

Global inequality user guide

Lucas Chancel

Amory Gethin

December 2017



GLOBAL INEQUALITY USER GUIDE

Lucas Chancel, Amory Gethin

Abstract

This user guide to detailed computed codes is meant to help researchers reproduce global inequality estimates available in the 2018 World Inequality Report using the “wid-global-inequality.stpr” Stata project. Interested readers should first read “The Elephant curve of global inequality and growth” (WID.world Working Paper 2017/20) and “Constructing a global income distribution brick by brick” (WID.world Technical Note 2017/5) before reading the present document.

Contents

1	Getting directly to the main data and results	2
2	General information	2
2.1	Installations	2
2.2	Main folders	2
2.3	Results folders	3
3	Preparation	4
3.1	Runwir program	4
3.2	Master file	4
3.3	Running R codes from Stata.....	5
3.3.1	Running R codes with rsource.....	5
3.3.2	Running R codes manually from R.....	5
3.3.3	Merging distributions from WID.world/Gpinter	6
3.3.4	Note on running time and geographic repartitions.....	6
4	Running the global inequality project step by step	6
4.1	Estimating global inequality	6

4.2	Reporting the main results.....	7
4.3	Decomposing within and between inequality.....	8
4.4	Comparing inequality in different regions and scenarios.....	8
4.5	Predicting inequality up to 2050.....	8

1 GETTING DIRECTLY TO THE MAIN DATA AND RESULTS

Researchers or journalists who are mainly willing to use the global inequality distributions produced by the global inequality project can directly go to:

Data/Main results

Regions-distributions.xlsx contains all the income inequality distributions of world regions from 1980 to 2016 in Purchasing Power Parity euros.¹

World-distributions.xlsx contains all global income inequality distributions from 1980 to 2016 for all scenarios tested in the global inequality project. The final global distribution used in the 2018 World Inequality Report, which combines all world regions in purchasing power parity, is available in the “PPP_4_Static” sheet.

Other folders contain the graphs produces in the PPP Scenario 4 (Static calibration) scenario. More information on these graphs in available in 2.3.

2 GENERAL INFORMATION

2.1 INSTALLATIONS

Running global inequality estimates requires Stata 14 and R version 3.4.1 or higher.

In Stata, the user needs to install the *rsource*, *dropmiss* and *renvars* commands² which are used all along the project. This can be done by running *ssc install [command]*.

In R, the user needs to install the *foreach*, *doParallel*, *parallel*, *xlsx*, *plyr*, *gdata* and *gpinter* packages. The *gpinter* package is available from <https://github.com/thomasblanchet/gpinter> or by running:

¹ The aggregation method used for this file is our baseline method: purchasing power parity, scenario 4, static calibration. For more information, see technical note.

² There are potentially other SSC commands used in the global inequality project. If at some point, Stata declares not recognizing a command, please try running *ssc install [command]*.

```
install.packages("devtools")
```

```
devtools::install_github("thomasblanchet/gpinter")
```

2.2 MAIN FOLDERS

The global inequality project contains the *wid-global-inequality.stpr* Stata project, all do-file used in this project on the “Do” folder, as well as 6 other main folders:

“Data” contains the raw data used all along the project to estimate global inequality.

- “Imported” contains the main WID database (“wid-final.dta”, produced on September 4th 2017) as well as subsets of the database, which are produced at the beginning of the project.
- “Added” contains data which are not directly included in the WID database at the time of the project. These include distributional national accounts for Africa (*africa.dta*), France (*france.dta*) and the Middle East (*middle-east.dta*), growth projections from the OECD (*projections.dta*) and country isocodes (*country-codes.csv*). This folder also includes templates used to produce formatted tables and figures during the project (*tablesgrowth.xlsx*, *tablesgrowth-detailed.xlsx*, “Geographic repartition templates”)

“Gpinter input” contains temporary folders used to export distributional data to excel and import them into R in order to merge them using Gpinter.

Subfolder names correspond to: “AS” for “Other Asia”, “AS_all” for “Asia”, “EU” for “Other Western Europe”, “EU_all” for “Europe”, “NA” for “Other North America”, “NA_all” for “US-Canada”, “SA” for “Other Latin America”, “WO” for “World”, “Between-countries equality” for the estimation of inequality within countries, “Within-countries equality” for the estimation of inequality between countries.

“Gpinter output” contains temporary folders used to export the merged distributions from R, and import them back into Stata. Subfolder names correspond to the same codes as above.

“R” contains the R codes or do-files necessary to merge distributions. “R codes” contains the same codes in .R format. The main folder contains a set of do-files used to merge subregional or regional distributions. These do-files are simply a copied version of the .R files which use *rsource* in order to run R codes from Stata. As Stata globals for user directories cannot be used with *rsource*, do-files are specific to one computer, which is why they are duplicated for the two main computers used by the WID.world team for this project (“Richard” and “Amory”). See 2.3 for more information.

“Results” contains the main results of the project, in market exchange rate and purchasing power parity for the five available scenarios and for the two calibration methods used (see below).

“ReadMe” contains this note as well as the *Global Inequality Technical Note* which provides detailed technical (rather than practical) information on the global inequality project.

2.3 RESULTS FOLDERS

Subfolders in the "Results" folder are all combinations of three parameters (see 2.2 for more detail):

- Purchasing Power Parity (PPP) / Market Exchange Rate (MER).
- Scenarios 1 to 5.
- Static / Dynamic calibration.

PPP/Combined contains figures comparing the evolution of inequality in different aggregates (in aggregates corresponding to scenarios 1 to 4).

In each of the results folders (in *PPP/Scenario 4 (Static Calibration)* which is our baseline scenario, for instance), subfolders are:

- *Figures and Tables*, which contains figures and tables summarizing the main results:
 - o *Geographic repartition* contains figures which provide information on the level and evolution across time of the geographic repartition of global income groups (non-empty for scenario 4 only).
 - o *Growth dynamics* contains global growth incidence curves and other figures providing information on growth dynamics across the world.
 - o *Growth tables* contains tables which show how different world regions grew between 1980 and 2016 and how inequality evolved in these regions.
 - o *Incomes shares* contains figures showing the evolution of global income shares across the world in different specifications (see 3.4 and technical note).
- *Gpinter input* contains all the distributional data used to produce the distribution of global income in this scenario. All files are formatted so that they can be used on WID.world/gpinter.
- *Gpinter output* contains all the merged distributions produced during the process of estimating global inequality. Folders names correspond to those explained in 2.1.
- *Projections* contains the raw distributional data used to predict inequality between 2016 and 2050 in three scenarios. BAU2016 correspond to own-country path, EU8016 corresponds to Europe's path, US8015 correspond to the United States' path (see 3.5).

The .dta files contained in the main folder correspond to temporary files that are used to produce the main results of the global inequality project. *gpinterized.dta* is the main results file, containing full distributional data for all world regions and for the world as a whole.

3 PREPARATION

3.1 RUNWIR PROGRAM

The first step for running the global inequality project is to open the *wid-global-inequality.stpr* project (located in the “Do” folder) in Stata and opening the *runwir-program.do* file. Then, the user should add his directory and R path in line 49 in the same way as it is done for the three other computers of the WID lab:

- The “user” global sets the user’s path.
- The “r_dir” global sets the path for Stata to run R using *rsource*. The user should look for a file named R.exe and copy the path to that file. For Windows users, the path is typically “C:\Program Files\R\R-3.4.1\bin\i386”.

You may then run *runwir-program.do*. This creates the *runwir* command, which allows you to define the mains parameters of the global inequality estimation.

3.2 MASTER FILE

After having set these two globals and ran *runwir-program.do*, you may open the *master.do* file, which enables you to define the method used to estimate global inequality using the *runwir* command. The *runwir* command takes three main options (see technical note):

- *ppp*: estimate global inequality using purchasing power parity estimates (1) or using market exchange rate estimates (0).
- *scenario*: combination of subregions used, from 1 to 5.
- *calibration*: infer inequality in unknown subregions using static calibration (0) or dynamic calibration (1).

The default scenario used to estimate global inequality is purchasing power parity, scenario 4 and static calibration, which locates results into the “*Results/PPP/Scenario 4 (Static calibration)*” folder and corresponds to:

runwir, ppp(1) scenario(4) calibration(0)

From the master file, you may simply define this folder by running the master up to line 20. If you want to let your computer loop over scenarios for specific results, however, you may choose combinations of parameters from the *ppp_mer*, *scenarios* and *calibrations* globals, and then run the whole master file until line 100. To compute global distributions, run lines 31 to 53; to produce figures and tables for these combinations, run lines 55 to 76; to compute projections of global inequality, run lines 78 to 98.

3.3 RUNNING R CODES FROM STATA

3.3.1 RUNNING R CODES WITH RSOURCE

In theory, the whole project can be fully reproduced using only the *master* file, which runs step by step all the do-files necessary to merge distributions. Yet, merging distributions requires Stata to run R codes, and the *rsource* command cannot use the path specified by the user in the *runwir-program* file.

To run R codes from Stata, the user has to make the following adjustments:

- In the “R” folder (or directly from the “99 – R” folder of the *wid-global-inequality.stpr* project), open all the do-files starting by “r-richard” and replace the existing path by your own path in *setwd()*. The user path should look like “C:/Users/.../Global inequality”. For “r-richard-subregions.do”, the path has to be changed twice, lines 11 and 62.
- In *1b - gpinterize-subregions.do*, line 132, change “C:/Users/r.clarke” to your path (“user” global).
- In *1e - gpinterize-regions.do*, line 149, change “C:/Users/r.clarke” to your path.
- In *1g - gpinterize-world.do*, line 157, change “C:/Users/r.clarke” to your path.
- In *4a - gpinterize-between-equality.do*, line 183, change “C:/Users/r.clarke” to your path.
- In *6d - gpinterize-projections.do*, line 93, change “C:/Users/r.clarke” to your path.

This should lead Stata to run all the do-files starting by “r-richard” from your computer, and it should enable R to recognize your path and merge distributions in the correct folders.

The *rsource* command should open the terminal and perform the commands specified in the corresponding do-file. Unfortunately, the command skips all calculations in case of mistake and does not inform the user if the program failed. All do-files calling R should take at least 5 minutes to run. If the terminal shuts down quickly (after less than one minute), it is a clear indication that *rsource* failed.

Unfortunately, *rsource* is very sensitive and may not work correctly on some computers. Another solution for merging distributions is presented below.

3.3.2 RUNNING R CODES MANUALLY FROM R

A second option for merging distributions is to run manually Stata until R is prompted, then run .R files from R, and finally come back to Stata. This process is more time-consuming but unfortunately there is no other straightforward method to reproduce the global inequality project without using *rsource*.

First, the user should open all .R files from the “R/R-codes” folder in his or her R software (such as [Rstudio](#)). All paths specified in *setwd()* should then be changed in all files, exactly as in 2.3.1.

Then, the user should alternate between Stata and R in the following way:

- Run Stata up to the “Gpinterize” section of *1b - gpinterize-subregions.do*, run *asia.R* and *europa.R* from R, and run Stata again starting from the “Import back” section.
- Stop Stata at the “Gpinterize” section of *1e - gpinterize-regions.do*, run *as_all.R*, *europa_all.R*, *na_all.R* and *sa_all.R* from R, and run Stata again starting from the “Import back” section.
- Stop Stata at the “Gpinterize” section of *1g - gpinterize-world.do*, run *world.R* from R, and run Stata again starting from the “Import back” section.
- Stop Stata at the “Gpinterize” section of *4a - gpinterize-between-equality.do*, run *world-between.R* from R, and run Stata again starting from the “Import back” section.
- Stop Stata at the “Gpinterize” section of *6d - gpinterize-projections.do*, run *projections.R* from R, and run Stata again starting from the “Import back” section.

3.3.3 MERGING DISTRIBUTIONS FROM WID.WORLD/GPINTER

Last but not least, the distribution can be merged online from wid.world/gpinter. The .xlsx files produced by Stata are formatted so that they can be directly uploaded on the application. The user can then use the “merge” option to reproduce the results of the global inequality project. Reproducing the whole project by this method step by step, however, is very time-consuming.

3.3.4 NOTE ON RUNNING TIME AND GEOGRAPHIC REPARTITIONS

Merging distributions takes between fifteen minutes (subregions) and one hour (world).

For merging all subregional distributions into a global distribution of income (using *r-richard-world.do* or *world.R*), the user has the possibility to include the computation of geographic repartitions, which provide information on the geographic composition of global income groups. For each global income group, the program returns the contribution of each subregion to this group (30% of Top 1% income earners may be European, for instance) and saves the result in “*percentiles-match-[year].xlsx*” files.

Geographic repartitions computations increase considerably the time needed to run the global inequality project (*world.R* with geographic repartitions can take over 4 hours to run), and are only meaningful for scenario 4. By default, geographic repartitions are deactivated from R and Stata programs, but can be activated by removing “#” signs from *world.R* (lines 71 to 75) or *r-richard-world.do* (lines 73 to 77). More specifically, the following lines of code are used:

```
thrmatch<-sapply(countries,function(x) mapply(match_perc,x,percentiles))
row.names(thrmatch)<-percentiles
colnames(thrmatch)<-countries
write.xlsx(thrmatch,paste(outputdir,"percentiles-match-",y,".xlsx",sep=""), row.names=T)
```


4 RUNNING THE GLOBAL INEQUALITY PROJECT STEP BY STEP

Below, the do-files used in the global inequality project are described in the order in which they should be used.

4.1 ESTIMATING GLOBAL INEQUALITY

0a – import-indices.do extracts information on PPP and exchange rates conversion factors, populations and national incomes from the WID database (*wid-final.dta*).

0b – import-regions.do uses *country-codes.csv* to define the composition of regions used in the project.

0c – aggregate-regions.do combines regions compositions, adult populations and national incomes to compute the adult populations and average national per adult of all subregions used in this project.

0d – import-shares-growth.do, which is later on used for global inequality projections along with *6b – import-shares-growth-BAU.do*, imports the share of growth captured by income group in several scenarios (France 1980-2015 and US 1962-1980 are finally not used).

1a – prepare-data.do imports distributional data from WID, adds countries not available from the database (Africa, France, Middle East), and adjusts distributions to obtain full distributional accounts for all countries or subregions from 1980 to 2016.

1b – gpinterize-subregions.do creates “Other Western Europe” by merging France, Germany and the UK, and “Other Asia” by merging India and China.

1c – create-other-subregions.do creates “Eastern Europe”, “Other North America” and “Other Latin America” by duplicating the distributions of “Other Western Europe”, the US and Brazil respectively.

1d – calibrate-subregions.do calibrates “Other Asia”, “Eastern Europe”, “Other North America”, “Other Latin America” and “Other Western Europe” by rescaling these distributions to the national income per adult of the corresponding aggregate (calibration = 1) or by predicting the evolution of inequality in these regions from the share of growth captured by income group and the evolution of the national income per adult (calibration =0).

1e – gpinterize-regions.do creates distributions for Europe, US-Canada, Latin America and Asia as a whole by merging subregions. Europe is the merged distribution of France, Germany, the UK, Other Western Europe and Eastern Europe in scenarios 4 and 5, or the merged distribution of France, Germany and the UK in scenarios 1, 2 and 3. US-Canada is the merged distribution of the US and Other North America. Latin America is the merged distribution of Brazil and Other Latin America. Asia is the merged distribution of China, India and Other Asia.

1f – calibrate-regions.do computes the average national income per adult of regions by aggregating national incomes and adult populations, and rescales the distributions of regions to their corresponding average national income per adult.

1g – gpinterize-world.do merges subregional distributions to produce a global distribution of income.

4.2 REPORTING THE MAIN RESULTS

2a – world-gic.do creates three global growth incidence curves (scaled by population, scaled by share of growth captured and scaled income group) for the 1980-2016 period.

2a – world-shares-growth.do plots the cumulative share of growth captured by income group between 1980 and 2016.

2b – growth-tables-detailed.do (only relevant for scenario 4) exports tables providing information on the total income growth, the average annual income growth and the share of growth captured by income group for all regions, subregions and countries included in the global inequality project.

2b – growth-tables.do (only relevant for scenario 4) exports tables providing information on the total income growth, the average annual income growth and the share of growth captured by income group for all regions included in the global inequality project.

2c – geographic-continuum-subregions.do (only relevant for scenario 4) creates excel files containing graphs which show the geographic repartition of global income groups in 1990 and 2016.

2c – geographic-repartition.do (only relevant for scenario 4) creates excel files containing graphs which show the evolution of the geographic repartition of the Bottom 50%, Middle 40%, Top 10%, Top 1% and Top 0.1% global earners. It also produces figures (stacked bars) indicating the geographic repartition of global income in 1980 and 2016.

2c – populations-shares.do plots the populations of regions as a share of world population in 2016.

2c – share-world-covered.do creates a table reporting the share of world population covered by each global inequality scenario.

2d – world-shares.do plots the evolution of the Top 10%, the Top 1% and the Bottom 50% shares of global income from 1980 to 2016.

4.3 DECOMPOSING WITHIN AND BETWEEN INEQUALITY

3a – within-generate-distribution.do uses the average national income per adult of all countries around the world to estimate income inequality between countries (what would global income inequality be if income was equally distributed across individuals *within* countries). It performs a simple percentile analysis on these aggregates (weighed by adult population) and returns a “within-equality” distribution of global income for the 1980-2016 period.

3b – within-world-shares.do plots the evolution of the Top 10%, the Top 1% and the Bottom 50% shares of global income from 1980 to 2016 for the “within-equality” distribution.

4a – gpinterize-between-equality.do uses subregions created in *1d-calibrate-subregions.do* to estimate a “between-equality” distribution of global income from 1980 to 2016. It replaces the average national income per adult of each subregion by the average national income per adult of the world. It then rescales the average income of each income group to match the average income, and merges all subregions as in *1g-gpinterize-world.do*.

4b – between-world-shares.do plots the evolution of the Top 10%, the Top 1% and the Bottom 50% shares of global income from 1980 to 2016 for the “between-equality” distribution.

4.4 COMPARING INEQUALITY IN DIFFERENT REGIONS AND SCENARIOS

5a – compare-between-within.do plots the evolution of the Top 10%, the Top 1% and the Bottom 50% shares of global income from 1980 to 2016 for standard, “between-equality” and “within-equality” distributions combined.

5b – world-combine-results.do compares the results obtained in the four main scenarios used for global inequality estimates. It produces graphs comparing growth incidence curves, cumulative shares of growth captured by income group, and Top 10%, Top 1%, Bottom 50% and Middle 40% shares for these four scenarios.

5c – compare-regions-shares.do creates figures comparing the evolution of Top 10% and Bottom 50% income shares in all world regions between 1980 and 2016.

5d – compare-regions-incomes.do creates a figures comparing the evolution of the average national income per adult in all world regions between 1980 and 2016.

5e – top-shares-regions.do creates a bar chat comparing Top 10% and Top 1% income shares in Brazil, the Middle East, South Africa, the US and Western Europe for the latest year available.

4.5 PREDICTING INEQUALITY UP TO 2050

6a – import-projections.do imports data on GDP and population growth rates from 2016 to 2050 and uses it to predict the evolution of the average national income per adult for each country of the world between these two dates.

6b – import-shares-growth-BAU.do imports the share of growth captured by income group in each region between 1980 and 2016.

6c – project-regions.do combines these shares of growth with predicted average national incomes per adult and available distributional data to project income inequality up to 2050 in all world subregions.

6d – gpinterize-projections.do merges the resulting distributional data created for the 2016-2050 period in all subregions to create global distributions of income inequality for this period in three scenarios: if all countries follow the same inequality trend as they did between 1980 and 2016; if all countries follow the same inequality trend as Europe did between 1980 and 2016; and if all countries follow the same inequality trend as the US did between 1980 and 2016.

6e – plot-scenarios-shares.do compares the evolution of the resulting global income shares in the three scenarios between 2016 and 2050.

6f – plot-scenarios-averages.do compares the evolution of the resulting averages of global income groups in the three scenarios between 2016 and 2050.