

Predicting Debt Distress in Low-Income Countries

Second Submission: PP_WLD_2026_712

Maria Reyes Retana
reproducibility@worldbank.org

This review verifies the reproducibility of the exhibits included in the paper “*Predicting Debt Distress in Low-Income Countries*” (Graf von Luckner, Horn, Kraay, Ramalho).

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit of the paper has been reproduced accurately.**
- The code was successfully executed on a new computer after:
 1. Setting the top-level path global in master do file.do.
 2. Creating a the python environment from requirements.txt or environment.yml. Installed: pandas <3.0 (pandas 3.x breaks a literal_eval column-assignment step), numpy, statsmodels, scikit-learn, seaborn, lxml, multiprocesspandas, openpyxl.
 3. Setting the BASE_PATH global at the top of both analysis notebooks (model_analyses_Mar2026.ipynb and appendix_exercises_Mar2026.ipynb) – the single path each notebook requires.
 4. Running model_analyses with depvar = 't12345' and apply_filter = 1 to reproduce Table 3 (the sensibility-filtered five-year models; the default apply_filter = 0 produces the unconstrained set instead).
- **Stability across runs:** Results show stability, producing the same results after two consecutive runs.
- **Runtime:**
 1. Stata: 2 minutes.
 2. Python: 4 minutes. model_analyses_Mar2026.ipynb and appendix_exercises_Mar2026.ipynb each complete in 2 minutes with RUN_HEAVY=False (the default, using precomputed result files as the README specifies). The cluster-only j_k_fold model-search notebook and the RUN_HEAVY=True branches (many cores, hours) were not run, matching the README’s own guidance not to rerun them for a standard replication.
- We conducted our reproducibility analysis based on the manuscript shared by the authors on August 5.

- **Verification Process and Data Handling:**

- The reproducibility package relies on 2 types of data: open and restricted data (available only from the data owners).
- Open data is included in the public reproducibility package.
- Reviewers used restricted data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package.
- *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the forthcoming/accessible/limited-access/restricted data can use this file to verify that the data has not been altered.

- **Reproducibility Summary:**

- **Data:** Some data is restricted and has not been included in the reproducibility package. For more details, refer to the README file.
- **Code:** All code files (from cleaning to analysis, Stata and Python) are included in the reproducibility package. The cluster-only `j_k_fold` model-search notebook and the `RUN_HEAVY=True` branches (many cores, hours) were not run.
- **Outputs:** All outputs checked in this review are generated by code included in the reproducibility package.
- **Dependencies environment:** The reviewers created a new environment (Stata `ado/` folder; Python conda environment) using package versions compatible with the code, since none was specified in the submission.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1. Build-up and resolution of arrears around external debt restructurings** Reproduced.
- **Figure 2. Debt distress episodes by signal and over time** Reproduced.
- **Figure 3. Error rates and loss functions for full model space** Reproduced.
- **Figure 4. Loss functions and model parsimony** Reproduced.
- **Table 1** (Top performing prediction models by number of predictors) Reproduced with manual edits. Exact match, every cell, against `model_analyses_Mar2026.ipynb` (cell computing `temp_result`, with the `apply_filter` toggle at its default value of 0). Not produced by the Stata do-file named "table 1...do" despite the filename – that script computes unrelated descriptive counts.
- **Table 2** (Best prediction models for one-year ahead debt distress prediction) Reproduced with manual edits. Exact match, every cell, against the *same* notebook cell as Table 1, with `apply_filter` set to 1. This toggle is undocumented in the README and the notebook.
- **Figure 5. Re-estimation of top models in larger data sample** Reproduced.

- **Figure 6. Error rates and loss functions for full model space – alternative algorithms** Reproduced.
- **Table 3** (Top performing prediction models for five-year ahead debt distress prediction) Reproduced with manual edits. Per the code's own markdown label, this table is produced by the same cell as Tables 1–2, rerun with the notebook's depvar toggle set to 't12345' instead of the default 't1'.
- **Table 4** (Confusion Matrix – BPM and LIC DSF Classification versus Actual) Reproduced with manual edits. Exact match on both confusion matrices (903/394/15/44 and 814/483/19/40) and all derived rates (FPR/FNR/loss function), against model_analyses_Mar2026.ipynb's confusion-matrix cell.

Results in the Annex

- **Table A1** (Sample coverage) Reproduced with manual edits. Independently re-derived in pandas from `licdsf_final_database.dta` using the same logic as `table a1...do` (per-country first/last year and observation count). Every one of the 79 countries printed in the manuscript matches exactly.
- **Table A2** (Sample coverage) Does not show analysis results.
- **Figure A1** (Share of LICs in external debt distress, 1970–2021) Reproduced. Verified against code output "FigureA2 _ Share of countries in debt distress.png".
- **Figure A2** Reproduced. Verified by the reviewer against code output "figure a1 _ external debt distress by signal country year all countries.png".
- **Figure A3** Reproduced. Verified against "figure A3 _ BOC vs depvar.png"
- **Figure A4** Reproduced. Verified against "figure a3 _ pc and external debt distress.png".
- **Figure A5** Reproduced. Verified by the reviewer against code output "figure a4 _ domestic and external debt distress episodes.png".
- **Figure A6** Reproduced. Verified by the reviewer against code output "figure a5 _ debt service country cases.png".
- **Figure A7** Reproduced. Verified by the reviewer against code output "figure a6 _ remittances.png".
- **Table B1** (Predictive performance by predictor variable) Reproduced. Exact match, every row, against `appendix_exercises_Mar2026.ipynb` (cell following the "Appendix table B1" markdown header). One label inconsistency: the manuscript's row "Credit History 2" has the exact values of the code's `credit_hist3` variable (labeled "Credit History 3" in code); the code has no row corresponding to a variable called "Credit History 2". Though the underlying numbers are correct, this is marked as reproducible.
- **Table B2** (Top performing prediction models for five-year ahead debt distress prediction) Reproduced with manual edits.
- **Table B3** (Top performing models by number of predictors – the HIPC Initiative) Reproduced.
- **Table B4** (Investigating coefficient signs on debt stock and debt flow measures) Reproduced.
- **Table B5** (Investigating coefficient signs on GDP per capita) Reproduced.

- **Figure B1** (GDP per capita and borrowing costs) **Reproduced**. Verified by the reviewer against code output "gdp_pc_borrowing_costs.png" (produced by Appendix B – Investigate gdp_pc.do).
- **Table B6** (Beta coefficients as a function of lambda) **Reproduced**.
- **Table B7, B8, B9** (Best models with 3/4/5 variables under different loss function weights) **Reproduced**.
- **Table C1** (Probit Regressions underlying the existing LIC DSF Model) **Reproduced**. Exact match, all four columns (GDP, Exports, Revenue, Exports), against Stata's "table 4 _ reproduce old probit regressions.do".
- **Table C2** (Definition of external debt distress episodes in LIC DSF and this paper) Does not show analysis results.
- **Figure B2** (Model performance of LASSO algorithm) **Reproduced**.
- **Table C3** (Confusion Matrix – BPM and LIC DSF Classification in 2017 LIC DSF Sample) **Reproduced**.

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: macOS 26.5.2 (Darwin)
- Processor: Apple M4 Pro
- Memory available: 24.0 GB
- Software version: StataNow 19.5 MP; Python 3.12