Distributional Analysis Using Microsimulations in Stata

Ercio A. Munoz

The World Bank (Poverty and Equity GP)

August 2022

Based on joint work with Paul Corral and Israel Osorio-Rodarte

Motivation

- What is the distributional effect of given macroeconomic policies or shocks? Ex-ante analysis often requires the combination of macro and micro modeling.

- Three prominent areas where this is the case are the study of the effects of trade reforms, the impact of financial crisis, and climate change mitigation policies.

- There is an extensive literature about these type of macro-micro models, which typically considers linking a Computable General Equilibrium (CGE) model to a microsimulation model. However, there are not readily available ways of implementing any of these models in Stata.

This presentation

- Briefly provide an overview of macro-micro modeling and microsimulations.

- Introduce a Stata command (ms_reweight) that implements a reweighting-based microsimulation that can be linked to a standard CGE model for top-down modeling.

- Show the command in use with an empirical example using a household survey from Chile.

OVERVIEW

Three options to include household heterogeneity into the modeling of long-term macroeconomic scenarios

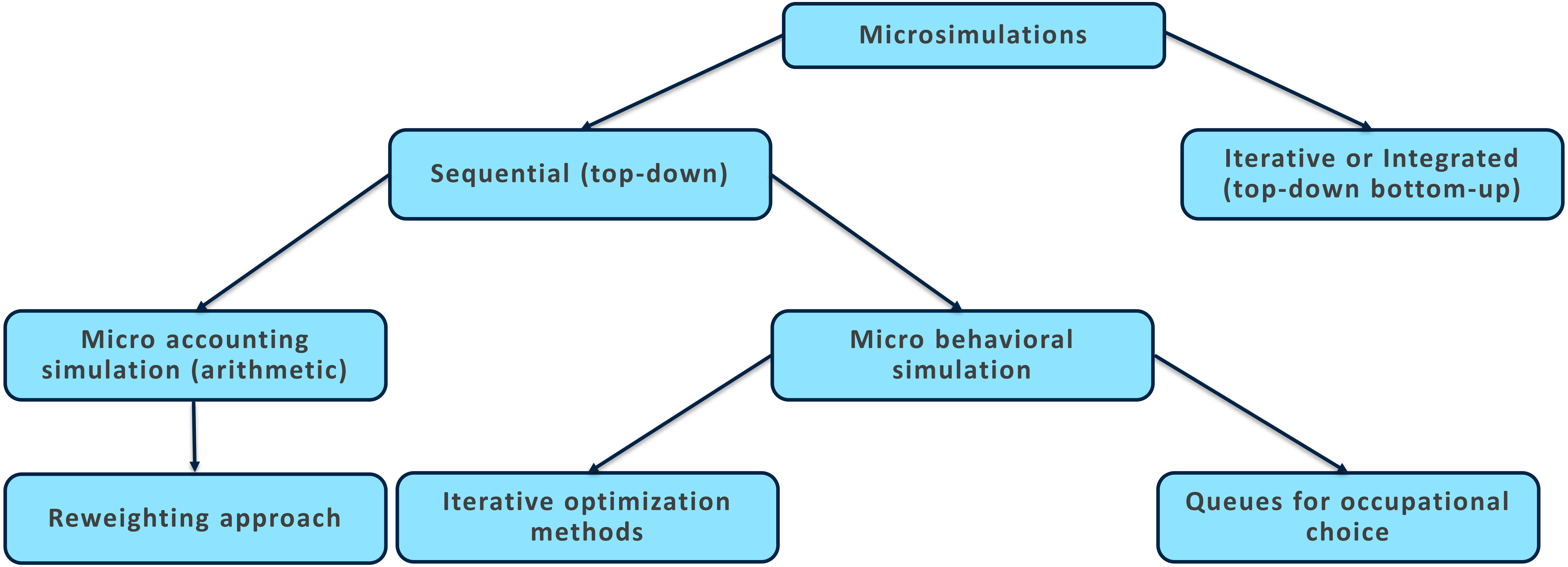
Explicit modeling of multiple household types within the CGE framework

Microsimulation of a large number of household types

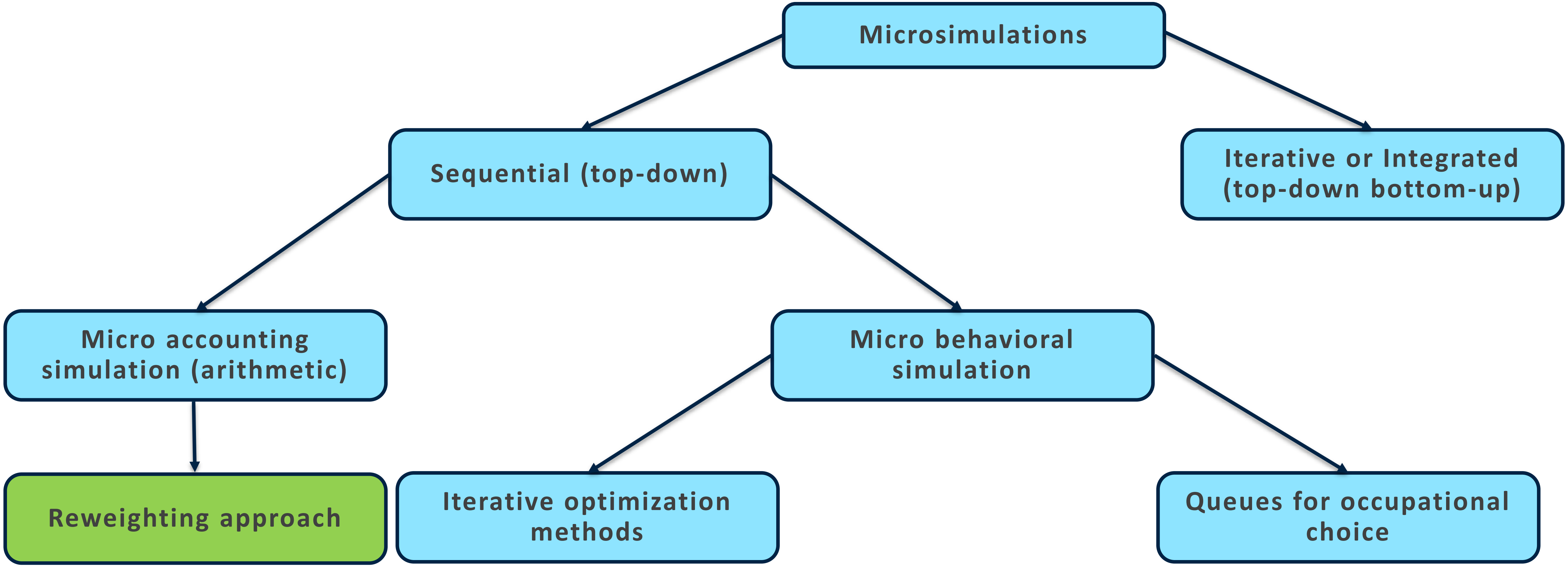
Direct modeling of the income distribution

See van Ruijven et al. (2015) for a review and assessment of these three options in the context of long-term climate change research.

Microsimulations: A diagram of (major) alternative approaches



This talk is about one approach within the sequential modeling



Sequential modeling (CGE-MS)

“Top” level: Macroeconomic model (i.e., CGE model)

↓

Linking aggregate variables (LAVs):

- Employment by sector and/or skill
- Wage by sector and/or skill
- Aggregate income growth
- Price changes (e.g., food vs. non-food)

↓

“Bottom” level: Household survey

REWEIGHTING-BASED MICROSIMULATION APPROACH

Some examples of this approach being applied in practice

- **Ferreira and Horridge (2006)** study potential distributional effects of the Doha round of trade negotiations in Brazil. They use an approach called “quantum weights method”.

- **Hérault (2010)** links a CGE model to a behavioral MS and one based on reweighting to compare their performance analyzing the effect of trade liberalization in South Africa.

- **Buddelmeyer, Hérault, Kalb, and van Zijl de Jong (2012)** links this type of MS model to a CGE model to assess the effects of climate-change mitigation policies in Australia from 2005 to 2030.

- **Vandyck and Van Regemorter (2014)** analyze distributional effects of increased oil excises in Belgium using a CGE linked to a reweighting-based MS.

- **Montaud, Pecastaing, and Tankari (2017)** study the effect of a possible deterioriation of weather conditions on Niger’s agriculture.

Reweighting-based microsimulation (CGE-MS)

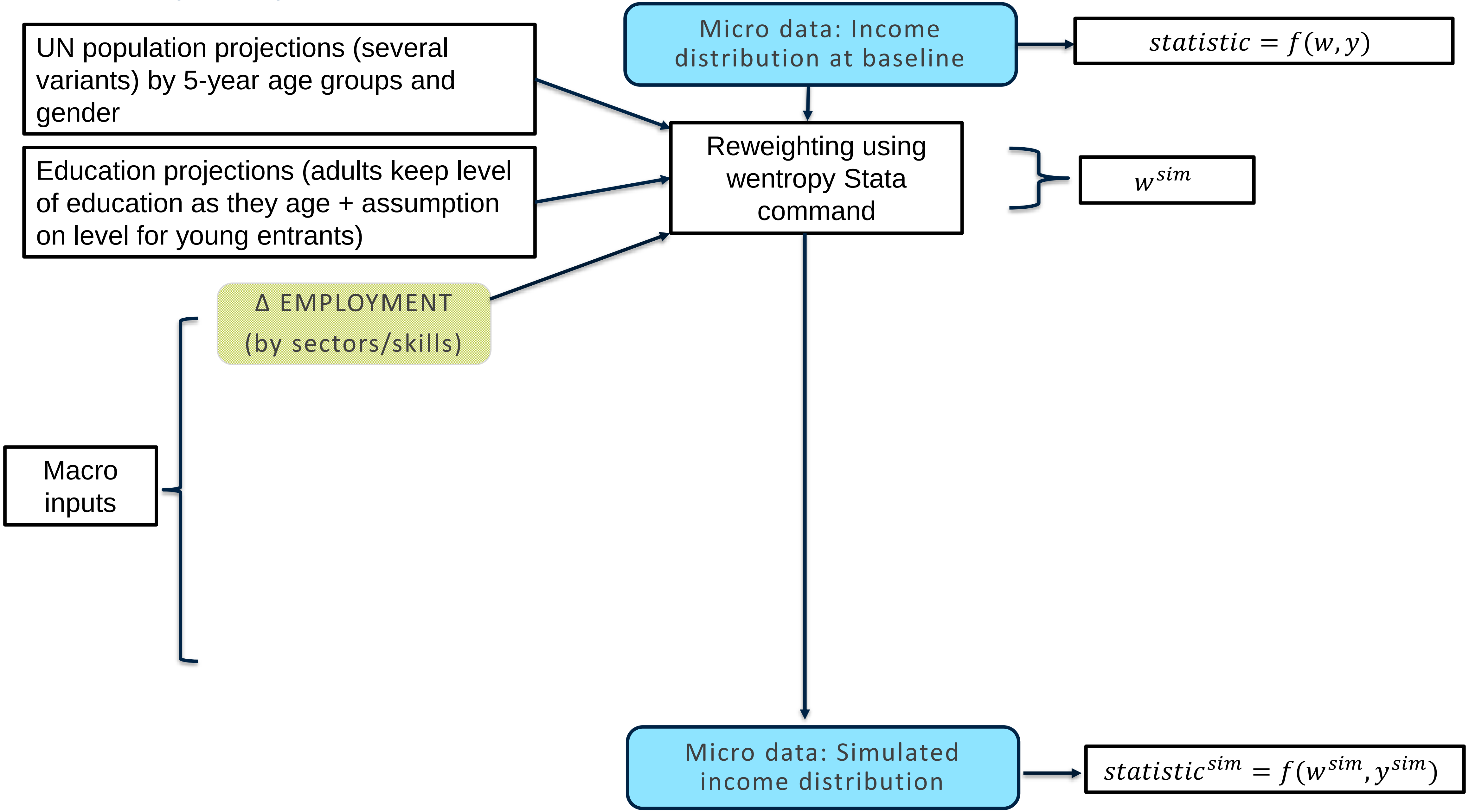
Micro data: Income distribution at baseline

$statistic = f(w, y)$

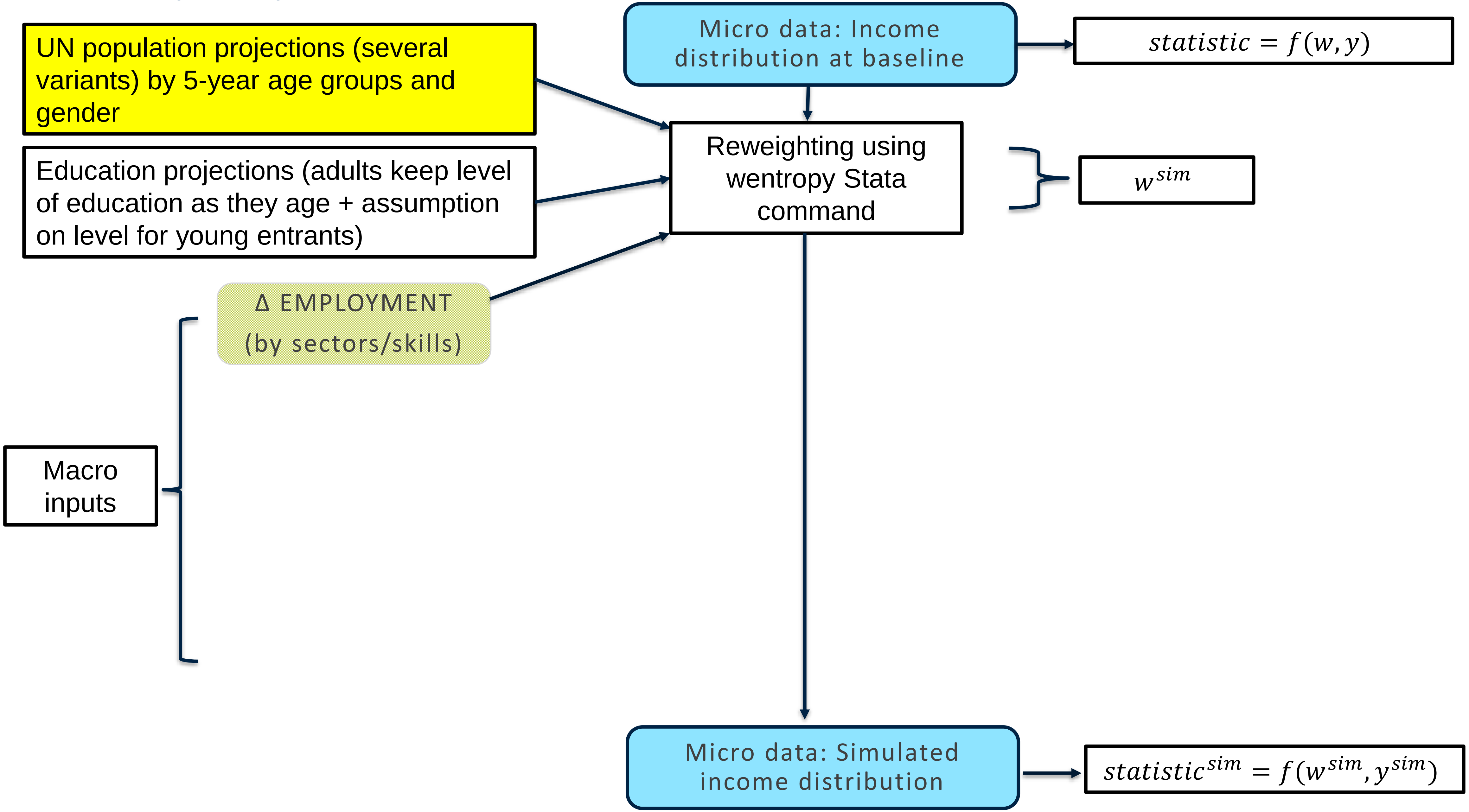
Micro data: Simulated income distribution

$statistic^{sim} = f(w^{sim}, y^{sim})$

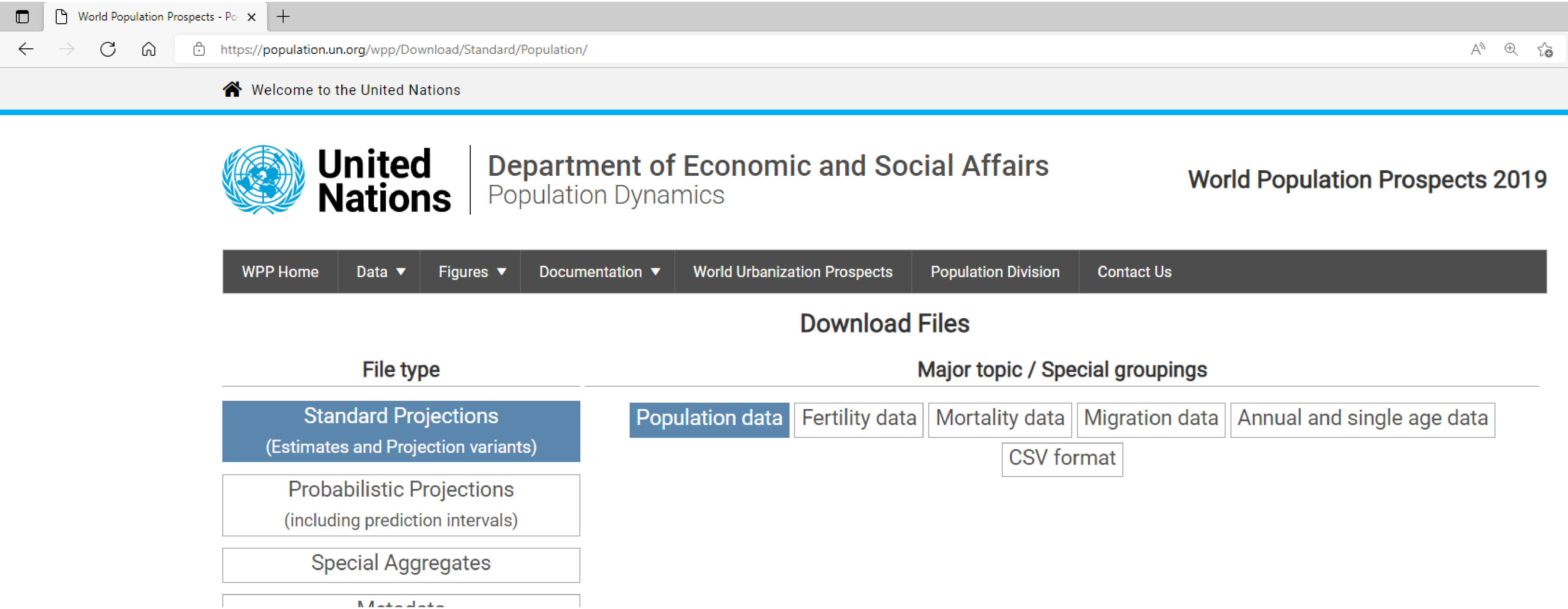
Reweighting-based microsimulation (CGE-MS)



Reweighting-based microsimulation (CGE-MS)

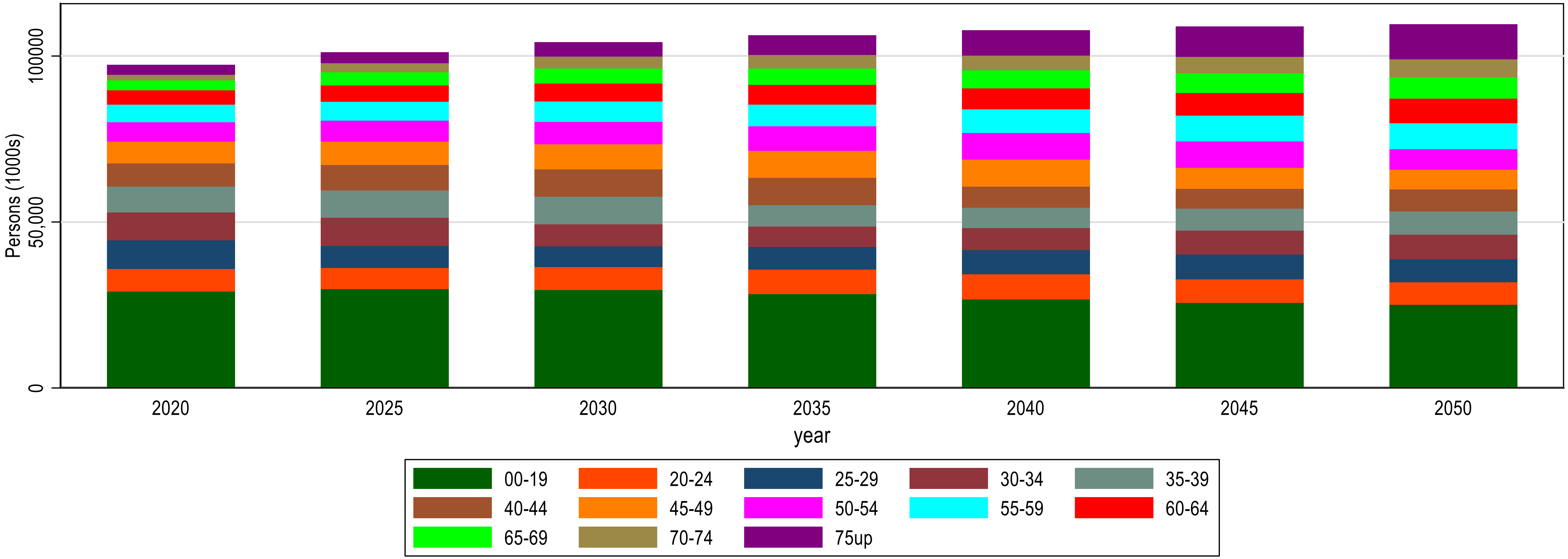


We can use data from UN population prospects:

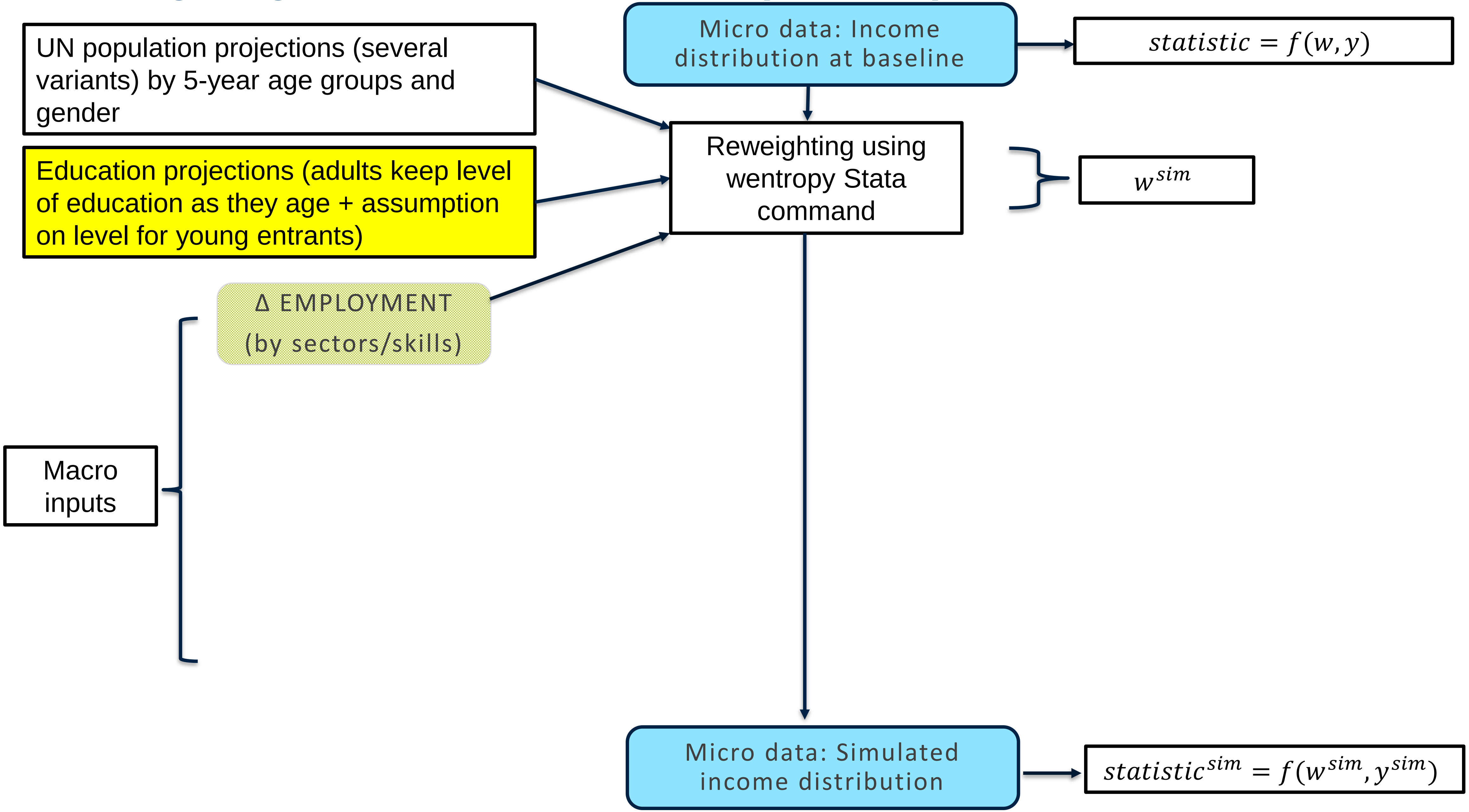


Population projections:

The idea is to accommodate the survey for countries with important projected demographic changes.



Reweighting-based microsimulation (CGE-MS)



Weights need to match population by age/gender/education

We need an assumption regarding the level of education of the new population.

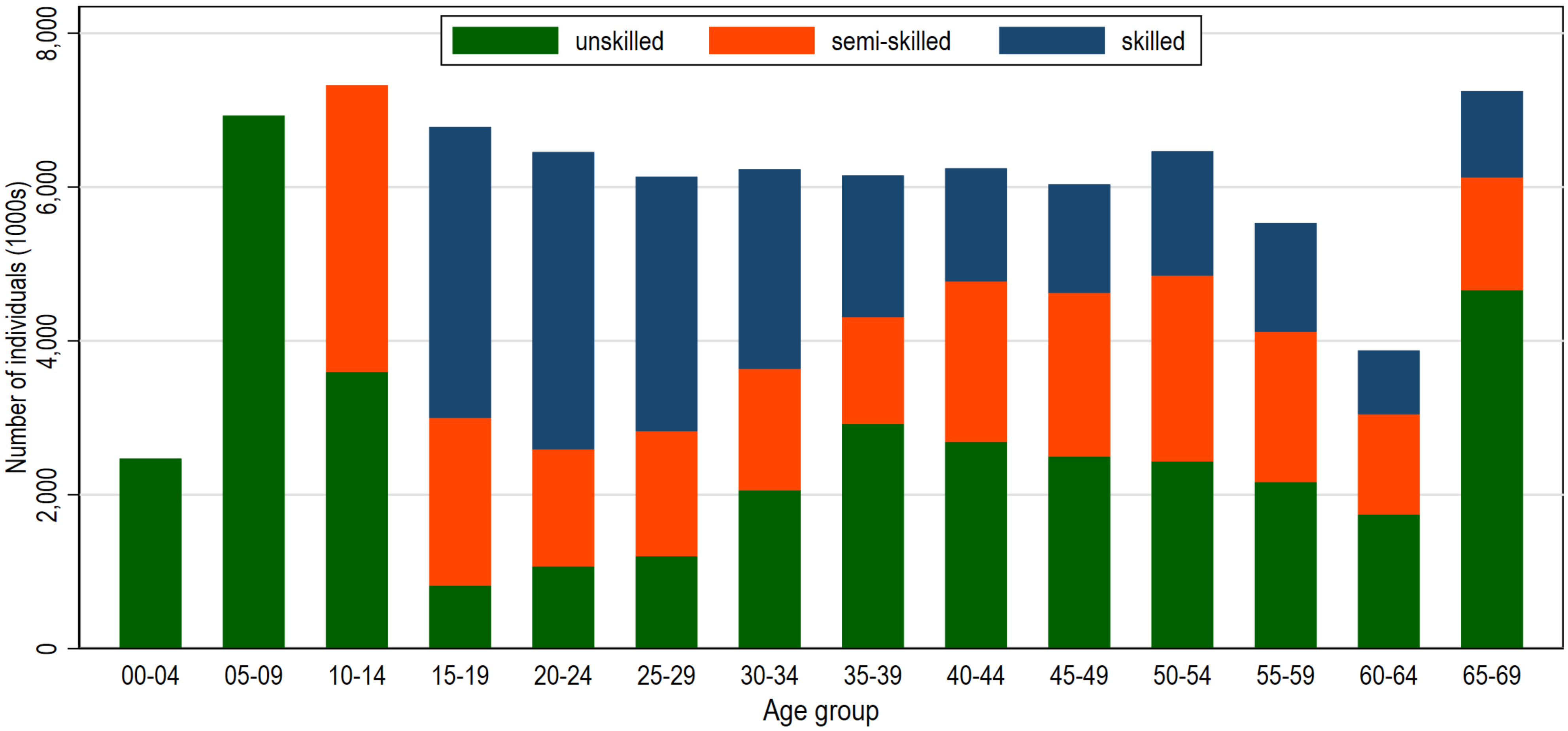
We assume that the new young cohort does not improve its education vis-à-vis the young cohort of the initial year (it keeps within-cohort school entrance and graduate rates constant over time).

	Skilled	Unskilled	Skilled	Unskilled
2000				
Young	60	40	P_1	P_2
Old	30	70	P_3	P_4
2030				
Young	60	40	$\hat{P}_1$	$\hat{P}_2$
Old	60	40	$\hat{P}_3$	$\hat{P}_4$

Source: Bourguignon and Bussolo (2013)

Weights need to match population by age/education

We use 00-29 as young cohort and the educational levels observed in the survey to create projections for the future.



Reweighting-based microsimulation (CGE-MS)

UN population projections (several variants) by 5-year age groups and gender

Education projections (adults keep level of education as they age + assumption on level for young entrants)

Δ EMPLOYMENT

(by sectors/skills)

Macro inputs

Micro data: Income distribution at baseline

$statistic = f(w, y)$

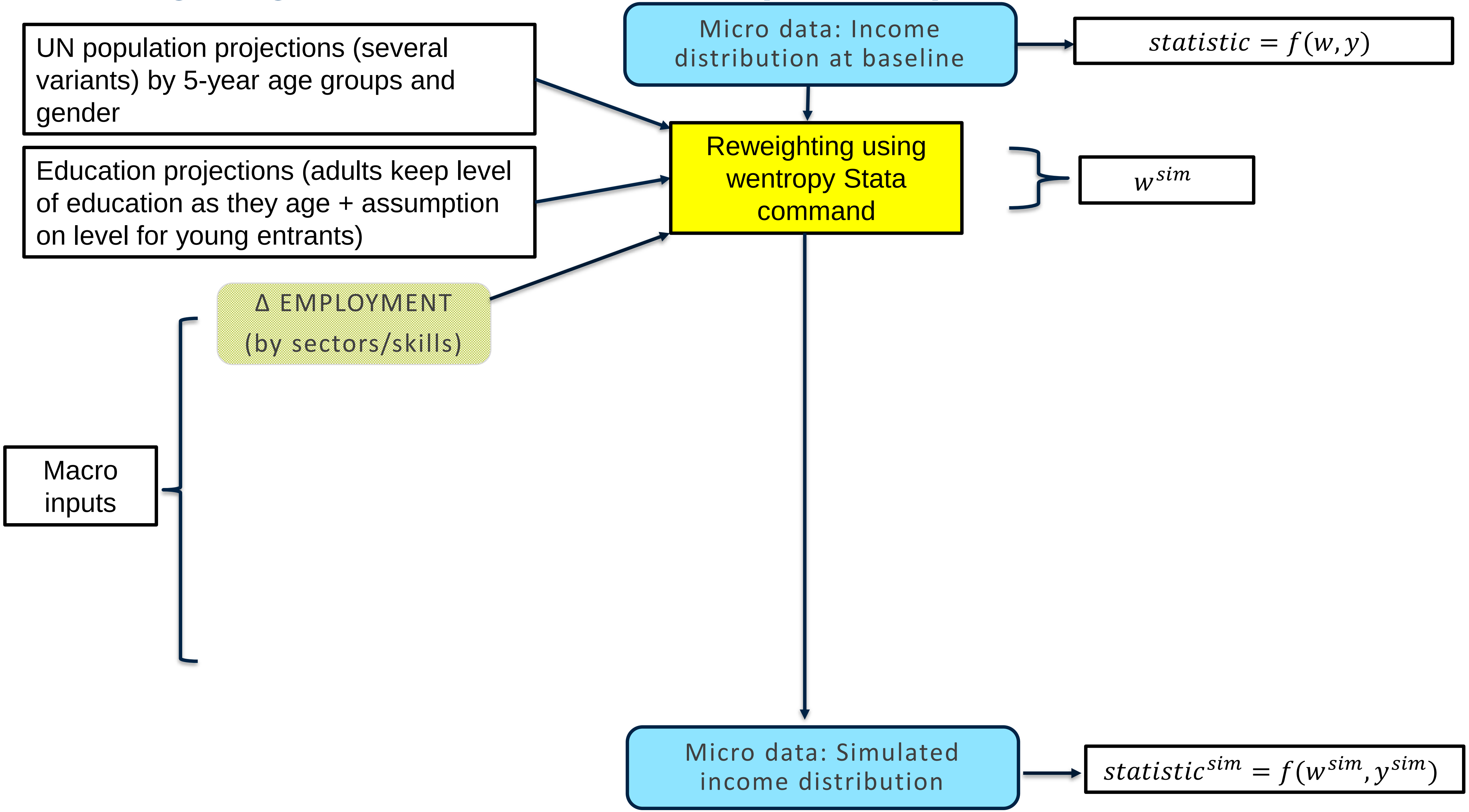
Reweighting using wentropy Stata command

w^{sim}

Micro data: Simulated income distribution

$statistic^{sim} = f(w^{sim}, y^{sim})$

Reweighting-based microsimulation (CGE-MS)



Recalibration of the weights to match population by gender, age, education, and sector

After we have all the population totals, the reweighting process can use maxentropy command in Stata or an improved version written by Paul Corral and Rodrigo Salcedo called wentropy.

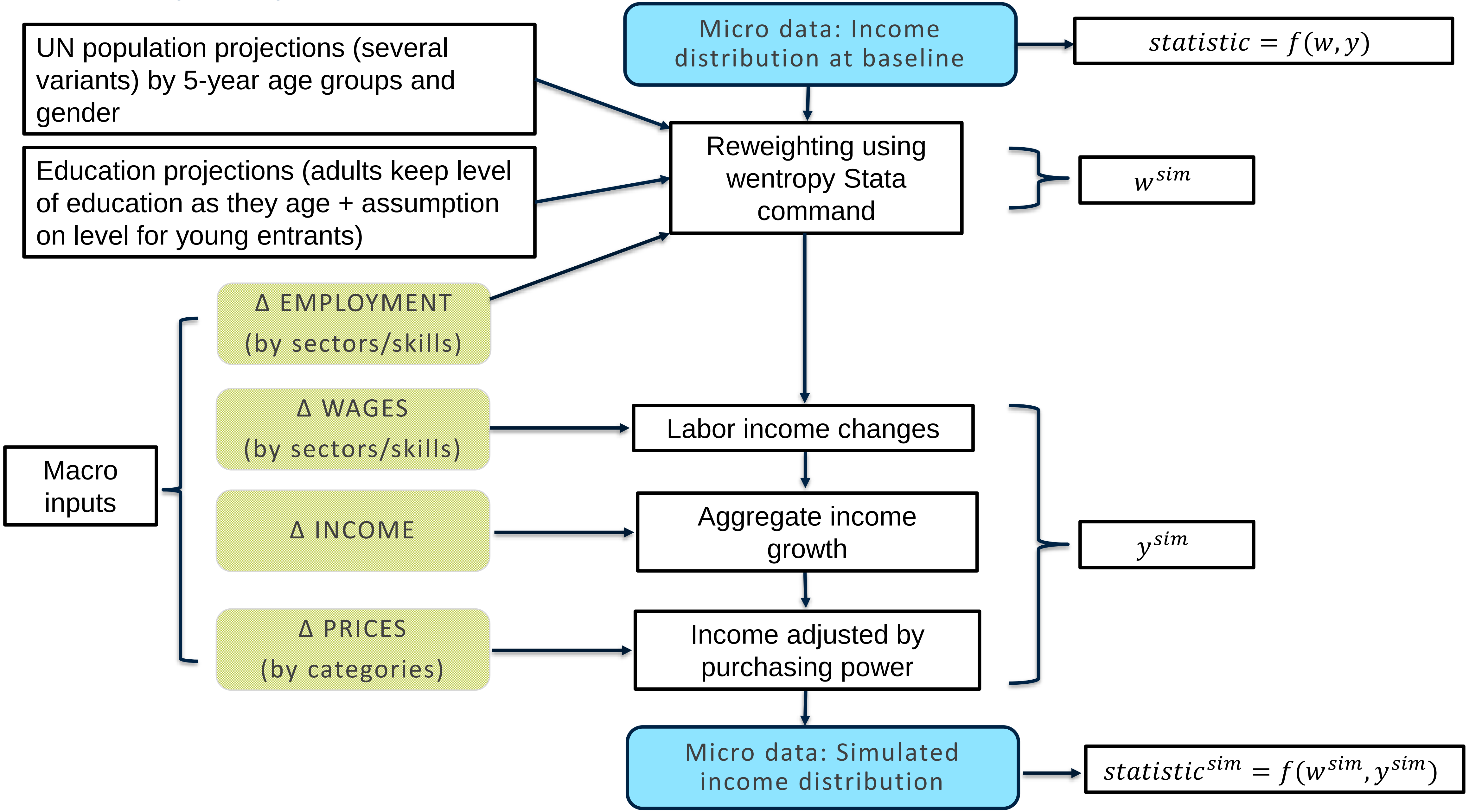
The Stata Journal (2010)
10, Number 3, pp. 315–330

An introduction to maximum entropy and minimum cross-entropy estimation using Stata

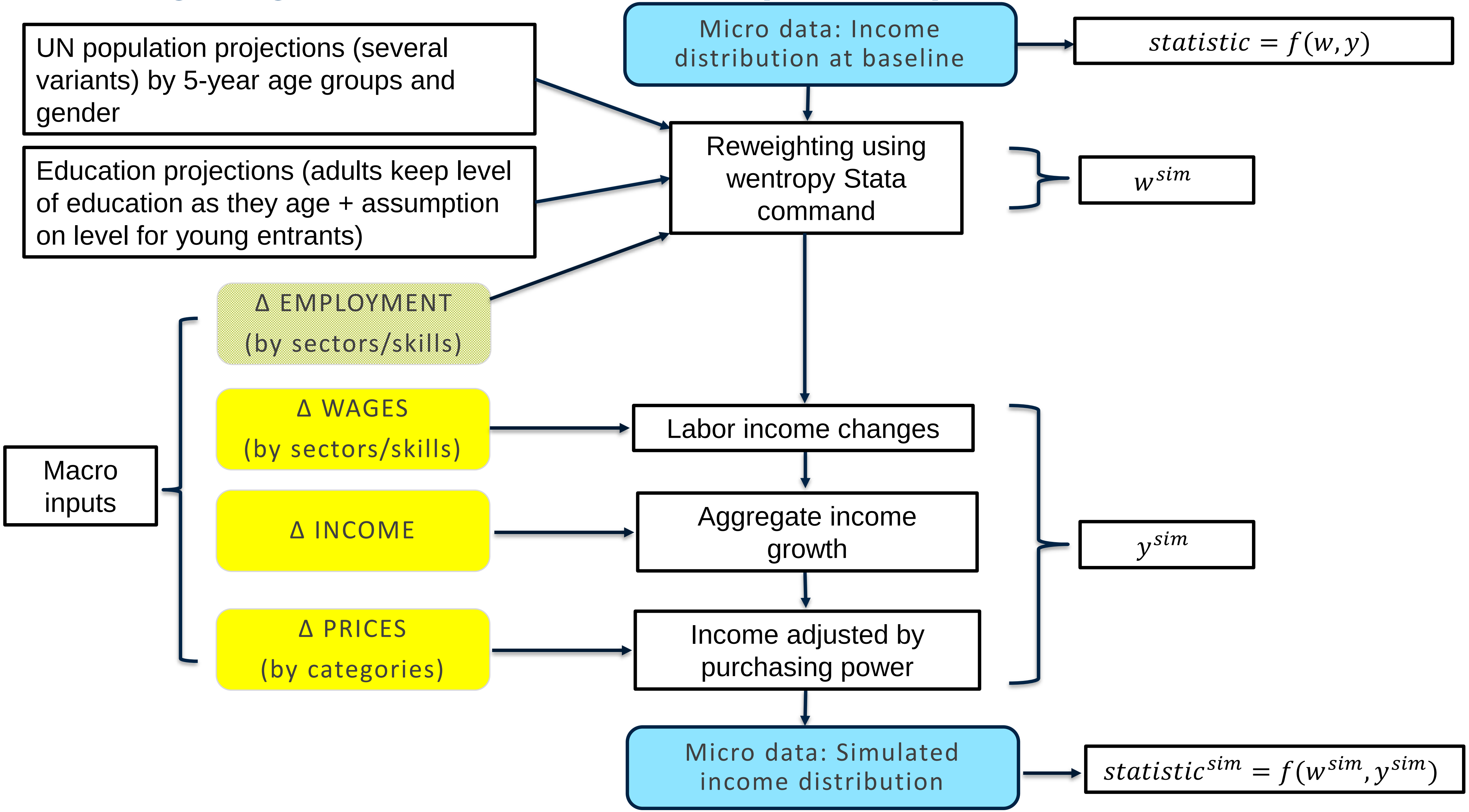
Martin Wittenberg
University of Cape Town
School of Economics
Cape Town, South Africa
Martin.Wittenberg@uct.ac.za

Abstract. Maximum entropy and minimum cross-entropy estimation are applicable when faced with ill-posed estimation problems. I introduce a Stata command that estimates a probability distribution using a maximum entropy or minimum cross-entropy criterion. I show how this command can be used to calibrate survey data to various population totals.

Reweighting-based microsimulation (CGE-MS)

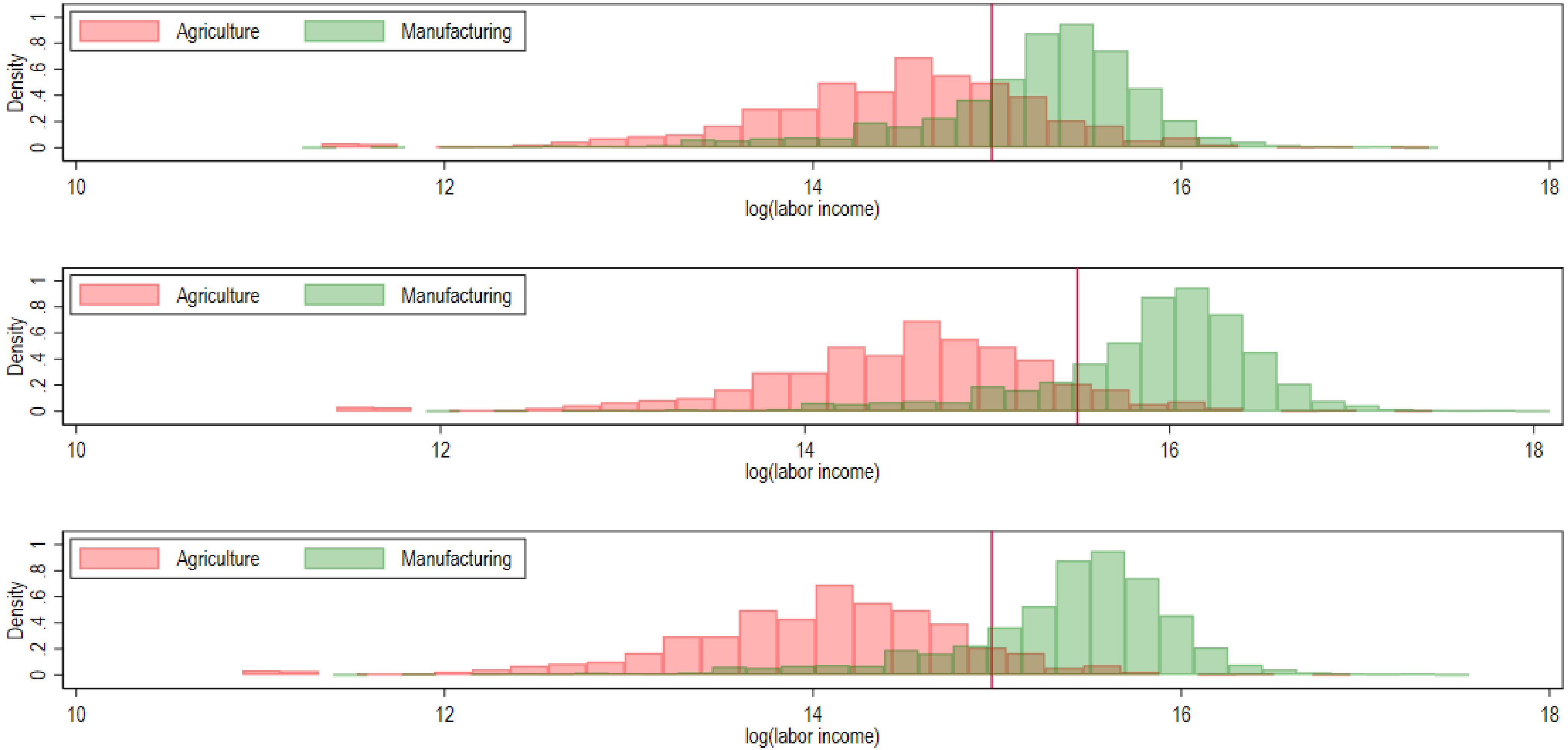


Reweighting-based microsimulation (CGE-MS)



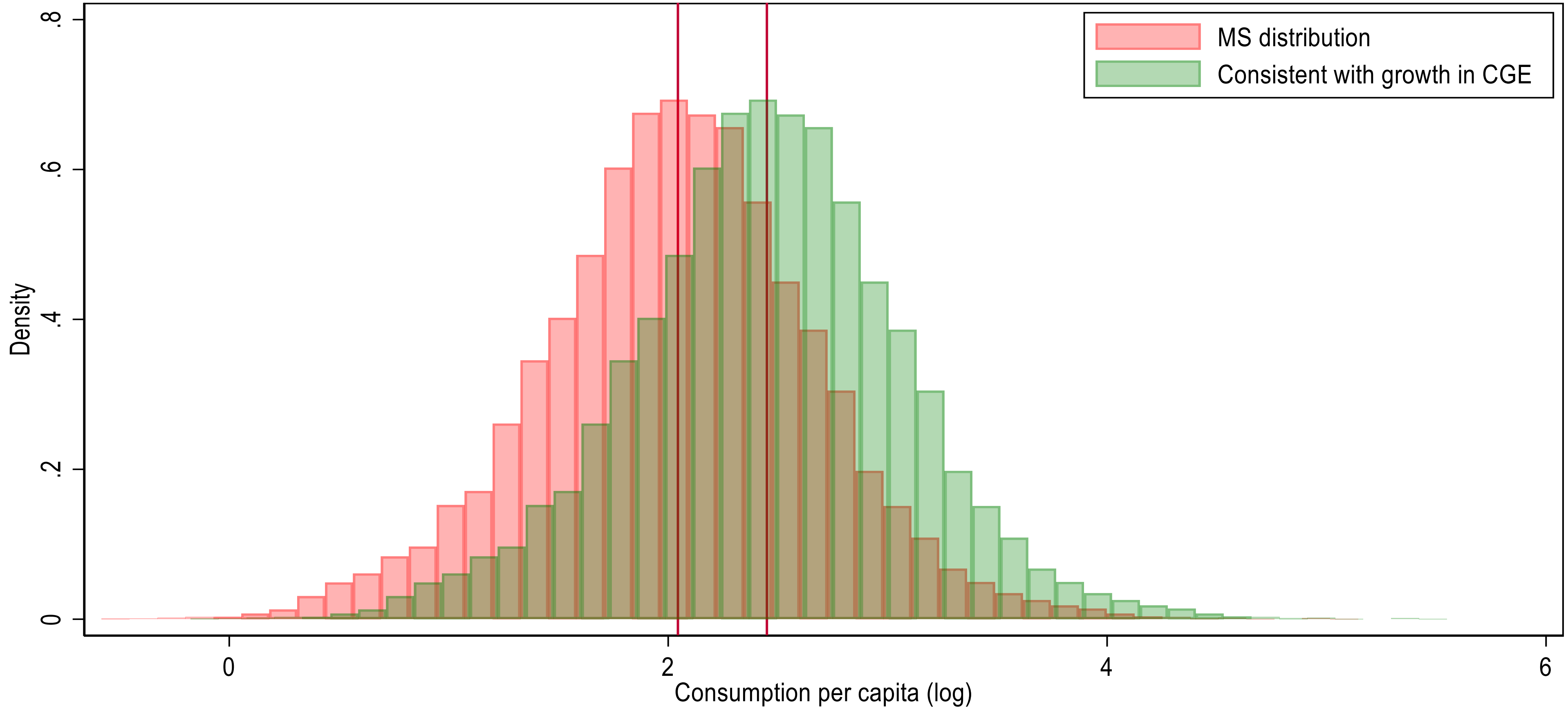
Change in wages by sectors and skills

We apply these changes by sector/skill to modify the distribution and then re-center the entire distribution back to the original mean. For example:

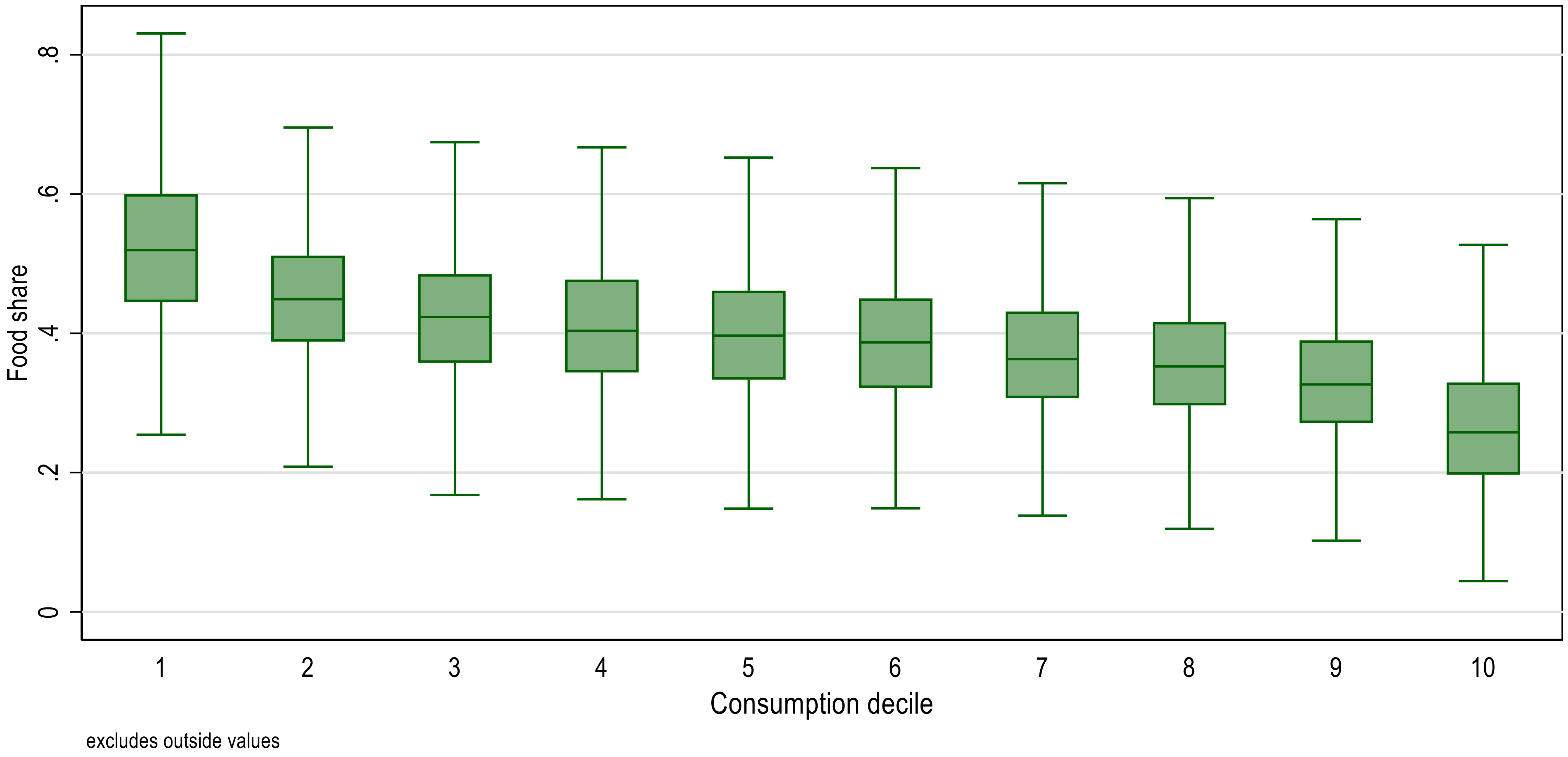


Growth in aggregate income requires just a shift

For example, we use the aggregate growth rate in income/consumption per capita from the CGE to shift the entire distribution of income/consumption from its original mean to a new mean consistent with it.



Food share by consumption decile

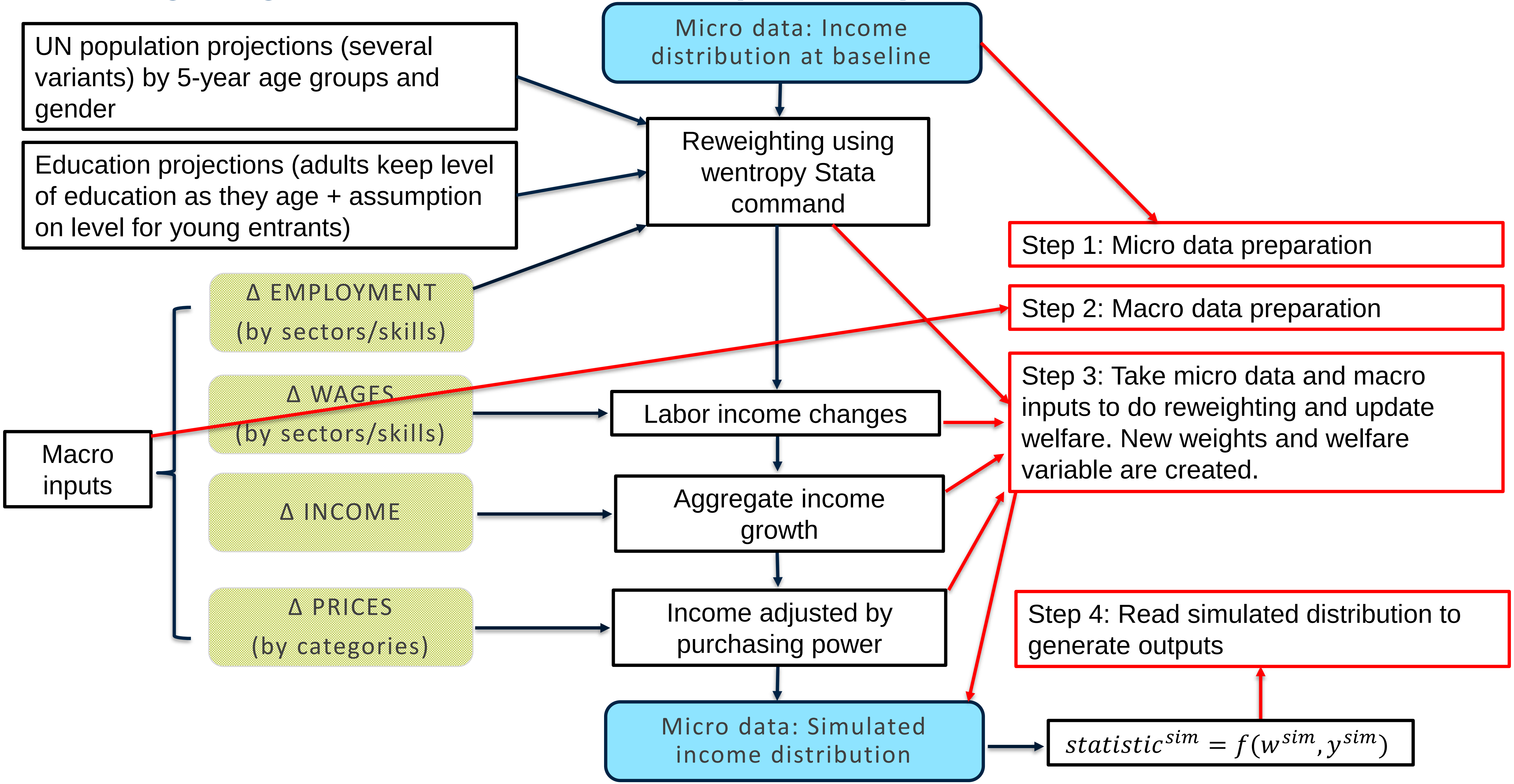


Impact of prices (food and non-food)

- We use food prices and non-food prices from the CGE to modify the distribution in the micro data. We create the weighted (using food share) change in prices by household and adjust income/consumption per capita by it.

- After we apply these changes, we re-center the distribution to the average consumption per capita obtained from the previous step.

Reweighting-based microsimulation (CGE-MS) in Stata



SYNTAX AND EMPIRICAL EXAMPLE

Syntax of the command:

ms_reweight, age(*varname*) education(*varname*) gender(*varname*) hhsize(*varname*)
hid(*varname*) iweights(*varname*) country(*varname*) iyear(*varname*) tyear(*varname*)
generate(*varname*) match(*varname*) popdata(string) variant(string) [pid(*varname*)
skill(*varname*) industry(*varname*) industryshares(*matrix*) targets(*matrix*) growth(string)
laborincome(*varname*) simlaborincome(*newvarname*) foodprices(*varname*)
foodshares(*varname*)]

Empirical example:

Preparing macroeconomic outputs:

```
.
. mat gen_edu_age_shares = ///
>  .0687158 \ .0201254 \ .0012683 \ .0024893 \ .0055047 \ .0094925 \ .0129582 \ .020409 \ ///
>  0 \ .0000199 \ .0316699 \ .0059426 \ .0090615 \ .0153785 \ .0192458 \ .0154524 \ ///
>  0 \ .0147977 \ .023008 \ .0755668 \ .053372 \ .051573 \ .0386678 \ .0194587 \ ///
>  .0122997 \ .0691335 \ .0220542 \ .0016508 \ .0024284 \ .0052396 \ .0068572 \ .0090823 \ ///
>  0 \ .0127094 \ .0000378 \ .0336497 \ .006714 \ .0083865 \ .0132019 \ .0156242 \ ///
>  0 \ .0125731 \ .0106471 \ .0221617 \ .071369 \ .0472987 \ .0431383 \ .033415

.
. mat growth_laborincome = 1922.653 , 1379.8 \ 4071.278 , 2105.882 \ 2320.895 , 1399.573 \ ///
>  3405.491 , 3915.768 \ 2480.845 , 1430.888 \ 2120.925 , 1242.761 \ 2812.062 , 1473.955 \ ///
>  3669.261 , 1442.874

.
. mat sectoral_targets = .0338954 , .0015263 \ .0106397 , .0001518 \ .0443509 , .0005247 \ ///
>  .0022176 , 4.12e-06 \ .0369428 , .0007995 \.0904014 , .0011197 \ .0309018 , .0002764 \ ///
>  .1374343 , .0011761

.
```

Empirical example:

Running the microsimulation:

```
.
. use Example_1998.dta,clear

.
. ms_reweight, age(age) edu(calif) gender(gender) hhsize(hsize) hid(hhid) iw(weight) ///
>  iyear(1998) tyear(2013) generate(wgtsim) match(HH) ///
>  country("CHL") popdata("Population_Example_1998") variant("Medium") ///
>  industry(industry) industryshares(sectoral_targets) skill(skilled) ///
>  targets(gen_edu_age_shares) ///
>  laborincome(labor_income) simlaborincome(sim_labor_income) growth(growth_laborincome)
Wentropy for country CHL in year 2013
The constraint matrix is

const2013[64,1]
          constraints
fc1a0010     .0687158
fc1c1120     .0201254
fc1a2130     .0012683
fc1a3140     .0024893
fc1a4150     .0055047
fc1a5160     .0094925
fc1a6170     .0129582
 fc1a70+      .020409
fc2a0010            0
fc2c1120     .0000199
```

Empirical example:

```
.
. ta industry [w=wgtsim] ,m
(frequency weights assumed)
```

Industry	Freq.	Percent
Agriculture and fishing	600,053	3.57
Mining and quarrying	181,708	1.08
Manufacturing	758,381	4.51
Electricity, gas and water	38,127	0.23
Construction	639,691	3.80
Commerce	1,552,931	9.23
Transport, storage and communication	528,743	3.14
Services and other	2,343,545	13.93
.	10,178,252	60.51
Total	16,821,431	100.00

```
. ta industry [w=weight] ,m
(frequency weights assumed)
```

Industry	Freq.	Percent
Agriculture and fishing	637,647	4.38
Mining and quarrying	80,381	0.55
Manufacturing	691,408	4.75
Electricity, gas and water	41,859	0.29
Construction	413,427	2.84
Commerce	952,476	6.55
Transport, storage and communication	387,306	2.66
Services and other	1,792,814	12.32
.	9,551,486	65.65
Total	14,548,804	100.00

```
.
. sgini labor_income [w=wgtsim]
(frequency weights assumed)

Gini coefficient for labor_income
```

Variable	v=2
labor_income	0.6039

```
. sgini labor_income [w=weight]
(frequency weights assumed)

Gini coefficient for labor_income
```

Variable	v=2
labor_income	0.5588

Thank you!