

Reproducibility Package

"Returns to Private Lending in Emerging Markets and Developing Economies"

World Bank Working Paper, July 2026

Authors: Shawn Cole, Cesaire Meh, Fanele Mashwama, Florian Mölders

Contents

- 1. Overview
- 2. Data Availability and Provenance Statement
- 3. Dataset Documentation
- 4. Computational Requirements
- 5. Package Structure
- 6. Instructions to Replicators
- 7. List of Tables, Figures, and Programs
- 8. Key Methodological Notes
- 9. License
- 10. Contact

1. Overview

This reproducibility package contains all code and publicly shareable data necessary to replicate the tables and figures in the working paper. The primary dataset — IFC loan cash flows — is **confidential** and subject to access restrictions. Replication of results therefore follows the **virtual verification** pathway described in Section 6.

2. Data Availability and Provenance Statement

2.1 Summary of availability

This paper uses two categories of data:

Category	Availability	Included in package
IFC Loan Cash Flow Database	Confidential — IFC internal	No
IFC Loan Terms Database	Confidential — IFC internal	No
IFC Default Database	Confidential — IFC internal	No
CEMBI / EMBI benchmark indices (JPMorgan)	Licensed — Bloomberg/Datastream subscription required; restricted redistribution	Yes, for virtual verification only (Data/Public/)
US Treasury total return series	Publicly available	Yes (Data/Public/)
IMF GDP growth data	Publicly available	Yes (Data/Public/)

Category	Availability	Included in package
World Bank classification data	Publicly available	Yes (Data/Public/)
ISO country code crosswalk	Compiled reference	Yes (Data/Public/)
IFC FCS / IDA classification lists	IFC-published	Yes (Data/Public/)
WDI macro indicators	Publicly available	Yes (Data/Public/)
World Bank IDS disbursement data	Publicly available (API)	Yes, cached (Data/Public/)

Because the primary datasets are confidential, full replication follows the **virtual verification** pathway (see Section 6).

2.2 Confidential data — access

The primary input dataset is the **IFC Loan Cash Flow Database** ([Loan cash flows anonymized.dta](#)), which records project-level cash flows for IFC's direct lending portfolio. This dataset contains anonymized but commercially sensitive information about individual borrowers and is held by the **International Finance Corporation (IFC)**. Access is restricted to authorized IFC staff.

A second confidential input is the **IFC Loan Terms Database** ([Loan terms anonymized.dta](#)), which records project approval dates and contractual loan tenures. A third confidential input is the **IFC Default Database** ([IFC default data anonymized.dta](#)), which records institution-level default events in IFC's portfolio.

Researchers seeking access for replication purposes should contact the corresponding author, Florian Mölders (fmoelders@ifc.org), who will coordinate the request internally with IFC. Access is granted at IFC's discretion and subject to a non-disclosure agreement. The World Bank's virtual verification pathway (Section 6, Option A) requires no data transfer to the reviewer.

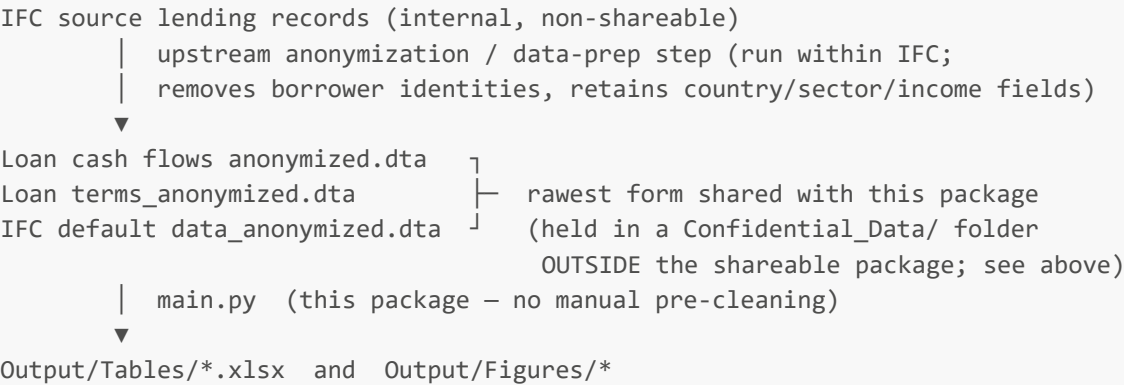
2.3 Data storage and stewardship

The three confidential IFC input datasets are held as anonymized Stata ([.dta](#)) files on **World Bank Group–managed OneDrive** (the authors' institutional Microsoft 365 tenant), which is access-controlled and backed up by WBG IT — not on a personal or unmanaged device. The corresponding author retains these files in his WBG OneDrive as the working copy used to produce the results in this package.

To allow the reproducibility package to be shared without exposing confidential data, the three [.dta](#) files are stored in a [Confidential_Data/](#) folder that sits **outside** (as a sibling of) the shareable package folder. The shared package's own [Data/Confidential/](#) directory is empty apart from a [README_CONFIDENTIAL.md](#) placeholder. [main.py](#) locates the confidential folder via, in order: the [IFC_CONFIDENTIAL_DIR](#) environment variable; a [Confidential_Data/](#) folder next to the package; or the in-package [Data/Confidential/](#) folder as a legacy fallback.

2.4 Data pipeline and rawest available form

The **rawest form** of the confidential data that the authors can share is the anonymized [.dta](#) files. They were produced upstream, within IFC, by an internal anonymization / data-preparation step applied to IFC's source lending records. The underlying non-anonymized source records remain within IFC's internal systems and cannot be redistributed. [main.py](#) runs end to end starting from the anonymized [.dta](#) files; no further manual cleaning is required or performed inside this package.



The upstream anonymization/data-preparation script is maintained internally at IFC and is **not distributed** with this package, because it operates on non-anonymized source records containing confidential borrower identifiers. Its role is limited to (i) anonymizing borrower/institution identifiers and (ii) exporting the relevant fields to Stata format; it performs none of the analytical sample construction, which is done transparently inside this package's scripts (deduplication, Filter A', blended-finance/non-USD/ short-tenor/equity exclusions — see Section 8). Replicators with authorized access to the anonymized `.dta` files can therefore reproduce every result from the rawest form the authors are permitted to share.

2.5 Public data

All non-confidential input files are included in the `Data/Public/` folder. Per-dataset sources, access dates, licenses, and citations are documented in Section 3.

3. Dataset Documentation

This section documents every dataset used in the analysis, following the [Social Science Data Editors' citation standards](#) and World Bank reproducibility guidelines.

3.1 IFC Loan Cash Flow Database (*Confidential*)

Attribute	Value
Filename	Loan cash flows anonymized.dta
Format	Stata (.dta), version 14+
Location	Held outside the shareable package (see Section 2.3); not included
Source	International Finance Corporation (IFC), World Bank Group
Description	Project-level cash flow records for IFC's lending portfolio, including disbursements, repayments, fees, and outstanding balances. Each row is a transaction for a specific loan tranche. Borrower identities are anonymized; country, sector, and income-group information is retained.
Time coverage	1992–2024 (analysis restricted to 2002–2024)
Unit of observation	Transaction (loan tranche × date × transaction type)
Access instructions	The dataset is confidential and cannot be redistributed. Researchers seeking access for replication should contact the corresponding author, Florian Mölders (fmoelders@ifc.org), who will

Attribute	Value
	coordinate the request internally with IFC. Access is granted at IFC's discretion, subject to a non-disclosure agreement. The World Bank's virtual verification pathway (Section 6) requires no data transfer to the reviewer.
Access date (authors)	Anonymized file received by the authors from an IFC colleague on May 13, 2025.
License / restrictions	Confidential; not for public release or redistribution.
Rights statement	Data were obtained by the authors in their capacity as IFC/World Bank Group staff for the purpose of this research.

Key variables:

Variable	Description
PROJECT_ID	Anonymized project identifier
TRANCHE_NBR	Loan tranche number within project
loan_id	Constructed as PROJECT_ID + "_" + TRANCHE_NBR
VALUE_DATE	Transaction value date (calendar)
TRANSACTION_TYPE_ID	Transaction type (e.g., Disbursement, Principal Repayment, Interest, Fee)
USD_AMT_ADJ	USD-adjusted transaction amount (positive = outflow from IFC, negative = inflow)
OBLIGATION_NBR	Sub-obligation identifier
PARTICIPANT_NBR	Participant identifier (for syndications)
iso3	ISO 3166-1 alpha-3 country code
COUNTRY_NME	Country name
BF	Binary: 1 if blended-finance project
SECTOR	IFC sector classification
INDUSTRY_GROUP	IFC industry group

Citation: International Finance Corporation (IFC). (2025). *IFC Loan Cash Flow Database* [Confidential dataset]. World Bank Group, Washington, D.C. Data accessed by authors in institutional capacity, 2024–2026.

3.2 IFC Loan Terms Database (*Confidential*)

Attribute	Value
Filename	Loan_terms_anonymized.dta
Format	Stata (.dta)
Location	Held outside the shareable package (see Section 2.3); not included
Source	International Finance Corporation (IFC), World Bank Group
Description	Loan-level terms records for IFC's lending portfolio, including project approval date and contractual loan tenure (years). Used to assign approval year and contracted tenor in Table 9 and

Attribute	Value
	supporting regressions.
Time coverage	Loans approved through 2024
Unit of observation	Loan tranche
Access instructions	The dataset is confidential and cannot be redistributed. Researchers seeking access for replication should contact the corresponding author, Florian Mölders (fmoelders@ifc.org), who will coordinate the request internally with IFC. Access is granted at IFC's discretion, subject to a non-disclosure agreement.
Access date (authors)	Anonymized file received by the authors from an IFC colleague on May 13, 2025.
License / restrictions	Confidential; not for public release or redistribution.
Rights statement	Data were obtained by the authors in their capacity as IFC/World Bank Group staff for the purpose of this research.

Key variables:

Variable	Description
PROJECT_ID	Anonymized project identifier
TrancheNumber	Loan tranche number
loan_id	Constructed as PROJECT_ID + "_" + TrancheNumber
ProjectApprovalDate	Date of project approval by IFC Board
LoanTenureyear	Contractual loan tenure in years

Citation: International Finance Corporation (IFC). (2025). *IFC Loan Terms Database* [Confidential dataset]. World Bank Group, Washington, D.C. Data accessed by authors in institutional capacity, 2024–2026.

3.3 IFC Default Database (*Confidential*)

Attribute	Value
Filename	IFC default data_anonymized.dta
Format	Stata (.dta)
Location	Held outside the shareable package (see Section 2.3); not included
Source	International Finance Corporation (IFC), World Bank Group
Description	Institution-year panel recording default events in IFC's portfolio. Contains default date, recovery information, and institution-level identifiers (anonymized). Used for Figures 2–3 and supporting regression tables.
Time coverage	1991–2024 (analysis restricted to 2002–2024)

Attribute	Value
Unit of observation	Institution × year
Access instructions	The dataset is confidential and cannot be redistributed. Researchers seeking access for replication should contact the corresponding author, Florian Mölders (fmoelders@ifc.org), who will coordinate the request internally with IFC. Access is granted at IFC's discretion, subject to a non-disclosure agreement.
Access date (authors)	Anonymized file received by the authors from an IFC colleague on May 13, 2025.
License / restrictions	Confidential; not for public release or redistribution.
Rights statement	Data were obtained by the authors in their capacity as IFC/World Bank Group staff for the purpose of this research.

Key variables:

Variable	Description
INSTITUTION_ID	Anonymized institution identifier
YEAR	Calendar year
DEFAULT_DATE	Date of default event (missing if no default)
default_dummy	Binary: 1 if institution defaulted in that year
iso3	ISO 3166-1 alpha-3 country code

Citation: International Finance Corporation (IFC). (2025). *IFC Default Database* [Confidential dataset]. World Bank Group, Washington, D.C. Data accessed by authors in institutional capacity, 2024–2026.

3.4 JPMorgan CEMBI Broad Diversified (*Licensed — restricted redistribution*)

Filename: `jpm_cembi_broad_diversified_total_return.xlsx` — overall and country sub-indices **Format:** Excel / CSV **Location in package:** `Data/Public/` **License / redistribution restrictions:** The CEMBI index data are proprietary to JPMorgan Chase & Co. and are distributed under a data license agreement. The files included in this package were obtained by the authors via a Bloomberg terminal subscription. **Independent replicators must obtain their own access** through Bloomberg, Refinitiv Datastream, or directly from JPMorgan. These files are included here solely to support the World Bank's virtual verification process; they may not be redistributed further. **Date accessed:** February 2026 **Coverage:** Monthly total return index levels; coverage begins January 2002 for the broad index. **Key column used:** `EM Debt Indices | CEMBI BROAD DIVERSIFIED | Cum Tot Ret Idx`

Citation: JPMorgan Chase & Co. (2026). *CEMBI Broad Diversified Total Return Index* [Data file]. Retrieved via Bloomberg terminal, February 2026.

3.5 JPMorgan EMBI Broad Diversified (*Licensed — restricted redistribution*)

Filename: `jpm_embi_broad_diversified_total_return.xlsx` **Format:** Excel / CSV **Location in package:** `Data/Public/` **License / redistribution restrictions:** The EMBI index data are proprietary to JPMorgan Chase & Co. and are distributed under a data license agreement. The files included in this package were obtained by the authors via a Bloomberg terminal subscription. **Independent replicators must obtain their own access** through Bloomberg, Refinitiv Datastream, or directly from JPMorgan. These files are included here solely to support the World Bank's virtual

verification process; they may not be redistributed further. **Date accessed:** February 2026 **Coverage:** Monthly total return index levels; coverage begins January 1994.

Citation: JPMorgan Chase & Co. (2026). *EMBI Broad Diversified Total Return Index* [Data file]. Retrieved via Bloomberg terminal, February 2026.

3.6 US Treasury 7-10 Year Total Return Series

Filename: `us_treasury_7_10yr_total_return_monthly.csv` **Format:** CSV **Location in package:** `Data/Public/` **Source:** Datastream (Thomson Reuters/LSEG) **Date accessed:** January 2026 **Coverage:** Monthly total return index for 7-10 year US Treasuries. **Usage:** US Treasury benchmark in Tables 2 and 3.

Citation: LSEG Datastream. (2026). *US Treasury 7-10 Year Total Return Index* [Data file]. Retrieved via Datastream, January 2026.

3.7 IMF Real GDP Growth Data

Filename: `imf_real_gdp_growth_annual.csv` **Format:** CSV **Location in package:** `Data/Public/` **Source:** IMF World Economic Outlook (WEO) database **URL:** <https://www.imf.org/en/Publications/WEO> **Date accessed:** January 2026 **Coverage:** Annual real GDP growth rates for all IMF member countries. **Usage:** Used to construct GDP downturn/crisis indicators in Table 9 and Figures 2–3.

Citation: International Monetary Fund. (2026). *World Economic Outlook Database* [Data file]. <https://www.imf.org/en/Publications/WEO>. Retrieved January 2026.

3.8 World Bank Country Analytical History (Income Classifications)

Filename: `wb_country_income_classification_historical.xlsx` **Format:** Excel **Location in package:** `Data/Public/` **Source:** World Bank Open Data **URL:** <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519> **Date accessed:** February 2026 **Coverage:** World Bank fiscal-year income group classifications (Low, Lower-Middle, Upper-Middle, High Income) from FY1988 onwards, for all World Bank member countries. **Usage:** Income group assignment for all main tables and figures. Analysis uses fixed 2005 classification (equivalent to FY2006 World Bank designation).

Citation: World Bank. (2026). *World Bank Country and Lending Groups — Historical Classification* [Data file]. World Bank Open Data. <https://datahelpdesk.worldbank.org>. Retrieved February 2026.

3.9 ISO Country Master File

Filename: `iso_country_codes_master.xlsx` **Format:** Excel **Location in package:** `Data/Public/` **Source:** Compiled from ISO 3166-1 standard and World Bank / IFC country name conventions. **Coverage:** Crosswalk of country name variants (column `Name Variants`) to ISO 3166-1 alpha-3 codes (column `iso`). **Usage:** Resolves country names to ISO3 codes in the IFC default database, which carries no native ISO3 field. Used by `Table_4_5_6_FCS_Analysis.py`, `Table_7_IDA_Analysis.py`, `Figure_2_3_Default_Rates_Downturns.py`, and `Table_8_Downturn_PME_Regressions.py`. The scripts also carry an internal `manual_fixes` fallback for full-form country spellings, so the default-rate panel is not silently truncated if a name variant is absent from this file.

Citation: International Organization for Standardization. *ISO 3166-1 Country Codes* [Reference standard]. Compiled with World Bank / IFC name variants.

3.10 IFC IDA/FCS Country List 2006–2025

Filename: `ifc_fcs_ida_country_classification_2006_2025.xlsx` **Format:** Excel **Location in package:** `Data/Public/` **Source:** International Finance Corporation (IFC), published annually **Coverage:** Country-year panel of

IDA eligibility status and FCS classification, 2006–2025. **Usage:** FCS and IDA exposure indicators (Tables 5–8).

Citation: International Finance Corporation (IFC). (2025). *IFC IDA/FCS Country Classification List 2006–2025* [Data file]. World Bank Group, Washington, D.C.

3.11 World Bank IDA Eligibility

Filename: `wb_ida_eligibility_country_year.xlsx` **Format:** Excel **Location in package:** `Data/Public/` **Source:** World Bank **Coverage:** Country-year panel of IDA eligibility status. **Usage:** IDA eligibility classification in Table 8.

Citation: World Bank. (2026). *IDA Eligibility* [Data file]. World Bank Open Data.
<https://www.worldbank.org/en/topic/debt/brief/ida-eligibility>. Retrieved February 2026.

3.12 World Bank IDS API (fetched at run time)

Series: `DT.DIS.DPNG.CD` — EMDE private non-guaranteed external debt disbursements (USD) **Source:** World Bank International Debt Statistics (IDS) via Data API v2 **URL:** <https://api.worldbank.org/v2/> **Usage:** Table 1 (column: IFC disbursements as share of EMDE private non-guaranteed debt). **Note:** Fetched automatically by `Table_1_Summary_Statistics.py` at run time. A cached copy of the most recent fetch is saved to `Data/Public/ids_cache_dpng.csv` and used as a fallback if the API is unavailable.

Citation: World Bank. (2026). *International Debt Statistics* [API series DT.DIS.DPNG.CD]. World Bank Open Data.
<https://api.worldbank.org/v2/>. Retrieved April 2026.

3.13 World Development Indicators — Macro Panel (cached)

Filename: `wdi_macro_indicators_annual.csv` **Format:** CSV **Location in package:** `Data/Public/` **Source:** World Bank World Development Indicators (WDI), fetched via `wbgapi` **Date accessed:** June 2026 **Series:**

- `NY.GDP.MKTP.KN` — Real GDP, constant LCU (`gdp_real`)
- `FP.CPI.TOTL` — Consumer price index, 2010 = 100 (`cpi`)
- `NY.GDP.DEFL.ZS` — GDP deflator (`deflator`)
- `PA.NUS.FCRF` — Official exchange rate, LCU per USD, period average (`fx_rate`)

Coverage: Country-year panel, 1990–2024, all WDI economies. **Usage:** Table A9 macro covariate regressions — annualized real GDP growth, currency depreciation, and inflation over each loan's life. **Note:** `TableA9_Macro_Covariate_Analysis.py` reads this cache; if the file is deleted, the script re-fetches it from the WDI API (requires `wbgapi`) and re-saves it.

Citation: World Bank. (2026). *World Development Indicators* [Data file]. World Bank Open Data.
<https://data.worldbank.org/>. Retrieved June 2026.

4. Computational Requirements

Software	Version tested	Notes
Python	≥ 3.10	Tested on 3.10 (Anaconda conda env)
numpy	≥ 1.24	
pandas	≥ 2.0	
scipy	≥ 1.10	

Software	Version tested	Notes
statsmodels	≥ 0.14	
matplotlib	≥ 3.7	
openpyxl	≥ 3.1	Excel output
requests	≥ 2.31	Live API calls (Table 1)
pyreadstat	≥ 1.2	Required to read .dta files
wbgapi	≥ 1.0	<i>Optional</i> ; only to regenerate the Table A9 WDI cache (Data/Public/wdi_macro_indicators_annual.csv)

Install all dependencies with:

```
pip install -r requirements.txt
```

Execution of all scripts takes approximately **20–40 minutes** on a standard laptop (most time is spent in the PME/IRR numerical optimization loops).

5. Package Structure

The confidential data lives in a **sibling folder next to the package**, so the package folder itself can be shared without exposing it:

Loan data/
shared)
├─ Confidential_Data/
here
| ├── IFC default data_anonymized.dta
| ├── Loan cash flows anonymized.dta
| └─ Loan terms_anonymized.dta
└─ Reproducibility/

← (your private working directory; NOT
← the three confidential .dta files live
← THIS package – the folder you share

Package contents (**Reproducibility/**):

Reproducibility/
|
├─ README.md
Statement and dataset documentation)
├─ LICENSE
Section 9)
├─ sha256_checksums.txt
├─ data_hash_report.csv
datasets (WB hash-tool format)
├─ requirements.txt
├─ environment.txt
the verified run
└─ main.py

← This file (incl. Data Availability
← CC BY 4.0 (with exceptions; see
← SHA256 hashes of all input datasets
← SHA256 hash report of all input
← Python dependencies (version ranges)
← Exact pinned package versions used for
← Master run script (writes

Output/run_log.txt with run timestamps)

- ├─ Tools/
 - ├─ generate_data_hash_report.py ← Regenerates data_hash_report.csv (port of the WB create_hash.R tool)
 - └─ verify_run.py ← Post-run checks: input hashes, output timestamps, manuscript comparison

- └─ Returns to Private Lending In EMDEs-Draft-July 22-2026_CLEAN.docx ← Working paper manuscript

- ├─ Code/ ← All analysis scripts

- ├─ Table_1_Summary_Statistics.py
- ├─ Table_2_Portfolio>Returns_Benchmarks.py
- ├─ Table_3_IRR_Comparison.py
- ├─ Table_4_5_6_FCS_Analysis.py
- ├─ Table_7_IDA_Analysis.py
- ├─ Table_8_Downturn_PME_Regressions.py
- ├─ Figure_1_TableA1_PME_by_Vintage_Year.py
- ├─ Figure_2_3_Default_Rates_Downturns.py
- ├─ TableA2_A5_Tenor_Loan_Performance.py
- ├─ TableA6_Sensitivity_Analysis.py
- ├─ TableA7_Country_Matched_PME.py
- ├─ TableA8_Regional_PME.py
- ├─ TableA9_Macro_Covariate_Analysis.py
- └─ TableA10_Start_Year_Sensitivity.py

- ├─ Data/

- ├─ Confidential/ ← EMPTY in the shared package (holds only README_CONFIDENTIAL.md)

- ├─ README_CONFIDENTIAL.md ← explains where the confidential data lives

- ├─ Public/ ← All publicly shareable / licensed inputs

- ├─ jpm_cembi_broad_diversified_total_return.xlsx
- ├─ jpm_embi_broad_diversified_total_return.xlsx
- ├─ us_treasury_7_10yr_total_return_monthly.csv
- ├─ imf_real_gdp_growth_annual.csv
- ├─ iso_country_codes_master.xlsx
- ├─ wb_country_income_classification_historical.xlsx
- ├─ wb_ida_eligibility_country_year.xlsx
- ├─ ifc_fcs_ida_country_classification_2006_2025.xlsx
- ├─ wdi_macro_indicators_annual.csv ← WDI macro cache (Table A9); re-fetched if deleted
- ├─ ids_cache_dpng.csv ← World Bank IDS API cache (Table 1); re-fetched if deleted

- └─ Output/

- ├─ Tables/ ← Table outputs (regenerated by main.py)
 - ├─ table_1_summary_statistics.xlsx
 - ├─ table_2_benchmark_comparison.xlsx
 - ├─ table_3_irr_comparison.xlsx
 - ├─ table_4_5_6_fcs_analysis.xlsx
 - ├─ table_7_ida_analysis.xlsx
 - ├─ table_8_downturn_pme_regressions.xlsx
 - ├─ table_A1_pme_by_vintage_year.xlsx
 - ├─ table_A2_A5_tenor_loan_performance.xlsx
 - └─ table_A6_sensitivity_analysis.xlsx

```

├── table_A7_country_matched_pme.xlsx
├── table_A8_regional_pme.xlsx
├── table_A9_macro_covariate_analysis.xlsx
├── table_A10_start_year_sensitivity.xlsx
├── Figures/
│   ├── Figure_1_PME_by_Vintage_Year.png / .pdf
│   ├── Figure_2_Default_by_Downturn_Type.png
│   └── Figure_3_Default_Crisis_vs_Normal.png
└── run_log.txt                                ← Timestamp log of the run (created by
main.py; no PII)

```

6. Instructions to Replicators

Option A — Virtual verification (standard pathway for confidential data)

Because the primary data are confidential, replication follows the World Bank's **virtual verification** protocol:

1. The authors run the full script on their machine (with access to the confidential data) while the reviewer observes via screen sharing.
2. The reviewer examines the resulting outputs and console logs.
3. No confidential data is transferred to the reviewer.

To arrange a virtual verification session, contact the corresponding author.

Option B — Independent replication with data access

Researchers who have obtained access to the IFC data under an NDA can replicate independently:

1. Make the three confidential files available in **one** of these ways:
 - **(a) Sibling folder (default):** place them in a `Confidential_Data/` folder located next to the package folder (i.e. `../Confidential_Data/` relative to this package):
 - `Loan cash flows anonymized.dta`
 - `Loan terms anonymized.dta`
 - `IFC default data anonymized.dta`
 - **(b) Custom location:** set the `IFC_CONFIDENTIAL_DIR` environment variable to the folder containing the three files (see below).

2. Install dependencies and run from the package root:

```

pip install -r requirements.txt
python main.py

```

To use a custom confidential-data location instead of the sibling folder:

```

# Windows (PowerShell)
$env:IFC_CONFIDENTIAL_DIR = "D:\secure\Confidential_Data"; python main.py

```

```
# macOS / Linux
IFC_CONFIDENTIAL_DIR=/secure/Confidential_Data python main.py
```

3. All outputs will be regenerated in **Output/Tables/** and **Output/Figures/**.

Path resolution. Public data paths resolve automatically relative to **main.py**. The confidential-data folder resolves in this order: (1) **IFC_CONFIDENTIAL_DIR** if set; (2) a **Confidential_Data/** folder next to the package; (3) the in-package **Data/Confidential/** folder (legacy fallback). Confidential data is kept **outside** the package so the package can be shared without exposing it.

Running a single script. **main.py** performs the resolution above once and passes the result to every script through the **DATA_DIR** environment variable. A script launched on its own does not inherit that and falls back to the in-package **Data/Confidential/** folder. If the confidential files live anywhere else (such as the sibling **Confidential_Data/** folder), set **DATA_DIR** — note the **trailing slash** — before invoking the script:

```
# Windows (PowerShell)
$env:DATA_DIR = "..\Confidential_Data\"; python Code\Table_3_IRR_Comparison.py

# macOS / Linux
DATA_DIR=../Confidential_Data/ python Code/Table_3_IRR_Comparison.py
```

IFC_CONFIDENTIAL_DIR is read by **main.py** (and by **TableA10_Start_Year_Sensitivity.py**), not by the other scripts, so use **DATA_DIR** for single-script runs. Running **python main.py** needs none of this.

7. List of Tables, Figures, and Programs

Script	Paper output	Output file(s) in Output/Tables/ or Output/Figures/
<code>Table_1_Summary_Statistics.py</code>	Table 1	<code>table_1_summary_statistics.xlsx</code>
<code>Table_2_Portfolio_Returns_Benchmarks.py</code>	Table 2	<code>table_2_benchmark_comparison.xlsx</code>
<code>Table_3_IRR_Comparison.py</code>	Table 3	<code>table_3_irr_comparison.xlsx</code>
<code>Table_4_5_6_FCS_Analysis.py</code>	Tables 4, 5, 6	<code>table_4_5_6_fcs_analysis.xlsx</code>
<code>Table_7_IDA_Analysis.py</code>	Table 7	<code>table_7_ida_analysis.xlsx</code>
<code>Table_8_Downturn_PME_Regressions.py</code>	Table 8	<code>table_8_downturn_pme_regressions.xlsx</code>
<code>Figure_1_TableA1_PME_by_Vintage_Year.py</code>	Figure 1, Table A1	<code>Figure_1_PME_by_Vintage_Year.png/.pdf</code> ; <code>Figure_1_PME_by_Vintage_Year_DualAxis.png/.pdf</code> (development variant, not in paper); <code>table_A1_pme_by_vintage_year.xlsx</code>
<code>Figure_2_3_Default_Rates_Downturns.py</code>	Figures 2–3	<code>Figure_2_Default_by_Downturn_Type.png</code> , <code>Figure_3_Default_Crisis_vs_Normal.png</code>
<code>TableA2_A5_Tenor_Loan_Performance.py</code>	Tables A2–A5	<code>table_A2_A5_tenor_loan_performance.xlsx</code>

Script	Paper output	Output file(s) in <code>Output/Tables/</code> or <code>Output/Figures/</code>
<code>TableA6_Sensitivity_Analysis.py</code>	Table A6	<code>table_A6_sensitivity_analysis.xlsx</code>
<code>TableA7_Country_Matched_PME.py</code>	Table A7	<code>table_A7_country_matched_pme.xlsx</code>
<code>TableA8_Regional_PME.py</code>	Table A8	<code>table_A8_regional_pme.xlsx</code>
<code>TableA9_Macro_Covariate_Analysis.py</code>	Table A9	<code>table_A9_macro_covariate_analysis.xlsx</code>
<code>TableA10_Start_Year_Sensitivity.py</code>	Table A10	<code>table_A10_start_year_sensitivity.xlsx</code>

Note on Table 2 structure: Table 2 reports the full portfolio against CEMBI Broad Diversified, EMBI Broad Diversified and US Treasury, split by the year of each loan's first disbursement into three blocks — pre-GFC 2002–2006, GFC-era 2007–2011, post-GFC 2012–2024 — that partition the sample exactly (694 + 765 + 1,429 = 2,888 loans). A final row reports the most recent five years, 2019–2024; it is a **subset** of the post-GFC block, shown indented below a rule, and is deliberately excluded from the partition assertion in the script. All three benchmark series cover 2002–2024 without gaps, so a single common-period sample serves every column and the former two-panel full-history/common-period structure is no longer needed. Table A10 reports how the aggregate-CEMBI figure varies with the sample start date and is cited alongside Table 2.

Note on Table A7 country-matched benchmarks (formerly Table 4): Table A7 retains every loan in a country with a CEMBI sub-index (1,891 loans, 40 countries) with its full cash-flow stream. Country sub-indices cannot be used raw. Four adjustments are applied in `build_country_index()`: (1) JP Morgan resets a sub-index to 100 when a country re-enters the index; each such reset (13 events) is chained so the series is continuous, with continuity imposed via the aggregate CEMBI's return over the gap; (2) months before a sub-index begins, inside interior gaps, and after it ends are filled with the aggregate CEMBI rescaled to meet the sub-index at the adjoining observation, so no cash flow is dropped for lack of a benchmark value; (3) every series is normalised to 100 at January 2002, because the pooled PME weights each cash flow by the inverse of its benchmark level and raw levels differ several-fold across countries by start date. Each constructed series thus measures the cumulative return on a dollar in EM corporate debt from 2002, switching to the country's own index when available. The aggregate CEMBI and Treasury series are single gap-free indices and need none of this. Panel B splits the identical sample by year of first disbursement into three blocks that partition it exactly. The figures are independently verified by `Audit_Table4_Truncation/Code/verify_independent.py`, a from-scratch reimplementa-tion that asserts cash-flow conservation and exits non-zero on any deviation.

Note on Table 8 PME regressions (downturn): Loan-level PME values are computed inline from the IFC cash flow data and the CEMBI index within `Table_8_Downturn_PME_Regressions.py`. The regression sample is restricted to fully settled loans (no positive outstanding balance at end-2024), with non-missing contracted tenor and income classification (World Bank 2005 benchmark).

Note on sample filter counts (Data Filters and Sample Adjustments): The sequential application of sample filters is documented in the paper. The filter logic is embedded in `Table_2_Portfolio>Returns_Benchmarks.py`, which applies the full filter stack and prints per-step loan counts to the console during execution.

8. Key Methodological Notes

- **PME (Public Market Equivalent):** All PME calculations use the Kaplan-Schoar (2005) method. The portfolio is scaled by the CEMBI Broad Diversified cumulative total return index unless otherwise noted.
 - **Sample period:** 2002-01-01 through 2024-12-31 (common period used across all main tables).
 - **Sample filters** applied consistently across all scripts:
 1. 6-key deduplication (removes double-posted accounting entries)
 2. Filter A': loans must have at least one **Disbursement** transaction
 3. Blended-finance exclusion (**BF == 1**)
 4. Non-USD disbursement exclusion
 5. Short-tenor exclusion (projects with date span ≤ 31 days)
 6. Equity-conversion exclusion (**Net Conversion to Equity** transaction present)
 - **Income classification:** Fixed 2005 World Bank income classification, sourced from [wb_country_income_classification_historical.xlsx](#) (sheet "Country Analytical History").
 - **FCS classification:** IFC FCS list 2006–2025 ([ifc_fcs_ida_country_classification_2006_2025.xlsx](#)). Note: this file uses IFC internal country codes (e.g. **ANG**, **BUI**, **SRM**) rather than standard ISO 3166-1 alpha-3 codes. [Table_4_5_6_FCS_Analysis.py](#) applies a built-in crosswalk to convert these to standard ISO3 before matching loans.
 - **Live API calls:** [Table_1_Summary_Statistics.py](#) fetches IDS disbursement totals from the World Bank Data API (v2) to compute the IFC share of EMDE private non-guaranteed debt. If the API is unavailable, a cached version ([Data/Public/](#) contains [ids_cache_dpng.csv](#) if previously fetched) is used automatically.
-

9. License

The package (code, documentation, and publicly shareable data files) is licensed under the **Creative Commons Attribution 4.0 International License (CC BY 4.0)** — see the [LICENSE](#) file. Exceptions: the confidential IFC datasets are the property of IFC and are subject to the access restrictions described in Section 2; the JPMorgan CEMBI/EMBI index files are proprietary to JPMorgan Chase & Co., are included solely for World Bank virtual verification purposes, and may not be redistributed further.

10. Contact

Florian Moelders International Finance Corporation / World Bank Group fmoelders@ifc.org

This package follows the [World Bank Reproducible Research Repository](#) guidelines. This README consolidates the package documentation, the Data Availability Statement (Section 2), and the dataset documentation (Section 3) into a single document.