

Distributional Effects of Market-Oriented Reforms in Uzbekistan: A Macro-Micro Simulation Approach

README

1. Overview

This package contains the Stata do-files used to simulate the distributional and poverty impacts of Uzbekistan's Presidential Decree UP-85 across several policy scenarios (Mono, Trade, Dom, Mono_Trade, All, All_FDI, BAU/base).

The approach follows a CGE-microsimulation ("top-down", GIDD-style) design: macroeconomic targets (employment, wages, prices, non-labor income) generated by a CGE model are read in as exogenous scenario targets, and then imposed on the 2024 Uzbekistan Household Budget Survey (HBS) microdata through a sequential simulation:

1. Reweight the HBS microdata to match Wittgenstein Centre population projections.
2. Estimate employment propensity (logit), sector affinity (multinomial logit) and a Mincer wage equation on the reweighted microdata.
3. Simulate individual employment status stochastically (rank-based on scaled probability scores) to match macro employment targets by skill.
4. Reallocate simulated workers across sectors (stayers / separations / hires) to match macro employment targets by skill-sector cell.
5. Assign simulated wages (Mincer-predicted + stochastic residual, rescaled to match macro wage targets by skill-sector cell).
6. Deflate/scale income (labor and non-labor) to match macro price and income (GDP) targets.
7. Produce distributional and poverty analysis (by scenario and by three poverty lines: national poverty line "npl", vulnerability line "vul", and the \$8.30 UMIC line "8_3"), including growth incidence curves, poverty headcounts by urban/rural, and wage distributions by gender/skill/sector.

The code is fully deterministic: a global seed (12345) is set in 00_Housekeeping.do, and each simulation/reallocation/wage-assignment step uses a further seed derived deterministically from the base seed and the survey year. Stata/MP is forced to run on a single processor to guarantee bit-identical results across machines.

2. Data Availability

Overall status:

- ☐ All data publicly available
- ☒ Some data cannot be made publicly available
- ☐ No data can be made publicly available

2.1 Data sources

(A) *Uzbekistan 2024 Household Budget Survey (HBS)*

- **Filenames:** /Data/Raw Data/UZB 2024 Household Budget Survey/uzb_hbs_2024_20241231_M00_weight.dta, ..._M01_roster.dta, ..._M02_migration.dta, ..._M03_education.dta, ..._M05_employment.dta, ..._M07_socben.dta, ..._M08_nonwage.dta, ..._M09_food.dta, ..._M10_nfa.dta, ..._M10_nfb.dta, ..._M10_nfc1.dta, ..._M10_nfc2.dta, ..._M10_nfd.dta, Poverty_Analysis_File@20241231.dta
- **Access:** Restricted.
- **Note:** The HBS 2024 microdata is not publicly available and is not included in this reproducibility package. Data was accessed internally via the World Bank Stata package datalibweb, which is only accessible to World Bank staff and consultants using Bank-managed credentials (see <https://github.com/worldbank/datalibweb>). There is no documented request procedure for external (non-World Bank) users to obtain this dataset. Interested readers — internal or external — should contact Ikuko Uochi (iuochi@worldbank.org) to inquire about access, an NDA, or a restricted-access arrangement. For World Bank staff/consultants seeking to reproduce this step: use datalibweb in Stata to browse the Uzbekistan country/collection tree and locate the 2024 HBS modules listed above (the exact datalibweb survey ID was not confirmed at the time of writing — verify with Ikuko Uochi or the ECA Poverty TSD team); place the resulting .dta files in the folder structure described in Section 6.
- **Access year:** 2025
- **Citation:** Statistics Agency of Uzbekistan. 2024. "Uzbekistan Household Budget Survey (HBS) 2024" [dataset]. Accessed 2025.

(B) *Wittgenstein Centre Population Projections (wcde_data.csv)*

- **Source:** Wittgenstein Centre for Demography and Global Human Capital (WIC) Wittgenstein Centre Data Explorer. Version 2.0 2018'
- **URL:** <https://dataexplorer.wittgensteincentre.org/wcde-v2/>
- **Access:** Public
- **Access year:** 2025
- **Citation:** Wittgenstein Centre for Demography and Global Human Capital (WIC). Wittgenstein Centre Data Explorer, Version 2.0 (2018). <https://dataexplorer.wittgensteincentre.org/wcde-v2/> . Accessed 2025.
- **Note:** To download the exact file:
Go to <https://dataexplorer.wittgensteincentre.org/wcde-v2/> and apply the following filters:
 1. Indicators: Indicator Type: Population & Human Capital Stocks; Indicator: Population Size by Education (000's)
 2. Geography: Country: Uzbekistan; Region: Leave blank
 3. Breakdown: Sex: Male Female (Do not select "Both"); Age: Select all age groups (0--4, 5--9, ..., 100+), except "All".
 4. Time Horizon: Scenario: Medium (SSP2); Year: Select 2020, 2025, 2030, 2035, 2040

Click Download and you will get a csv file named "wcde_data.csv"

(C) CGE model results — intermediate dataset

- **Filename:** Microsim_inputs_240326.xlsx (sheets: Mono, Trade, Dom, Mono_Trade, All, All_FDI, and "real income (expenditure) pc")
- **Access:** Public
- **Note:** The Excel file is produced by Alhassane Camara (World Bank), as the output of a separate CGE model and dataset calibrated for Uzbekistan under Presidential Decree UP-85 scenarios. The construction of this file (the CGE model itself) is out of scope for this Stata package; it is used here as a documented, exogenous macro-target input. The MANAGE model is available in a publicly accessible repository at <https://github.com/worldbank/MANAGE-WB>. The model documentation is available at the following source: <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099543511132534267>. The version used in this study builds on the same open version mentioned above. However, we employ the GTAP-Power extension of the GTAP 11 Database, which disaggregates the single electricity sector into transmission & distribution and individual generation technologies (coal, gas, oil, nuclear, hydro, wind, solar, etc.). This provides more granular power-sector representation than the standard GTAP 9 Database (free) used in the open version and also incorporates the updated cost and reference-year data of the version 11 vintage. This file (microsim_inputs_240326) has been produced on March 24, 2026. The data are freely available and may be shared alongside the package. Interested readers with specific questions may contact acamara8@worldbank.org. The data are stored in the "Intermediate Data" folder. They are not generated by the code but are required as input to run it.
- **Citation:** Camara, Alhassane (World Bank). 2026. "CGE macro-target inputs for Uzbekistan under Presidential Decree UP-85 scenarios" (Microsim_inputs_240326.xlsx) [dataset].

2.2 Statement about Rights

- ☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- ☒ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permissions are documented in the LICENSE.txt file.

3. Requirements

- Stata 18 or higher
- Stata packages (user-written, install from SSC before running):
 - winsor2 (used to trim extreme wage/income values in the HBS cleaning and baseline-income steps)
 - labutil (used to order/label provinces in the distributional-analysis graphs)

Reproducibility settings (already built into the code; no action needed):

- set rng mt64 (Mersenne Twister 64-bit random number generator)
- set seed 12345 (global seed)
- set sortseed 123456 (sort seed)
- set type double (avoids float rounding differences)
- set processors 1 (forces single-threaded mode on Stata/MP to prevent

parallel sorting from producing different orders)

- Per-year seed in simulation loops: $\text{seed} = 12345 + \text{year}$

Operating system: developed using Windows-style paths (backslashes, e.g. "\$WorkDir\Code\..."). Replicators on macOS/Linux will need to either run under Windows/a Windows VM, or adjust path separators throughout (replace "\" with "/"; Stata accepts "/" on all platforms if paths are edited consistently).

4. Folder Structure

Create the following folders BEFORE running anything. All folders must exist even if empty — Stata will not create them automatically.

```
Reproducibility Package\                                <- your root folder ($WorkDir)
|
+-- Code\                                                <- all do-files go here
|   +-- ado\
+-- Data\
|   +-- Raw Data\
|       +-- UZB 2024 Household Budget Survey\
|           +-- uzb_hbs_2024_20241231_M00_weight.dta
|           +-- uzb_hbs_2024_20241231_M01_roster.dta
|           +-- uzb_hbs_2024_20241231_M02_migration.dta
|           +-- uzb_hbs_2024_20241231_M03_education.dta
|           +-- uzb_hbs_2024_20241231_M05_employment.dta
|           +-- uzb_hbs_2024_20241231_M07_socben.dta
|           +-- uzb_hbs_2024_20241231_M08_nonwage.dta
|           +-- uzb_hbs_2024_20241231_M09_food.dta
|           +-- uzb_hbs_2024_20241231_M10_nfa.dta
|           +-- uzb_hbs_2024_20241231_M10_nfb.dta
|           +-- uzb_hbs_2024_20241231_M10_nfc1.dta
|           +-- uzb_hbs_2024_20241231_M10_nfc2.dta
|           +-- uzb_hbs_2024_20241231_M10_nfd.dta
|           +-- Poverty_Analysis_File@20241231.dta
|       +-- UZB 2024 Wittgenstein Population Projection\
|           +-- wcde_data.csv
|   +-- Intermediate Data\
|       +-- Microsim_inputs_240326.xlsx                <- CGE results (must exist beforehand;
|                                                         produced by Alhassane Camara)
|                                                         Required sheets: Mono, Trade, Dom,
|                                                         Mono_Trade, All, All_FDI,
|                                                         real income (expenditure) pc
|   +-- Cleaned Data\
+-- Output\
    +-- Graphs\                                         <- all figures exported here
    +-- Tables\                                         <- poverty timeseries Excel exported here
```

5. Edit Main_script.do

CHANGE 1 — Change the username (lines 5 and 7)

CHANGE 2 — Change the working directory path (line 8)

6. Step 3 — Run

Open Stata, navigate to your code folder, and run:

```
do "Main_script.do"
```

The script runs the full pipeline in order:

```
01  Clean HBS 2024 microdata
02  Clean Wittgenstein population projections
03  Read and clean CGE results
04b Reweight microdata to match population projections
05  Estimate employment propensity, sector affinity, wage coefficients
06  Simulate employment status
07  Simulate sectoral reallocation
08  Assign wages to simulated employment
09  Deflate income and adjust non-labor income
10b Generate cross-scenario distributional figures (runs 3x: npl, vul, 8_3)
10c Generate Full Reform figures (runs 3x: npl, vul, 8_3)
10d Generate FDI comparison figure
```

A datasignature check is run at the end of Main_script.do; compare the signature against the one recorded by the authors to confirm the run is stable/identical.

7. List of Exhibits

The provided code reproduces:

- ☐ All tables and figures in the paper
- ☐ All numbers provided in text in the paper
- ☒ Selected tables and figures only, as explained below

All figures are saved as .png files in Output\Graphs\

File	Paper figure
Fig5_Deflated_income_by_scenario	Figure 5
Fig6_UMIC_poverty_rate_by_scenario	Figure 6
Fig7_Income_distribution_shift	Figure 7
Fig8_Poverty_rate_by_region	Figure 8
Fig9_Wage_by_skill_full_reform	Figure 9
Fig10_Wage_by_gender_full_reform	Figure 10
Fig11_Wage_by_sector_full_reform	Figure 11

Fig12_Poverty_by_location_UMIC	Figure 12
Fig13_Growth_incidence_curves	Figure 13
Fig14_Poverty_by_location_NPL	Figure 14
Fig15_Poverty_by_location_vulnerability	Figure 15
Fig16_Vulnerability_rate_by_scenario	Figure 16
Fig17_National_poverty_rate_by_scenario	Figure 17
FigA51_Wage_by_skill_all_scenarios	Figure A5.1
FigA52_Wage_by_gender_all_scenarios	Figure A5.2
FigA61_Employment_by_sector_all_scenarios	Figure A6.1
FigA62_Wage_by_skill_all_scenarios	Figure A6.2
FigA63_Wage_by_sector_all_scenarios	Figure A6.3
FigA64_Real_wage_all_scenarios	Figure A6.4
FigA65_Poverty_by_urban_rural_all_scen	Figure A6.5
FigA81_Income_paths_full_reform	Figure A.8.1
FigA82_Income_FDI_vs_All	Figure A.8.2

Table A7.1, A7.2, A7.3

Tables A7.1, A7.2, and A7.3 are reproduced following these instructions: Three Excel files with poverty time series by scenario is also saved in Output\Tables\Percentage_change_timeseries.xlsx for each of the three poverty lines (npl, vul, 8_3) when running the code.

These three files have the information for Tables A7.1 (npl), A7.2(vul) and A7.3 (8_3).

To reproduce the content in the tables:

1. Open each file "Percentage_change_timeseries.xlsx"
2. Select all data in Sheet1, including headers
3. Select Insert -> pivot table
4. Select the following: Columns -> scenario, Rows -> year, Values -> pp_poor

Table 1

It presents the baseline macroeconomic projections used as inputs to the MANAGE model; the model does not generate these projections. They are drawn from the 2025 Macro Poverty

Outlook (MPO) and differ from the updated projections now available at <https://mtimodelling.worldbank.org/livempodata/mpodata.html>.

To update the variables reported in Table 1, follow these steps:

Go to <https://mtimodelling.worldbank.org/livempodata/mpodata.html> and select UZB. Select and download the following variables:

- UZBBNCABFUND CD — Current Account Balance Value Millions USD
- UZBGGBALOVRLCN — Overall Fiscal Balance Value Millions LCU
- UZBNECONGOVTKN — Government Consumption Volume Millions LCU
- UZBNECONPRVTKN — Private Consumption Volume Millions LCU
- UZBNVAGRTOTLKN — Agriculture Volume Millions LCU
- UZBNVINDTOTLKN — Industry Volume Millions LCU
- UZBNVSRVTOTLKN — Services Volume Millions LCU
- UZBNYGDPMKTPCD — Gross Domestic Product at Market Price Value Millions USD
- UZBNYGDPMKTPCN — Gross Domestic Product at Market Price Value Millions LCU
- UZBNYGDPMKTPKN — Gross Domestic Product at Market Price Volume Millions LCU

Open the Table 1_Baseline Projections.xlsx Excel file and:

a. Paste the raw data into the Baseline MPO sheet, beginning in cell A1. Table 1 will be generated automatically beginning in cell O1. Note that the current Table 1 was prepared using an earlier version of the MPO projections and the same procedure described above.

Table 2

It shows the estimate of markups in monopolistic sectors used as inputs to the MANAGE model; the model does not generate these projections. To use them as input, we weighted the empirically estimated markups by the sectoral VA share. To get the data in Table 2, below are the steps: (1) Empirical estimate of markups using survey data (Stata program can be provided upon request). (2) Open the Table 2-estimated Markups.xlsx Excel file and report the estimated markups in column C. The final results will be generated in column E.

Table 3

It shows the estimated forgone export tax revenue, used as inputs to the MANAGE model; the model does not generate these projections. To use them as input, we transform the estimated revenue as a share of 2024 GDP. To get the data in Table 3, below are the steps: (1) Empirical estimate of forgone export tax revenue using survey data and UNCOMTRADE (Stata program can be provided upon request). (2) Open the Table 3-estimated forgone export tax revenue.xlsx Excel file and report the estimated forgone revenue in column C. The final results will be generated in column D.

Tables A4.1–A4.9

These tables are entirely generated from MANAGE model outputs. To replicate them, follow the steps below:

For Tables A4.1 and A4.7–A4.9, open the Table A4.1_A4.7-9.xlsx file and enter the MANAGE model outputs from the following cells and worksheets: cell B2 in the Table A4.1 worksheet for real GDP at market prices (annual growth rate by scenario); cell E3 in the Table A4.7 worksheet for sectoral value added (annual deviation from the baseline); cell A1 in the Table A4.8 worksheet for demand components (deviation from the baseline); and cell B2 in the Table A4.9 worksheet for real GDP at market prices (deviation from the baseline). The final Tables will automatically be generated in the same sheets.

For Tables A4.2–A4.6, open the Table A4.2-A4.6.xlsx file and enter the cumulative deviations from the business-as-usual (BAU) scenario generated by the MANAGE model. The reported variables are employment by skill level and sector, wages by skill level and sector, and the consumer price index. The workbook contains six worksheets corresponding to the scenarios analyzed: (a) Mono cumul_chx: market reforms in monopolistic sectors. (b) Trade cumul_chx: trade liberalization. (c) Dom cumul_chx: preference for domestic procurement. (d) Mon_Trđ cumul_chx: combined reforms in monopolistic sectors and trade liberalization. (e) Allcumul_chx: combined implementation of all reforms. (f) allfdicumul_chx: combined implementation of all reforms and an increase in foreign direct investment (FDI). Enter each scenario's outputs in the corresponding worksheet. The final tables will be generated automatically in the Mono cumul_chx worksheet.

8. Code Description

00_Housekeeping.do — Installs required packages (winsor2, labmask), sets the working-directory globals (by username), clears memory, fixes the random-number generator/seed, forces single-processor execution, and defines the folder-structure globals used by every other script. Included (via include) at the top of every numbered script and inside Main_script.do.

01_Clean_UZB_Microdata_HBS_2024.do — Reads the raw HBS 2024 modules (roster, weights, education, employment, non-wage income, social benefits, migration, food/non-food consumption, poverty analysis file) and merges them into a single cleaned individual/household-level dataset -> Cleaned_UZB_2024_HBS.dta.

02_Clean_Wittgenstein_UZB_Population_Projection_data.do — Reads and cleans the Wittgenstein Centre population-projection CSV (wcde_data.csv) into age-group/sex population targets by year -> Cleaned_Wittgenstein_Population_Projection.dta.

03_Read_and_clean_CGE_results.do — Reads the CGE macro targets (produced by Alhassane Camara) from Microsim_inputs_240326.xlsx (one sheet per policy scenario: Mono, Trade, Dom, Mono_Trade, All, All_FDI) and builds macro target datasets used downstream: employment by skill/sector, wages by skill/sector, prices, and income/GDP targets.

04b_Reweight_microdata_to_match_population_projection.do — Reweights the cleaned HBS microdata so that its demographic structure (age/sex cells) matches the Wittgenstein population projections for the target year(s) -> Reweighted_scaling_microsurvey.dta.

05_Estimate_Employment_Propensity_and_Sector_Affinity_and_Wage_Coefficient.do — Estimates, on the reweighted microdata: (i) a logit model of employment probability (hirability) as a function of age, skill, sex, province; (ii) a multinomial logit of sector affinity

(affinity); and (iii) a Mincer wage equation (mincer: ln_wage on experience, experience², skill, sex, province, sector). Coefficients/predictions feed into steps 06-08.

06_Simulate_employment_status.do / 06a_Function_Simulate_employment_status.do —

Uses the estimated employment-propensity scores, rescaled to match macro employment-by-skill targets, to stochastically assign simulated employment status to each individual (rank-based assignment on the scaled probability score, with a deterministic per-year seed) -> one dataset per scenario, Simulated_employment_microsurvey_scenario_.dta.

07_Sectoral_reallocation.do / 07b_Function_Sectoral_reallocation.do —

Classifies simulated employment transitions (stayers/EE, separations/EU, hires/UE) and reallocates workers across sectors to match macro employment-by-skill-sector targets ->

Simulated_sectoral_allocation_scenario_.dta.

08_Assign_wages_to_simulated_employment.do / 08a_Function_Wage_assignment.do —

Predicts a wage for each simulated worker from the Mincer equation plus a stochastic residual (drawn from the regression's error distribution), then rescales wages within skill-sector cells to match macro wage targets -> Simulated_Wage_Assignment_scenario_.dta.

09_Deflate_income_and_adjust_nonlabor_income.do /

09a_Function_Deflate_income_and_adjust_nonlabor_income.do — Scales non-labor income to match macro income/GDP targets and deflates total income using food/non-food price indices from the CGE macro price targets, producing real income ->

Simulated_Adjusted_income_scenario_.dta.

10b_Distributional_Income_Analysis_percentage.do — Main distributional/poverty analysis script, run once per poverty line (npl, vul, 8_3, looped in Main_script.do). Produces most of the numbered figures (Fig5-Fig17, FigA51-FigA82) and the exported poverty/income table.

10c_Distributional_Income_Analysis_Full_reform_effect_only.do — Focuses on the "All"/full-reform scenario only; produces wage-by-skill/gender/sector and poverty-by-location figures (Fig9-Fig12, Fig14-15).

10d_Distributional_Income_Analysis_FDI_included.do — Compares the "All" vs. "All_FDI" scenarios; produces FigA82 and related poverty statistics under the three poverty lines.

Main_script.do — The one-button entry point. Sets the housekeeping path and username, runs scripts 01-09 in order (quietly), then loops the distributional analysis (10b, 10c) over the three poverty lines and runs 10d once, ending with a datasignature check for stability verification.