_index_{ct} + \alpha_i + \gamma_t + \varepsilon_{it}$$

- **Dependent Variable:** ROA (return on assets)

- **Independent Variable:** Democracy Index (e.g., V-Dem, Freedom House)
- **Controls:** Firm and year fixed effects
- **Interpretation:** Estimates the average effect of institutional democracy on firm profitability

Example 2: Cultural Distance and Internationalization

Research Question:
Are firms from individualistic societies more likely to engage in international expansion?

Model Specification:

```
intl_sales_share_it = β0 + β1 * individualism_c + β2 * gdp_pc_ct
+ α_i + γ_t + ε_it
```

- **Dependent Variable:** Share of international sales (%)
- **Independent Variable:** Cultural dimensions (e.g., individualism, uncertainty avoidance)
- **Controls:** GDP per capita, firm fixed effects
- **Use Case:** Understanding how cultural traits shape cross-border strategies

Example 3: Political Instability and Board Structure

Research Question:
Do firms adapt their board structure when facing political instability in their institutional environment?

Model Specification:

```
board_size_it = β0 + β1 * political_instability_ct + β2 *
firm_size_it + α_i + γ_t + ε_it
```

- **Dependent Variable:** Board size
- **Independent Variable:** Political regime volatility (e.g., V-Dem instability index)

- **Controls:** Firm size, firm/year fixed effects
- **Use Case:** Governance adaptation under risk and uncertainty

6.2 Analytical Techniques

The panel and multilevel structure of the dataset allows for a range of modeling strategies:

Methodology	Application
Fixed Effects / Random Effects	Control for unobserved firm or country heterogeneity
Hierarchical Linear Models (HLM)	Model nested structures (e.g., firms within countries)
Difference-in-Differences (DiD)	Estimate the impact of policy shocks or regime changes
Interaction Models	Capture joint effects of firm- and country-level predictors (e.g., ROA ~ intl_assets × regulatory_quality)

6.3 Recommended Tools and Packages

Depending on your preferred statistical environment, we suggest the following packages:

R

Package	Use Case
plm	Panel data with fixed/random effects
lme4	Multilevel and hierarchical models
tidyverse	Data cleaning, transformation, plotting
mice	Multiple imputation for missing data

Python

Package	Use Case
pandas	Data manipulation and reshaping

Package	Use Case
linearmodels	Panel regressions (e.g., fixed effects)
statsmodels	OLS, logistic, DiD, multivariate models
pyjanitor	Data cleaning and NA-handling

Stata

Command	Use Case
xtreg	Fixed/random effects panel models
mixed	Multilevel/hierarchical models
mi	Multiple imputation routines

6.4 Best Practices

- Always **inspect missingness** and **temporal coverage** before estimating models.
- Use **robust standard errors clustered by firm or country**, depending on model structure.
- Where appropriate, **standardize variables** (e.g., GDP, cultural scores) for interpretability.
- Clearly **document modeling choices**, especially in multilevel designs or DiD frameworks.

7. Tips for Responsible Use

This section provides guidance for using the dataset ethically, accurately, and effectively. It includes recommendations on citation practices, technical integration, and interpretive caution, as well as information on where to seek support.

7.1 Citation and Attribution

Proper citation of the dataset ensures academic transparency and gives credit to the contributors. When using the dataset in publications, presentations, or teaching materials, please cite as follows:

> **Morgulis-Yakushev, S., Kang, O., Yildiz, E., Safari, A., & Melén Hånell, S.** (2025). *Global MNC Panel Dataset (2010–2024): Firm–Country Data on Strategy, Institutions, and Contexts.* Mälardalen University & Uppsala University. Version 1.0.
>
> You may also include a reference to the accompanying technical report:
>
> **Morgulis-Yakushev, S., Kang, O., Yildiz, E., Safari, A., & Melén Hånell, S.** (2025). *Technical Appendix: Data Construction and Sources for the Global MNC Panel Dataset*. Mälardalen University.

7.2 Merging and Data Integration

When merging the firm-level and country-level datasets:

- Use **iso3c** and **year** as the composite key.
- Be aware of **potential discrepancies in country naming conventions** across external datasets.
- Always **verify merge success** using summary statistics (e.g., n() by country-year).
- Consider keeping the original **merge indicator** (e.g., merge_status) to track incomplete joins.

For additional merge integrity:

- Ensure that variable names do not conflict across datasets.
- Use unique firm identifiers if extending the data with external sources.

7.3 Interpretation Guidelines

Some variables—especially those related to institutions, culture, and governance—require nuanced interpretation:

- **Indices are not directly observable variables**. They are often based on surveys, expert assessments, or composite models. Interpret them **as proxies**, not exact measurements.
- **Cultural dimensions** (e.g., Hofstede scores) represent average national tendencies. They should not be interpreted at the individual level.
- **Aggregated country-level indicators** may not capture subnational variation or contextual heterogeneity within large countries.
- When using **interaction terms** (e.g., firm × country), carefully consider multicollinearity and centering strategies.

7.4 Contact and Support

For questions, feedback, or requests for collaboration, please contact the project team:

Sergey Morgulis-Yakushev
Principal Investigator
sergey.morgulis-yakushev@fek.uu.se

Olivia Kang
Project Lead
olivia.kang@mdu.se

We welcome academic collaboration, reproducibility studies, and suggestions for future extensions. Please let us know if you publish work using the dataset so we can include it in our impact records.

8. Licensing and Use Conditions

This section outlines the licensing terms, permitted uses, and responsibilities associated with accessing and analyzing the dataset. It is essential that all users adhere to these conditions to ensure ethical, secure, and academically appropriate use of the data.

8.1 Academic-Only Usage

This dataset is distributed strictly for **non-commercial, academic purposes**. Authorized uses include:

- Academic research projects
- Doctoral dissertations and master's theses
- Peer-reviewed journal publications
- Conference presentations
- Graduate-level teaching and training

Any commercial, consultancy, or proprietary application of the dataset is **explicitly prohibited** without prior written consent from the project team and host institution.

8.2 Restrictions on Redistribution

Users are **not permitted to redistribute raw datasets** (firm-level or country-level) outside their authorized academic setting. This includes:

- Posting files on public repositories or websites
- Sharing the full dataset with third parties not covered under institutional license or agreement
- Republishing or reformatting the dataset for distribution

Instead, interested collaborators should be directed to the project contact listed below to request access via proper channels.

8.3 Institutional Access and Licensing

All firm-level data derived from **Refinitiv Eikon** and **Orbis Global (Bureau van Dijk)** were accessed through **Mälardalen University's institutional subscriptions**. Users

from other institutions must ensure they have legitimate academic access to these underlying sources for replication or extension purposes.

Country-level data were sourced from publicly accessible and academically validated databases (e.g., World Bank, V-Dem, EPI). Licensing conditions for these sources are subject to their respective terms of use.

8.4 Contact for Access and Inquiries

To request access, clarify permissions, or inquire about licensing, please contact:

Olivia Kang
Project Lead
olivia.kang@mdu.se

All users are encouraged to confirm their intended use with the project team, especially when planning publications or data sharing beyond their immediate academic environment.