

README for Reproducibility Package for “Neither Too Little, nor Too Late: Restructuring as a Reflex for Outcomes.”

- This README file is designed to be included in reproducibility packages to provide directions for replicating the results in the research paper, “**Neither Too Little, nor Too Late: Restructuring as a Reflex for Outcomes.**”

Contents

1. [Overview](#)
2. [Data Availability](#)
3. [Instructions for Replicators](#)
4. [List of Exhibits](#)
5. [Requirements](#)

Overview

1. Overview of the Replication Package

This replication package includes raw data files, scripts for cleaning and merging the data, and programs used to reproduce the main results, figures, and appendix materials. Replicators should begin by reviewing the folder structure, then proceed by executing the master script (**main.R** and **main.do**), which calls all subsidiary scripts in the correct order.

2. Folder Structure and Contents

- /code/ – R scripts and Stata do-files for cleaning, analysis, and plotting
- /data/raw/ – source data files
- /data/final/ – processed .csv / .dta files created by the cleaning scripts, plus a few .csv files that required no cleaning
- /data/txt/ – auxiliary restructuring papers scraped from the World Bank *Documents & Reports* portal
- /data/rap/ – auxiliary *Results and Performance of the World Bank Group* reports scraped from the World Bank *Documents & Reports* portal
- /figure/ – paper figures
- /table/ – paper tables
- codebook.xlsx – codebook for the main analysis dataset (twfe.dta)

3. Steps to Run the Replication

- Set the main project path in the master scripts (line 19 in main.R; and line 2 in main.do).

- Install required packages (see the Software Requirements section).
- Run the cleaning scripts to produce final datasets.
- Run the main analysis script (first **main.R**; then **main.do**) to produce the text-analysis outputs as well as the TWFE and PanelMatch estimates.
- Final outputs save to /figure/ and /table/.

Data Sources

- **Filename 1:** /data/raw/v_dim_portfolio_project_isr.csv
- **Source:** Internal World Bank Databricks environment (restricted access)
- **Databricks path:** prd_corpdata/dm_operation_gold/v_dim_portfolio_project_isr
- **Description:** This dataset provides a comprehensive overview of project evaluations, capturing key performance ratings and historical data to support strategic analysis and decision-making.
- **Access date:** August 18, 2025

- **Filename 2:** /data/raw/v_dim_project.csv
- **Source:** Internal World Bank Databricks environment (restricted access)
- **Databricks path:** prd_corpdata/dm_operation_gold/v_dim_project
- **Description:** This dataset provides a project dimension view.
- **Access date:** August 18, 2025

- **Filename 3:** /data/raw/v_dim_restructuring.csv
- **Source:** Internal World Bank Databricks environment (restricted access)
- **Databricks path:** prd_corpdata/dm_operation_gold/v_dim_restructuring
- **Description:** This dataset captures the sequential and fiscal details of project restructuring, including approval authority and various change indicators.
- **Access date:** August 18, 2025

- **Filename 4:** /data/final/rap_rest_freq.csv
- **Source:** Authors' construction (public access)
- **URL:** <https://documents.worldbank.org/en/publication/documents-reports>
- **Description:** This dataset provides counts of restructuring-related terms in the *Results and Performance of the World Bank Group* reports.
- **Construction date:** September 22, 2025

- **Filename 5:** /data/final/txt_meta.csv
 - **Source:** Authors' construction (public access)
 - **Method:** World Bank API
 - **Description:** This dataset provides metadata for each restructuring paper.
 - **Construction date:** September 19, 2025
-
- **Filename 6:** /data/final/twfe.dta
 - **Source:** Authors' construction (restricted access)
 - **Source:** Internal World Bank Databricks environment
 - **Description:** This dataset is used for the main analysis. It is constructed based on files 1, 2, and 3.
-
- **Filename 7:** /data/final/proj_active.dta
 - **Source:** Authors' construction (restricted access)
 - **Source:** Internal Standard Reports environment
 - **Description:** This dataset denotes active portfolios for each fiscal year.
 - **Construction date:** October 29, 2025

Data Availability

This study draws on two categories of data: (1) internal World Bank operational datasets that cannot be made publicly available, and (2) publicly accessible project documents from the World Bank's *Documents & Reports* portal.

Restricted Internal Data.

Three operational datasets used in this study—**v_dim_portfolio_project_isr.csv**, **v_dim_project.csv**, and **v_dim_restructuring.csv** (located in the project's raw/ folder)—are accessible only to users with internal World Bank Databricks permissions. Because of institutional confidentiality rules governing operational systems, these files cannot be shared directly.

The final analysis dataset (**twfe.dta**) is a merged and cleaned file constructed from the three internal csv files listed above. Since it is derived entirely from restricted data, this dataset also cannot be publicly released.

World Bank staff can request Databricks access via eServices. Select "ITSDA Platforms Service Request," then, under Service Details, choose the "Platforms" tab and select "Databricks Lakehouse." Link:

<http://databricks.worldbank.org>

proj_active.csv is constructed from the Standard Reports system (<https://standardreports.worldbank.org>) and cannot be shared.

To construct this dataset, World Bank staff must download two standard reports (Path: Standard Reports → Operations → Portfolio → Reports):

1. P5.6 – Details

Variables used: Project ID, Approval FY, Bank Approval Date, Revised Closing Date, and Parent Project ID (projects listed as children under a parent were excluded).

2. P5.8 – Closed Project Detailed Evaluation Ratings

Variables used: Project ID, Approval FY, Bank Approval Date, Project Closing Date, and Project Closing FY.

For both reports, any missing dates were manually completed using information from the Operations Portal.

Replicators should then append the two reports to form a unified dataset covering both active and closed projects. For each fiscal year X, a project is coded as active if:

Approval FY $\leq$ X AND Closing FY $\geq$ X.

This rule is applied across all fiscal years to generate the annual activity indicator.

The dataset was constructed as of 29 October, and Standard Reports are updated daily. We'd note that some discrepancies may arise if the data is regenerated later – for example, projects active on 29 October but later closed or cancelled may no longer appear.

Publicly Available Data/Documents.

1. All project restructuring documents and related operational files referenced in the study are publicly available through the World Bank *Documents & Reports* portal:

<https://documents.worldbank.org/en/publication/documents-reports>

These documents can be located by searching using the keyword “**restructuring.**”

The metadata for each document was constructed using the World Bank API. We have attached the R code (**code/txt_meta.R**), which replicators can use to reproduce the process. We note that, over time, the returned rows may differ as the underlying data are updated.

2. *Results and Performance of the World Bank Group* referenced in the study are publicly available through the World Bank *Documents & Reports* portal:

<https://documents.worldbank.org/en/publication/documents-reports>

These documents can be located by searching using the report title (**data/rap/**).

The authors then use the PDF tool to count occurrences of the words “**restructure**” and “**restructuring**,” as well as the total word count.

Statement about Rights

- We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Instructions for Replicators

New users should follow these steps to run the package successfully: Users must first have access to all data files if they are not included in the reproducibility package. They should go to the mentioned links, download the listed files, and place them in the data folder.

Update the following files with your directory paths

- `main.R/main.do`
- Ensure all required software and dependencies are installed as listed in the [Requirements](#) section.
- Run the `main.R/main.do` file.

List of Exhibits

The provided code reproduces:

- All tables and figures in the paper

Requirements

Computational Requirements

- **macOS Tahoe 26.1**
- **Chip: Apple M2 Pro**
- **Memory: 16 GB**

Software Requirements

- **Stata version 18**
 - `reghdfe`
 - `estout`
- **R 4.4.3**
 - `haven` 2.5.4
 - `tidyverse` 2.0.0
 - `PanelMatch` 3.1.1
 - `hrbrthemes` 0.8.7
 - `tidytext` 0.4.2
 - `textstem` 0.1.4
 - `widyr` 0.1.5
 - `igraph` 2.1.4
 - `ggraph` 2.2.1
 - `fixest` 0.12.1
 - `stargazer` 5.2.3

Memory and Runtime and Storage Requirements

The replication package is approximately 1.4 GB. Running all analyses end-to-end takes roughly 40 minutes on a MacBook Pro with an M2 Pro chip (Tahoe 26) and 16 GB of RAM.