

From Shadows to Society: Medium-Run Effects of Regularization on Forced Migrants in Colombia (June 2026 final draft).

Author: *Sandra V. Rozo.*

Use: This replication package is prepared for the World Bank's Reproducible Research Repository verification process. It is not intended for public release. The proprietary VenRePS microdata are governed by a data-use agreement with the Inter-American Development Bank and are forthcoming on the World Bank Microdata Library; reviewers acquire them through the channels described in ``data/DATA_AVAILABILITY.md``.

This folder contains the data, code, and pre-built outputs that reproduce every exhibit referenced in the manuscript (``main.tex`` and ``appendix.tex`` in ``Overleaf/Overleaf/``).

Manuscript-to-code integrity: Every ``*.tex`` exhibit in ``Overleaf/Overleaf/tables/`` is byte-identical to its counterpart in ``outputs/tables/`` or ``outputs/appendix/tables/``. The manuscript reflects the current state of the code and data. Running ``master.do`` regenerates every table and figure cited in the body and the online appendix.

Stability: The pipeline is deterministic. Running ``master.do`` twice produces byte-identical ``*.tex`` outputs and identical values for the analysis-sample variables (``trust_index`` = 13,298 non-missing, ``mh_index`` = 13,570). PDF figures embed creation timestamps so PDF byte-diffs across runs will differ; the figure content is stable.

Folder layout

replication/

code/ Stata do-files (master.do runs the full pipeline)

data/ analysis sample + raw VenRePS inputs (proprietary)

outputs/ regenerated tables (.tex) and figures (.pdf/.png) for the LaTeX project

To replicate

1. Install Stata 17 or later.
2. Install user-written commands once: ``rdrobust``, ``rddensity``, ``estout``, ``coefplot``.
3. Open Stata, ``cd`` into ``code/``, then run ``do master.do``.

4. Outputs land in ``outputs/tables/``, ``outputs/figures/``, ``outputs/appendix/tables/``, and ``outputs/appendix/figures/``. The LaTeX project under ``Overleaf/Overleaf/`` pulls every exhibit from these locations.

Expected runtime is roughly two to three hours on a recent laptop. The pre-built outputs shipped in this package are produced by the same ``master.do`` and let a replicator compile the manuscript without re-running Stata first.

Data availability

The proprietary microdata (``VenReps_Panel_w1&w2.dta`` and ``PEP_Panel_Clean.dta``) are governed by the data-use agreement with the Inter-American Development Bank. Bona fide researchers can request access directly from the IDB. The analysis-ready file ``PEP_Analysis_Sample.dta`` shipped with this package is sufficient to reproduce every table and figure in the paper without re-running the data-construction step (``00_BuildPanel.do``). Researchers with access to the raw VenRePS files can rebuild ``PEP_Analysis_Sample.dta`` from scratch.

Pipeline overview

The ``master.do`` script runs the following stages in order:

- 1. Data construction.** ``00_BuildPanel.do``, ``01_SampleAnalysis.do``, ``03_OutcomeConstruction.do`` build the analysis sample from the raw VenRePS panel and persist the JEEA-style thematic indices (including the mental-health composite and the prosocial-behaviors composite).
- 2. JEEA short-run replication.** ``02_JEEA_Replication.do`` reproduces the short-run estimates of Ibanez et al. (2025) on the Wave-1 sample as a baseline cross-check.
- 3. Identifying assumptions.** ``04_Descriptives.do``, ``05_Validity.do``, ``06_FirstStage.do`` produce the McCrary density test, the joint balance test, and the first-stage discontinuity reported in Table 2 of the body.
- 4. Headline fuzzy-RD estimates.** ``07_RD_W1_W2.do`` produces the wave-by-wave fuzzy LATE estimates for every outcome. ``07b_*`, ``07c_*`, ``07d_*` build the coef-plot, RD-plot, and AER-style table inputs.
- 5. Attrition and Lee bounds.** ``08_Attrition.do``, ``08_LeeBounds_W2.do``, ``08b_LeeBoundsPlot.do`` produce the attrition diagnostics and the Lee-bound trimming bounds.
- 6. Main-paper figures and tables:** ``11_PaperFigures.do`` and ``11b_PaperTables.do`` assemble the body's seven figures and ten tables.

7. Appendix tables and figures: `12\_AppendixFigures.do`, `12\_FiscalAnalysisW2.do`, `12b\_AppendixTables.do`, plus the diagnostic scripts (`14\_\*`, `15\_\*`, `16\_\*`, `18\_\*`, `20\_AppendixA\_Comprehensive.do`) build the per-domain RD tables and the fiscal-balance figure.

8. Mechanisms: `15\_ProsocialMediation.do`, `21\_ServicesMediation.do`, `22\_\*`, `22b\_\*`, `22c\_\*`, `22d\_\*`, `29\_DurationTests.do` produce the mediation tables, the duration-interaction tests, the cohort-gradient test, and the renewal descriptive piece.

9. Robustness: `26\_ConsumptionRobustness.do` produces the consumption-attenuation robustness panel for Online Appendix D.

Exhibit inventory

The manuscript references 10 body tables, 5 body figures, 30 appendix tables, and 4 appendix figures. The pre-built versions of all 49 exhibits are in `outputs/`.

Body

- **Tables:** `tab2\_validity`, `tab3\_wellbeing`, `tab4\_services`, `tab5\_labor`, `tab6\_integration`, `tab7\_trust`, `tab8\_food`, `tab9\_housing`, `tabA\_balance\_full`, `mccrary\_pvalue` (snippet).

- **Figures:** `fig1\_first\_stage`, `fig3\_rd\_headline`, `fig5\_leebounds`, `fig7\_bw\_sweep`, `fig\_fiscal`.

Appendix

- **Tables:** `tabA\_attrition\_balance`, `tabA\_attrition\_summary`, `tabA\_consumption\_robustness`, `tabA\_desc\_\*` (13 domain blocks), `tabA\_fd\_attrition`, `tabA\_fd\_leebounds`, `tabA\_fd\_validity`, `tabA\_fiscal\_w2`, `tabA\_leebounds\_full`, `tabA\_panel\_subsample`, `tabA\_rd\_full\_discrim`, `tabA\_rd\_full\_education`, `tabA\_rd\_full\_migration`, `tabA\_rd\_full\_remittance`, `tabA\_reduced\_form`, `tabA\_services\_mediation`.

- **Figures:** `fig\_hist\_consumption`, `fig\_hist\_labor\_income`, `density\_w1`, `density\_w2`.

What was removed in this revision

The Robustness, Power-and-Minimum-Detectable-Effects, Romano-Wolf, Complier-Characterization, and Ibanez-Replication appendices have been removed from the manuscript. The do-files that produced their exhibits (`10\_Robustness.do`, `10b\_RobustnessPlots.do`, `10c\_PlaceboCutoffs.do`, `16\_FormalProbit.do`,

`19\_ComplierCharacterization.do`, `19\_RomanoWolf.do`, `23\_DurationTriple.do`,
`23\_MainExhibits.do`, `24\_AppendixFRobustness.do`, `27\_TSPI\_Duration.do`) are not
shipped with this package. They are preserved in the earlier working folder
`07\_Revision/replication\_materials/code/` if they need to be revived.

The mediation table for the housing-quality index (former Panel C of
`tabA\_services\_mediation`) has also been removed. The augmented Wave-2 mediation
table now reports only the integration and prosocial-behavior panels and is organized into
two theoretical sub-blocks within each panel: (a) the service-access cascade (SISBEN
enrollment, subsidized health insurance, and government transfers, the three components
of the JEEA service-access index) and (b) material precarity (housing quality, food security,
and the physical-health PCA index).

The panel-based dynamic specifications appendix has been replaced with a shorter
explanation in `5.1` of the body: PEP-RAMV take-up is essentially time-invariant after the
2018 rollout, so the within-individual change in program status is mechanically near zero
and the first stage of the within-individual fuzzy RD collapses. The panel-FD diagnostics
that supported that explanation (`tabA\_fd\_validity`, `tabA\_fd\_attrition`,
`tabA\_fd\_leebounds`) remain in the appendix.