PEA Stata Code Manual

I. Introduction

The new PEA 3.0 framework provides an integrated analysis of countries' progress in advancing on the updated World Bank mission statement: *ending extreme poverty and boosting shared prosperity on a livable planet*. Key to the PEA revamp is a standardization effort to help to streamline the analysis.¹

This manual provides step-by-step guidance to World Bank staff for using a newly developed PEA Stata package, which automates the production of trends in poverty, shared prosperity and inequality, as well as their drivers and profiles. The package generates a full set of tables (pea tables) and figures (pea figures) to aide the core analytics in PEAs (shown in Annex 1: Full list of tables and figures). A smaller set of tables and figures are for inclusion in a standardized data annex (pea core) to ensure consistency and comparability across countries.

All outputs are automatically created by running the three commands, with minimum manual inputs needed from users. Parts of the codes can nevertheless be tailored to county contexts, for instance by specifying country-specific education or industry groups. The codes facilitate standardization of core analysis across countries and reproducibility of results. This guide explains how to install, set up, and execute the codes in Stata by following the below steps:

- Step 1: Installing and setting up the pea Stata package
- Step 2: Preparing and harmonizing survey data
- Step 3: Create additional variables
- Step 4: Run the pea commands

II. Installation and Setup

Step 1: Install the pea Stata Package

Option 1: Downloading and setting up the package

The PEA package can be downloaded from <u>Github</u> (click on "Code" on the top right and "Download ZIP"). Proceed to store the unzipped file in your folder of choice. The file location needs to be specified in the code later on (see Annex 2: Example code (adopath +)).²

The package includes the three main commands (pea core, pea tables, pea figures) which produce one Excel output file each. Additional ado-files allow each table and figure to be created separately. The whole package needs to be downloaded in order to produce the full output. This is because some of the ado-files run calculations and prepare and download data in the background. Ensure internet access is enabled, as additional data files are downloaded automatically from pip or datalibweb.

A data sub-folder needs to be extracted from the package and stored in the correct location on your device. A few additional datasets are required to run the full package, which are included in the pea

¹ See the <u>PEA website</u> for resources and tools for PEAs.

² An integration to be downloaded directly from Stata for automatic installation is planned.

package.^{3,4} The sub-folder within the package can be found here: "/pea/Stata/personal/**pea**". Importantly, this **pea** sub-folder (including the Scorecard_Summary_Vision folder) needs to be moved into the personal folder of your Stata system directory (which can be found by entering "sysdir" in Stata): e.g. "c:/ado/personal/**pea**".

Option 2: Cloning the Github repository

Cloning the Github repository to your local repository has the advantage that any updates to the package can easily be merged into the cloned repository.

The data folder will still need to be moved to your personal Stata folder on your device (see option 1). Furthermore, the paths within the do-files will have to be adjusted according to your local repository (see Annex 2: Example code (adopath +)).

Step 2: Access and Prepare Survey Data

The pea Stata package runs on survey data harmonized according to the Global Monitoring Database (GMD).⁵ Proceed to one of the following next steps.

(2a) Using pre-harmonized GMD data

If the survey data is already included in the GMD, access it in Stata using the datalibweb command (see Step 2 in Annex 2: Example code).⁶ Using datalibweb has the advantage of retrieving multiple years of data easily and consistently. You may proceed to Step 3.

(2b) If survey data needs to be harmonized and added to GMD

If the latest survey is not already harmonized in the GMD, country teams should work with the regional stats team to add the latest survey data into the database. If using multiple rounds of data for analysis, the new data needs to be appended to previous datasets (which can be accessed through datalibweb) before running the pea command, to include trends and changes in the tables and figures.

³ The following are included in the data folder: CLASS.dta, which includes country classifications, updated in July every year (https://github.com/GPID-WB/Class/tree/master/OutputData); exposure_vulnerability_2021.dta, which contains data needed to calculate exposure and vulnerability to climate-related hazards (accessed from the reproducibility package of the Poverty, Prosperity and Planet Report 2024); UNESCO.dta, which is used in the calculation for the multidimensional poverty measure (accessed through datalibweb); Scorecard_Summary_Vision, which contains excel files for each corporate scorecard indicator (accessed from World Bank Scorecard website.)

⁴ The package also installs several additional Stata commands. The following Stata packages are installed automatically when running the pea code for the first time (and don't need to be installed separately): apoverty, ineqdeco, svylorenz, fastgini, glcurve, alorenz, povdeco, fs, groupfunction, drdecomp, adecomp, pip, schemepack, colorpalette, geoplot, palettes, colrspace, moremata. The command sedecomposition cannot be installed through Stata's SSC and is therefore included in the pea package (pea/Stata/plus/s). The sedecomposition package has been accessed from here: https://github.com/vavalomi/stata_tools/blob/master/sedecomposition/sedecomposition.ado ⁵ For more information, visit the GMD website.

⁶ If this is your first time accessing datalibweb, visit https://datalibweb2.worldbank.org/home to create a token to access datalibweb through Stata. See here for datalibwebguidelines, and the D4G website for more instructions.

Step 3: Create additional variables and check data

Once all surveys have been GMD-harmonized (step 2), only the poverty lines, PPP-adjusted welfare aggregates and national welfare aggregates need to be created manually. The specific variables that need to be created are PPP-adjusted and constant welfare aggregates, the welfare aggregate that is underlying the national poverty rates, and the used poverty lines. In the example codes these variables are called: welfareppp, pline215, pline365, pline685, natwelfare, natline (and natline2—in case that there is a second national poverty line). Step 3 in Annex 2: Example code provides an example code to construct these additional variables. The PEA package also includes example do-files that can be used as templates to run the codes ("pea/Stata/Example").

If users want tables to show standard-errors, the data needs to be 'survey set' in Stata, before running the code. In the example in Annex 2: Example code, the line "svyset ..." in Step 3 survey sets the data accordingly.

The pea commands have an option "setting (GMD)", which creates an additional set of variables. If the option setting (GMD) is not specified, variables "married" and "nowork" need to be created. Table 1 displays code to create these, and the other additional variables (such as "welfppp").

Table 1 lists all variables that are needed to run the pea commands in Stata. It is critically important to check and confirm that all variables in the GMD-harmonized dataset take on the correct variable type, value labels, and only the allowed values. Specifically, all variables should be numeric and variable labels and value labels need to be defined. This is to ensure proper presentation and minimal post-editing in tables. For example, the value label of the subnational region variable should be in the format in which region names should be displayed in the table output (Annex 2: Example code shows an example in Step 3).

The type, value labels and values of all variables need to be consistent across all surveys. This should be carefully checked, in particular when appending new survey data to pre-harmonized GMD data. Household identifiers might be repeated between surveys over time, even though it is not panel data. In this case, please create year-specific identifiers.

Note that even when not all variables are in the dataset or correctly defined, the pea commands will still run and produce all tables and figures where these variables are not required.

Table 1: List of variables used by the PEA command

Description	Name	Values	Code	
1) Variables tha	1) Variables that need to be created			
Welfare aggregate (divided by CPI and ICP, and 365 days)	welfppp	integer	gen welfppp = welfare/cpi2017/icp2017/365 (note: welfare should be in percapita terms)	

⁷ Importantly, verify that the correct deflators are used.

⁸ This is only relevant for Table 1, and Table A.1.

Motionalwalfara	naturalfara	intogor	con naturalfara - V Nata: The
National welfare	natwelfare	integer	gen natwelfare = X Note: The
aggregate			construction of the national welfare
			aggregate will depend on the
			country context.
National poverty	natline	integer	gen natline = X if year==
line (LCU)			
Second national	natline2	integer	gen natline2 = X if year==
poverty line (LCU)		_	
Poverty line: \$2.15	pline215	2.15	gen pline215 = 2.15
per day (PPP)			Sem painted and
Poverty line: \$3.65	pline365	3.65	gen pline365 = 3.65
per day (PPP)	punesos	3.00	gen punesos – 5.05
	nlinoCOF	C OF	don plinoCOE – C OE
Poverty line: \$6.85	pline685	6.85	gen pline685 = 6.85
per day (PPP)			
Subnational ID.	subnatvar	integer (labeled)	e.g. encode subnatid,
Make sure the			gen(subnatvar)
variable is numeric			
and harmonized			
over all years.			
Comparability of	comparability_peb	"Yes" or "No"	This variable defines comparability
national welfare			of national poverty for Figure C.1,
variable			following the same methodology as
			used for PEB reporting.
If forecast data is ad	ded to core Figure 1:		acca for i Eb reperang.
Years with forecast	year_fcast	integer	These variables need to be
data	yeai_icasi	IIILEGEI	
41414		tuta da u	appended to the survey data to
Forecast of	natpov_fcast	integer	show forecasts in Figure C.1.
national poverty			
rates			
Forecast of GDP	gdp_fcast	integer	
per capita			
·		led if setting(GMD) is not	specified (see next section for more
details on th	e code)		
Household head	head	0=Not household	gen head = relationharm==1 if
dummy		head; 1=Household	relationharm~=.
		head	
Married dummy	married	0=not married;	gen married = marital==1 if
,		1=married	marital~=.
Not working	nowork	0=working; 1=not	gen nowork = lstatus==2
dummy		working	lstatus==3 if lstatus~=.
	e needed to obtain th		
Household	hhid	•	
	IIIIIU	integer	
identifier	m: al	:mtowou	
Individual identifier	pid	integer	
Year of survey	year	integer (4 digits)	
Poverty specific	weight_p	integer	
weights			

Welfare aggregate	welfare	integer	
(LCU) Relation to household head	relationharm	1=Household head; 2=Spouse; 3=Child; 4=Parents; 5=Other relative; 6=Non- relative	
Age of individual	age	integer	
Sex of individual	male	0=female; 1=male	
Household size	hsize	integer	
Marital status of individual	marital	1=married; 2=never married; 3=living together; 4=divorced/separated; 5=widowed	
Highest level of education completed (4 categories)	educat4	1=No education; 2=Primary (complete or incomplete); 3=Secondary (complete or incomplete); 4=Tertiary (complete or incomplete)	
Urban indicator	urban	1=urban; 2=rural	
Currently in school dummy	school	0=not currently in school; 1=currently in school	
Labor force Status	lstatus	1=employed; 2=unemployed; 3=not in labor force	
Sector/industry of employment (4 categories) [NOTE: For some countries this is industrycat4_year]	industrycat4	1=Agriculture; 2=Industry; 3=Services; 4=Other	
Sector/industry of employment (10 categories) [NOTE: For some countries this is industrycat10_year]	Industrycat10	1=Agriculture; 2=Industry; 3=Services; 4=Other	
Employment status	empstat	1=Paid employee; 2=Non-paid employee; 3=Employer; 4=Self- employed	

Improved water access	imp_wat_rec	0=No; 1=Yes	
Improved sanitation facility	imp_san_rec	0=No; 1=Yes	
Access to electricity	electricity	0=No; 1=Yes	
Ownership of a television	tv	0=No; 1=Yes	
Ownership of a car	car	0=No; 1=Yes	
Ownership of a cellphone	cellphone	0=No; 1=Yes	
Ownership of a computer	computer	0=No; 1=Yes	
Ownership of a fridge	fridge	0=No; 1=Yes	
Comparability between survey rounds	Comparability	Integer	

Step 4: Run pea commands

Once the pea package, data and additional variables have been prepared, users can follow the example in Annex 2: Example code under Step 4 to run the commands. The PEA package includes example do-files that can be used as templates to run the codes ("pea/Stata/Example"). Note that the prepared data and the adopath to the pea package need to be called every time before a pea code is run (as in the example).

The syntax of the three main programs inside the pea code is outlined below. Stata help files are available for all commands (refer to Annex 1: Full list of tables and figures, for a list of all help files and table and figure numbers). Each table or figure can also be produced separately, using the table/figure specific command (see Step 4 in Annex 2: Example code). All tables and figures include standardized tables notes which can be customized as needed. The core appendix tables are a subset of the full list tables. The following codes produce the core appendix tables and figures:

- Figure C.1: pea figureC1
- Table C.1: pea tableC1
- Table C.2: pea table5
- Figure C.2: pea figure3b
- Figure C.3: pea figure4
- Table A.1: pea table1
- Table A.2: pea tableA2
- Table A.3. pea tableA2
- Table A.4a. pea table4
- Table A.4b. pea table4

Table 2 provides details on the main options that need to be specified to produce output. Option setting (GMD), when specified, automatically creates all the input variables needed to produce the entire core analytics output. Furthermore, when specified, a number of additional options do not need to be specified. These options are listed in Table 3. The setting (GMD) option works only in conjunction with the main pea commands, and not the table or figure specific commands. For the table or figure specific commands, those options in Table 3 need to be specified that relate to the specific table or figure. Which options are required for which table or figure is specified in the help files (accessible through help pea core, help pea tables, help pea figures, help pea table10, help pea figure2, etc.). Not specifying setting (GMD) can be useful when users want to use categories and values for certain variables that differ from the harmonized default values listed in Table 1.

Every output, both figures and tables, are saved in an Excel sheet, along with table and figure notes. When the excel () option is not specified, a temporary Excel file will be created, which is automatically opened after the code is completed. Users can specify where the file should be saved with the excel () option. However, the in the option specified Excel file needs to already exist (it can be empty) prior to running the pea commands.

1. pea core

This first code (in Stata: pea core) generates a standardized data annex which is expected to be included in all PEAs. The annex includes main poverty and shared prosperity indicators, as well as multidimensional and sub-group (e.g. by age or education) poverty rates and poverty profiles. The code also benchmarks core statistics for a selected group of peer countries, the country's region and income group averages. Peer countries can be specified by the user and a benchmarking tool is available to support the selection of countries. Growth incidence curves and the Datt-Ravallion decomposition complement the core outputs. ¹⁰

```
pea core [weight] [if exp] [in exp] [,Year(varname numeric)

NATWelfare(varname numeric) NATPovlines(varlist numeric)

PPPWelfare(varname numeric) PPPPovlines(varlist numeric)

PPPyear(integer) svy std(string) SETting(string) BYInd(varlist numeric) ONELine(varname numeric) ONEWelfare(varname numeric)

comparability_peb(varname string) YRange(string) YRange2(string)

year_fcast(varname numeric) natpov_fcast(varname numeric)

gdp_fcast(varname numeric) missing latest within3 BENCHmark(string)

spells(string) excel(string) save(string)]
```

2. pea tables

The second set of outputs is produced using the pea tables command. These expand on the core outputs, providing more detailed statistics on poverty and shared prosperity over the last years, the extent of vulnerability to poverty, poverty profiles as well as drivers of welfare changes. The latter examines how

⁹ Refer to Annex 1: Full list of tables and figures, for a list of all help files and table and figure numbers.

¹⁰ See <u>benchmarking tool</u> and <u>guidance note</u> to identify benchmarking countries.

inclusive economic growth has been to those in the bottom of the income distribution, as well as decompositions of welfare changes into within and between group shifts.

```
pea tables [weight] [if exp] [in exp] [, Country(string) Year(varname numeric) natwelfare(varname numeric) natpovlines(varlist numeric) pppwelfare(varname numeric) ppppovlines(varlist numeric) PPPyear(integer) svy std(string) oneline(varname numeric) onewelfare(varname numeric) setting(string) byind(varlist numeric) latest within3 earnage(integer) missing benchmark(string) spells(string) excel(string) save(string)]
```

3. pea figures

The third output includes a series of figures that are produced using the pea figures command. Figures include a comparison of poverty rates and GDP per capita with other countries, changes in poverty and inequality over time, population composition by income decile, profiles, and exposure to risk from climate-related hazards. Options that are specific to the pea figures command are listed in Table 4. Each Excel sheet containing a figure also includes the raw data that is used to create the figure (in row 100 of the sheet).

```
pea figures [if] [in] [weight], [natwelfare(varname)
natpovlines(varlist) pppwelfare(varname) ppppovlines(varlist)
PPPyear(integer) year(varname) setting(string) byind(varlist)
oneline(varname) onewelfare(varname) missing country(string)
within(integer) combine comparability(varname) benchmark(string)
spells(string) noequalspacing scheme(string) palette(string)
welfaretype(string) yrange() ineqind() excel(string) save(string)]
```

¹¹See the <u>Corporate scorecard website</u> and <u>CCG dashboard</u> on the risk from climate-related hazards indicator.

Table 2: Main options

Setting	Description	Entry or variable		
1) Required options				
Country()	PEA country	3-letter country code		
natwelfare()	Welfare aggregate in current or constant LCU	welfare (or the national welfare aggregate)		
natpovlines()	National poverty line, multiple entries allowed	natline		
pppwelfare ()	Welfare aggregate in PPP terms	welfareppp		
ppppovlines ()	International poverty lines in PPP terms	pline215 pline365 pline685		
pppyear()	specifies which year PPPs are based on. Default is 2017	Integer (such as 2017)		
svy	Triggers the inclusion of standard-errors in Table 1 and Table A.1. Importantly, the data needs to be 'svy set' in Stata before the code is run. See Annex 2: Example code (step 3).			
std(string)	Specifies whether standard-errors are displayed in the same cell as the main statistic or in a separate cell to the right. Default is inside.	inside; right		
year()	Year variable in survey dataset	year		
onewelfare()	Main welfare variable used for comparisons (i.e. Table 1, Table 7, Table 14, Table A1, Table A2, Table A4, Table A4, Figure C1)	welfareppp		
oneline()	Main poverty line used for comparisons (i.e. Table 1, Table 7, Table 14, Table A1, Table A2, Table A4, Table A4, Figure C1)	pline685		
spells()	Years to be used when calculating or showing changes across periods (such as Growth Incidence Curves), multiple entries allowed separated by semicolon	year1 year2; year2 year3		
benchmark()	Countries to be benchmarked against, multiple entries allowed	3-letter country codes		
setting() or options in Table 3)	If GMD option is specified, harmonized variables are created, and additional options (Table 3) in do not need to be specified.	GMD		
2) Optional options				
aggregate()		groups benchmark		
byind()	Geographic units for disaggregation, multiple entries allowed. Note that while optional, it is used in many tables.	urban subnatvar		
urban()	Variable that indicates the urban indicator	urban		
comparabilit y_peb()	Only relevant for core Figure 1: This variable denotes which survey rounds are comparable over time, following PEB notation.	Yes No		

year_fcast()	Only relevant for core Figure 1: To show now- and forecasts in the figure, insert the variable with forecast years here.	
natpov_fcast ()	Only relevant for core Figure 1: To show now- and forecasts in the figure, insert the variable with forecast national poverty here.	
gdp_fcast	Only relevant for core Figure 1: To show now- and forecasts in the figure, insert the variable with forecast GDP here.	
missing	Missing values are reported separately. Otherwise, they are treated as non-existent.	
minobs()	Specifies the minimum number of observations required to display a cell value (relevant for Tables A2, 2 and 3)	integer
vulnerability ()	Specifies the number that is multiplied with the main poverty line to define vulnerability to poverty (relevant for Table 7). Default is 1.5.	Any number
earnage()	Specifies the age cut-off for working status for the economic composition (Table 14).	integer
latest	Includes only the most recent available data.	
Within3	Limits analysis to data from benchmark countries within 3 years of the target year.	
excel()	Path of directory where and name under which output file is saved. If it is not specified, the file is saved in "Temp". Note that if the option is specified, the excel file needs to exist before the code is run (empty).	Path and file name
save	Specifies the file path for saving results.	

Additional options, if setting (GMD) is not specified. Note that these are relevant for the pea core, pea tables and pea figures, and all individual tables and figures commands.

hhhead(varname numeric) edu(varname numeric) married(varname numeric) school(varname numeric) services(varlist numeric) assets(varlist numeric) hhsize(varname numeric) hhid(string) pid(string) industrycat4(varname numeric) industrycat10(varname numeric) lstatus(varname numeric) empstat(varname numeric) relationharm(varname numeric)

Table 3: Additional options if setting(GMD) is not specified

Setting	Description	Default under setting(GMD)
age()	Age of individual	age
male()	Sex of individual	male
married()	Marital status (variable is based on marital)	married
services()	Services to be included in the analysis. Default are access to improved water, access to improved sanitation, access to electricity.	<pre>imp_wat_rec imp_san_rec electricity</pre>

assets()	Assets to be included in the analysis	tv car cellphone computer fridge
hhsize()	Household size	hsize
relationh arm()	Relation to household head	relationharm
head()	Household head indicator (Variable is based on variable relationharm)	head
hhid()	Household identifier	hhid
pid()	Individual identifier	pid
edu()	Education level	educat4
school()	Currently in school dummy	school
Istatus()	Dummy if respondent does not work (Variable is based on variable Istatus). Please make sure the value = 1 refers to the respondent not working.	nowork
industryc at4()	Industry categories	industrycat4
Industryc at10()	1-digit Industry categories	Industrycat10
empstat	Employment status of respondent	empstat

Table 4: Additional optional figure options

Setting	Description	Entry or variable name
within()	Specifies the number of years before and after the pea survey year, to define which surveys from other countries should be used (e.g. in scatter plots on inequality). Default is 3, and value should be less than 10.	integer
combine	When specified, figures with multiple panels are combined to one figure with only one legend.	
comparability()	This variable denotes which survey rounds are comparable over time. Non-comparable survey rounds are not connected in figures.	comparability
noequalspacing	When specified, figures show gaps between years on the x-axis proportional to their distance. Default is to display constant gaps between years, regardless of how far away years are. The option can be useful if gaps between survey-years are large.	
scheme()	Sets the scheme, specifying the overall look of the figures. Default is "white_tableau".	string
palette()	Sets the color palette for figures. Default is "tab20". See Annex 2: Example code for an example of a custom set of colors.	Either <i>string</i> (e.g. <i>vividis</i>) or list of colors.
welfaretype()	Can be used to specify whether the survey uses income or consumption to calculate welfare. Figures showing scatters of inequality display different symbols for countries with consumption or income aggregates.	CONS or INC

yrange()	Users can specify the range of the y-axis. The range must be entered in Stata figure format.	Example: "0(10)100"
Yrange2()	For core Figure 1, users can specify the range of the RIGHT y-axis (GDP per capita).	Example: "0(10)100"
Ineqind()	Users can specify which inequality indicators to show in Figure 9a. Entry options are Gini Theil Palma Top20. Default are all four.	Any combination of: Gini Theil Palma Top20
bar	Users can specify this option to display Figures 1, 9a and 10a as bar graphs instead of line graphs. Warning: All selected years will be shown in the figures, regardless of whether they are comparable or not.	
trim()	Specifies percentiles below and above which growth incidence curves are trimmed.	Example "trim(3 97)"
Idpl()	Affects only Figure 4: If national poverty lines are different within one year, specify the variable grouping the poverty lines here. This is relevant for the Shorrocks-Kolenikov decomposition.	Example: "idpl(urban)"
relativechange	Affects only Figure 12: When specified, changes in mean welfare and changes in inequality are presented relative to changes in the prosperity gap. The default is that contributions are not presented in relative terms, the addition of the two are the change in the prosperity gap.	

III. Finding the right outputs for PEAs

The following lists the tables and figures that can be used to answer the questions raised in the <u>proposed</u> PEA structure.

- 1.1. How have poverty and shared prosperity evolved in recent years?
 - What is the Prosperity Gap for the country and has it improved in recent years?
 - Table 1, Table A1, Figures 10a-d, Figure 12
 - What are the trends in income/consumption inequality using the Gini index, income shares/income share ratios, and the Palma ratio?
 - Table 8, Figures 9a-b
 - Benchmarking of poverty, inequality, and Prosperity Gap with regional, structural, and/or aspirational peers. Teams may use the PEA benchmarking tool to identify comparator countries.
 - Table A3, Table 10, Figure 2, Figure 9c, Figure 10d
 - What is the extent of multidimensional poverty? Examine level of deprivations, degree of overlap
 across dimensions, and compare these patterns with monetary poverty.
 - Table 6a, Table 6b, Figures 14a-c
- 1.2. How many people are at risk of falling back into poverty and is there a stable middle class?
 - What share of the population is at high risk of falling into poverty and how does it compare to the share of poor? How has that share evolved over time? What is the share of the population that is at high risk from climate-related hazards (briefly discuss climate-related vulnerability)?

- Table 7, Table 9, Figure 15
- 1.3. What are the distinct characteristics of the poor compared to the non-poor?
 - What are the levels and trends of poverty and inequality by urban and rural areas and key regions?
 Examine poverty rates and number of poor. If available, present results from poverty maps.
 - Table 2, Table A2, Figure 1
 - Has the gap in poverty and inequality narrowed or widened over time?
 - Table 1, Table A1, Table 8, Figure 1, Figures 3a-b, Figure 9a
 - What are the demographic and socioeconomic characteristics (age, sex, education, access to basic services, household size and composition, location) of the poor, those at high risk of falling into poverty, and the middle class? Focus on notable differences; using regressions can help find relevant factors.
 - > Tables 3a-c, Figures 7a-b
 - How do the profiles vary by household type, defined with an approach that classifies households by their demographic and economic composition to better understand women's experiences across the life cycle?
 - Table A4a, Table 4a, Table 14, Figure 8, Figure 16
- 2.1. Is growth conducive to poverty reduction and economic inclusion?
 - How is GDP growth associated with poverty reduction and has there been a change in the growth-poverty elasticity? Benchmark against peers.
 - Figure 6
 - What does the growth incidence curve (GIC) suggest about how the benefits of growth are distributed, i.e., how does income/consumption growth vary across the distribution? Has there been a change in the GIC over time? (Required figure 2).
 - Figure 3a
 - What is the relative contribution of growth and redistribution to poverty changes? (Present the Datt-Ravallion decomposition in required figure 3).
 - Tables 12a-b, Figures 4a-b
 - What is the role of urban-rural growth patterns and their contribution to poverty changes? The GIC by urban/rural areas can help to contrast patterns. A Huppi-Ravallion decomposition can be used to examine whether poverty reduction came mainly through a growth effect (i.e., higher incomes/consumption within urban or rural areas) or a population shift effect (i.e., rural-to-urban migration where people move from high-poverty rural areas to low-poverty urban areas).
 - Table 13a, Figure 3b, Figure 5a
- 2.2. What is the role of labor and nonlabor incomes behind changes in poverty and inequality?
 - A Huppi-Ravallion decomposition can help understand which households have made more
 progress, depending on the main income sources or sectors of activity (detailed income data is
 not required unlike the above Shapley decomposition). For example, if the decomposition is
 conducted with information on household's primary sector of activity, the analysis would indicate
 whether changes in income and poverty are mainly attributed to households working in
 agriculture or non-agriculture (suggesting within-sector productivity increases or other factors

that benefit the specific sector) or "population shifts" which capture households moving sectors, usually from agriculture to non-agriculture (suggestive of structural transformation).

- Table 13b, Figure 5b
- 2.3. Has growth translated into more and better jobs for the poor and vulnerable populations?
 - What is the extent of labor market engagement—both extensive (participation) and intensive (hours worked)—across the distribution and among key population groups? Required Table 2.
 - Table A4b, Table C2, Table 4b, Table 5

IV. Caveats and FAQ

- Note that the nominal welfare variable (welfare) in the harmonized GMD data is not necessarily the welfare aggregate that is used to construct national poverty rates and lines. Key differences can include spatial deflation or the welfare aggregate being transformed in adult-equivalent rather than per-capita terms. To obtain correct national poverty estimates, the analyst needs to construct the correct respective national welfare aggregate, and pass it into the welfare() or onewelfare() options.
- Importantly, when the nominal welfare aggregate is passed into the onewelfare() option, growth incidence curves will show growth in welfare in nominal, not in in real terms. In this case, it is advised to redo the growth incidence curves separately (e.g. pea table11, welfare (welfppp) spells (2018 2022) year (year) setting (GMD) graph).
- The notes of Tables 3b, 3c, 14a A4a, and Figures 7a, 7b, specify that education levels refer to "highest level attended, complete or incomplete". If this does not correspond to the data used in the analysis, please change the notes of these outputs accordingly.
- When there are multiple survey periods, make sure that welfare in LCU and national poverty lines are both expressed in current prices or both in constant prices.
- Note that when nominal welfare in current prices is specified in the <code>oneline()</code> option, growth incidence curves will not reflect changes in real-terms. Therefore, specifying welfare in constant terms in the <code>oneline()</code> is encouraged, or re-doing the growth incidence curves separately in constant terms (pea table11, pea figure3) is recommended.
- It is important that the correct temporal and spatial deflators are used.
- Poverty numbers are based on population numbers from the respective surveys. Note that this
 may cause slight discrepancies between poverty numbers reported in PIP and in PEBs, which are
 based on population from WDI.
- Labor force related variables can have different recall periods. For example, in GMD, some survey data use the variable *Istatus* (which has a 7-day recall period) and some use *Istatus_year* (which has a 1-year recall period). It is up to country teams to decide whether variables are comparable over time (e.g. even when recall periods are different). For the example of labor force status, if different recall periods are deemed to be comparable, one variable would need to be defined which has both variables' values (e.g. replace lstatus = lstatus_year if lstatus==.)

- Furthermore, labor force related variables (Istatus and empstat) often have many missing values.
 These variables affect Table A.4a, Table A.4b, Table 13b, Table 14a, Table 14b, Figure 5b and Figure
 16. It is advised to inspect the data before running the commands for these tables. Note that missing observations can lead to different poverty rates, for example for the sectoral Huppi-Ravallion decomposition.
 - Please make sure the value = 1 of the Istatus variable refers to the respondent not working.
- Table 13b and Figure 5b, which apply Huppi-Ravallion decompositions on the sectoral level, will
 only be produced if the specified sector variable (default is industrycat4) will be non-missing for
 the years that are requested. Please ensure that this is the case in your data.
- The MPM is retrieved from the GMI. New surveys in the GMD might not be immediately included in the GMI (updated usually in April and October).
- For growth incidence curves disaggregated by urban and rural areas, the calculation is performed for the welfare distribution within that area only.
- The "Other" sector in the variable industrycat4 is based on a country-specific harmonization from the GMD which can vary across countries.
- Note that not all surveys of a country in the GMD are comparable over time, due to updated
 methodologies. The variable "comparability" indicates which surveys are comparable. When
 specified, the comparability option in the pea figures commands shows breaks
 between non-comparable survey spells.
- Please ensure that all variables have proper value labels. Core variables, such as educat4 or male, not having labels may cause errors.

FAQ

Q1: How can I check whether the main poverty lines and welfare aggregates are correct?

✓ Use the following code to check whether the poverty lines and welfare aggregates reproduce the poverty rates under the international poverty lines in PIP, and the correct national poverty rate.

```
// international poverty
gen ipoor = welfppp < pline215
bys year: sum ipoor [aw = weight_p]
// national poverty
gen poor = natwelfare < natline
bys year: sum poor [aw = weight_p]</pre>
```

Q2: Why are my tables missing some indicators?

✓ Check if all required variables (Step 3) are present in the dataset.

✓ Ensure numerical variables are correctly formatted (encode string variables if needed).

Q3: Why are some of the core tables not produced?

✓ Check if all required variables (Step 3) are present in the dataset.

✓ If the svy option is specified, please make sure that the data is actually svy set.

Q4: Can I use country-specific variables instead of GMD harmonized variables for education, sector, etc?

✓ Yes, not specifying the option setting (GMD) allows to insert country-specific variables in the options specified in Table 3. For example, the globally harmonized education variable in the GMD database takes on only 4 values (educat4), but a different, country-specific variable can be inserted in option edu().¹² See example 2b under Step 4 in Annex 2: Example code.

Q5: How do I define a custom color palette for figures?

✓ Use the palette () option, see example under Step 4 in Annex 2: Example code.

Q6: How do I include multiple years for analysis?

✓ Make sure that all variables are consistently specified over all years

✓ Options years () and spells () specify which years are used in the analysis. Use spells (year1 year2; year3 year4), to select which pairs of years should be used to display growth rates (for example in growth incidence curves).

Q7: I get an error that datalibweb or pip cannot be accessed, what can I do?

✓ Make sure that your datalibweb token is active (see Step 2a, footnote 10). It needs to be renewed monthly.

 \checkmark PIP or datalibweb could be temporarily not accessible. You can test by running 'pip, clear' or 'dlw' in Stata.

Q8: Figure 2 is not produced with the pea figures command. Why?

¹² In the GMD, education by default follows the International Standard Classification of Education (ISCED) mappings.

✓ Figure 2 is a scatter plot of poverty rates of the PEA country against benchmark countries. It will not be produced if the national poverty line is passed in the option oneline(), as it is not comparable across countries.

✓ Either enter a PPP-adjusted poverty line in the oneline() option of the pea figures command, or produce figure 2 separately. For example: pea figure2 [aw=weight_p], c(PHL) year(year) onew(welfppp) onel(pline215) benchmark(VNM BGD PAK NPL IDN)

Annex 1: Full list of tables and figures

Core and standardized data annex (pea core):

- Figure C.1. Trends and nowcast of the national poverty rate
 - help pea figureC1
- Table C.1. Key Poverty, Shared Prosperity and Labor Market Indicators
 - o help pea tableC1
- Table C.2. Key Poverty, Shared Prosperity and Labor Market Indicators
 - o help pea table5
- Figure C.2 Growth Incidence Curves
 - o help pea figure3b
- Figure C.3 Datt-Ravallion decomposition
 - help pea figure4
- Table A.1. Core poverty and equity indicators
 - help pea table1
- Table A.2. Poverty indicators by subgroup
 - help pea tableA2
- Table A.3. Benchmarking of poverty and inequality
 - o help pea table10
- Table A.4a. Profiles of the poor
 - o help pea table4
- Table A.4b. Demographic and economic household typologies
 - o help pea table4

Full set of tables (pea tables):

- Table 1. Core poverty indicators
 - o help pea table1
- Table 2. Core poverty indicators by geographic areas
 - o help pea table2
- Table 3a. Subgroup poverty rates by sex and age-group
 - help pea table3
- Table 3b. Subgroup poverty rates by education (age 16+, %)
 - help pea table3
- Table 3c. Subgroup poverty rates of household head (%)
 - o help pea table3
- Table 4a. Demographic profiles of the poor
 - o help pea table4
- Table 4b. Labor force profiles of the poor
 - o help pea table4
- Table 5. Key labor market indicators by population group
 - o help pea table5
- Table 6a. Multidimensional poverty: Multidimensional Poverty Measure (World Bank) (%)
 - o help pea table6

- Table 6b. Multidimensional poverty: Multidimensional poverty components (%) (World Bank)
 - o help pea table6
- Table 7. Poverty vulnerability
 - help pea table7
- Table 8. Core inequality indicators
 - o help pea table8
- Table 9. Scorecard Vision Indicators
 - o help pea table9
- Table 10. Benchmarking of poverty and inequality
 - help pea table10
- Table 12a. Decomposition of poverty changes: growth and redistribution Datt-Ravallion decomposition
 - help pea table12
- Table 12b. Decomposition of poverty changes: growth and redistribution Shorrocks-Kolenikov decomposition
 - help pea table12
- Table 13a. Huppi-Ravallion decomposition (rural/urban)
 - o help pea table13
- Table 13b. Huppi-Ravallion decomposition (sectoral)
 - o help pea table13
- Table 14. Demographic and economic household typologies
 - o help pea table14
- Table 15. Distribution of welfare by deciles
 - o help pea table15

Full set of figures (pea figures):

- Figure 1: Poverty rates by year lines.
 - o help pea figure1
- Figure 2. Poverty and GDP per capita in benchmark countries
 - help pea figure2
- Figure 3a. Growth Incidence Curves over time
 - o help pea figure3a
- Figure 3b. Growth Incidence Curves by area
 - o help pea figure3b
- Figure 4a. Datt-Ravallion decompositions
 - o help pea figure4
- Figure 4b. Shorrocks-Kolenikov decompositions
 - o help pea figure4
- Figure 5a. Huppi-Ravallion decomposition (rural/urban)
 - help pea figure5
- Figure 5b. Huppi-Ravallion decomposition (sectoral)
 - o help pea figure5
- Figure 6. GDP Poverty elasticity
 - help pea figure6

- Figure 7a. Share of poor and population by demographic groups
 - o help pea figure7a
- Figure 7b. Poverty rates by demographic groups
 - help pea figure7b
- Figure 8. Poverty rates by sex and age groups
 - o help pea figure8
- Figure 9a. Inequality by year
 - o help pea figure9a
- Figure 9b. Gini and GDP per capita scatter
 - o help pea figure9b
- Figure 9c. Benchmark countries ranked by Gini index
 - o help pea figure9c
- Figure 9d. Welfare percentiles by year
 - o help pea figure9d
- Figure 10a. Prosperity gap by year and area
 - o help pea figure 10a
- Figure 10b: Prosperity gap (line-up) and GDP per capita scatter
 - o help pea figure 10b
- Figure 10c: Prosperity gap (survey) and GDP per capita scatter
 - o help pea figure 10c
- Figure 10d: Prosperity gap over time in benchmark countries (line-up)
 - o help pea figure 10d
- Figure 12. Decomposition of growth in prosperity gap
 - o help pea figure 12
- Figure 13. Distribution of welfare by deciles
 - help pea figure13
- Figure 14a. Multidimensional Poverty Measure, by components, versus benchmark countries
 - o help pea figure 14
- Figure 14b. Multidimensional poverty and poverty rates, versus benchmark countries
 - o help pea figure 14
- Figure 14c. Multidimensional poverty and poverty rates, contributions
 - o help pea figure 14
- Figure 15: Climate risk and vulnerability
 - o help pea figure 15
- Figure 16: Profiles of the poor by demographic and economic composition
 - help pea figure16

Annex 2: Example code

The following code produces the main tables, the standardized data annex and figure outputs for Guinea-Bissau.

```
// Set path where the survey data is stored, or should be stored if
accessing from datalibweb
global pea path "C:/Users/data/ "
// Step 2: Access data (either from survey, or from datalibweb)
datalibweb, country(GNB) year(2018 2021) type(qmd) mod(all) clear
// Step 3: Preparation of additional variables
// Welfare aggregate to compute international poverty rates
gen welfppp = welfare/cpi2017/icp2017/365
gen pline215 = 2.15
gen pline365 = 3.65
gen pline685 = 6.85
// National welfare & poverty line: Adjust according to the PEA country
gen natwelfare = welfare
gen natline = 298083.5 if year == 2021
replace natline = 271071.8 if year == 2018
la var pline215 "$2.15 per day (2017 PPP)"
la var pline365 "$3.65 per day (2017 PPP)"
la var pline685 "$6.85 per day (2017 PPP)"
la var natline "National poverty line (2017 LCU)"
// Comparability national poverty for Figure C.1 (PEB)
// define here which years are comparable for national poverty rates
// years which are comparable should have value "Yes", as in the PEBs
gen comparability peb = "Yes"
// Cleaning of region variable: Adjust according to the PEA country
split subnatid, parse("-") gen(tmp)
replace tmp2 = ustrlower( ustrregexra( ustrnormalize( tmp2, "nfd" ) ,
"\p{Mark}", "" ) )
replace tmp2 = " bolama/bijagos" if tmp2 == " bolama bijagos"
replace tmp2 = proper(tmp2)
encode tmp2, gen(subnatvar)
// Additional labeling for table and figure notes
                    "Welfare (constant 2017 PPP)"
label var welfppp
label var natwelfare "Welfare (local currency)"
local lbl: value label educat4
label define `lbl' 1 "No education" 2 "Primary" 3 "Secondary" 4
"Tertiary", modify
// Survey set the data, to get correct standard errors
svyset psu [w= weight p], singleunit(certainty)
// Add now- and forecast data (including the three relevat fcast
variables)
```

```
append using "$pea path/data/GNB forecast.dta"
save "${pea path}/data/GNB GMD clean.dta", replace
// Step 5: Run pea commands
// la. Core tables and figures code:
//Set ado path
adopath + "${pea path}/ado/pea/Stata/plus"
//Call data
use "${pea path}/data/GNB GMD clean", clear
//Run code with options
pea core [aw=weight p], c(GNB) year(year)
                                                                    ///
           natw(natwelfare) natp(natline)
                                                                    ///
           pppw(welfppp) pppp(pline365 pline215 pline685)
                                                                    ///
           onew(welfppp) oneline(pline365)
                                                                    ///
           byind(urban subnatvar)
                                                                    ///
           benchmark(CIV GHA GMB SEN)
                                                                    ///
           setting(GMD) missing
                                                                    ///
           spells(2018 2021)
                                                                    ///
           aggregate (groups)
                                                                    ///
           comparability peb(comparability peb)
                                                                    ///
           year fcast(year fcast)
                                                                    ///
           natpov fcast(natpov fcast)
                                                                    ///
           gdp_fcast(gdp_fcast)
                                                                    ///
           yrange(20(20)80)
                                                                    ///
           yrange2(300000(50000)500000)
                                                                    ///
           svy std(inside)
// 1b. Core tables code without setting(GMD):
//Set ado path
adopath + "${pea path}/ado/pea/Stata/plus"
//Call data
use "${pea path}/data/GNB GMD clean", clear
//Run code with options
                                                                    ///
pea core [aw=weight p], c(GNB) year(year)
           natw(natwelfare) natp(natline)
                                                                    ///
           pppw(welfppp) pppp(pline365 pline215 pline685)
                                                                    ///
           onew(welfppp) oneline(pline365)
                                                                    ///
           byind(urban subnatvar)
                                                                    ///
           benchmark(CIV GHA GMB SEN)
                                                                    ///
           comparability peb(comparability peb)
                                                                    ///
           missing spells (2018 2021)
                                                                    ///
                                                                    ///
           svy std(inside)
           age(age) male(male) hhhead(head) edu(educat4)
                                                                    ///
           urban(urban) married(married) school(school)
                                                                    ///
           services (imp wat rec imp san rec electricity)
                                                                    ///
```

```
///
           assets(tv car cellphone computer fridge)
           hhsize(hsize) hhid(hhid) pid(pid)
                                                                   ///
           industrycat4(industrycat4)
                                                                   ///
           industrycat10(industrycat10)
                                                                   ///
           relationharm(relationharm)
                                                                   ///
           lstatus(nowork) empstat(empstat)
// 2. Main tables code:
//Set ado path
adopath + "${pea path}/ado/pea/Stata/plus"
//Call data
use "${pea path}/data/GNB GMD clean", clear
//Run code with options
pea tables [aw=weight p], c(GNB) year(year)
                                                                   ///
           natw(natwelfare) natp(natline)
                                                                   ///
           pppw(welfppp) pppp(pline365 pline215 pline685)
                                                                   ///
           onew(welfppp) oneline(pline365)
                                                                   ///
           byind(urban subnatvar) vulnerability(2)
                                                                   ///
           benchmark(CIV GHA GMB SEN)
                                                                   ///
           setting(GMD) missing
                                                                   ///
           spells(2018 2021)
                                                                   ///
           svy std(inside)
// 3a. Figures code:
//Set ado path
adopath + "${pea path}/ado/pea/Stata/plus"
//Call data
use "${pea path}/data/GNB GMD clean", clear
//Run code with options
pea figures [aw=weight p], c(GNB) year(year)
                                                                   ///
           natw(natwelfare) natp(natline)
                                                                   ///
           pppw(welfppp) pppp(pline365 pline215 pline685)
                                                                   ///
           onew(welfppp) oneline(pline215)
                                                                   ///
           byind(urban) benchmark(CIV GHA GMB SEN AGA)
                                                                   ///
           spells(2010 2018; 2018 2021)
                                                                   1//
           setting(GMD) urban(urban) within(3)
                                                                   ///
           comparability(comparability) welfaretype(CONS)
                                                                   ///
           combine
// 3b. Figures code but bars instead of lines and non-equal spacing
between years:
//Set ado path
adopath + "${pea path}/ado/pea/Stata/plus"
//Call data
use "${pea path}/data/GNB GMD clean", clear
//Run code with options
pea figures [aw=weight p], c(GNB) year(year)
                                                                   ///
           natw(natwelfare) natp(natline)
                                                                   ///
```

```
pppw(welfppp) pppp(pline365 pline215 pline685)
                                                                   ///
           onew(welfppp) oneline(pline215)
                                                                   ///
                                                                   ///
           byind(urban) benchmark(CIV GHA GMB SEN AGA)
           spells(2010 2018; 2018 2021)
                                                                   ///
           setting(GMD) urban(urban) within(3)
                                                                   ///
           comparability(comparability) welfaretype(CONS)
                                                                   ///
           combine
                                                                   ///
           bar noequalspacing
// 4. Running a single figure and defining own color palette
use "${pea path}/data/GNB GMD clean", clear
adopath + "${pea path}/ado/pea/Stata/plus"
local custom palette = "#337ab7 #5cb85c #5bc0de #f0ad4e #d9534f #e6e6e6
#286090 #449d44 #31b0d5 #ec971f #c9302c"
pea figure2 [aw=weight p], c(GNB) year(year)
                                                                   ///
           onew(welfppp) onel(pline215)
                                                                   ///
           benchmark(CIV GHA GMB SEN) palette(`custom palette')
// 5. Running a single table with LCU poverty line
use "${pea path}/data/GNB GMD clean", clear
adopath + "${pea path}/ado/pea/Stata/plus"
// When running single tables, setting(GMD) is not called, and additional
variables need to be created
gen head = relationharm==1 if relationharm~=.
la def head 1 "HH head"
la val head head
gen nowork = lstatus==2|lstatus==3 if lstatus~=.
label define nowork 0 "Working" 1 "Not working (unemployed or out of
labor force)"
label values nowork nowork
gen married = marital==1 if marital~=.
pea table4 [aw=weight p], welfare(natwelfare) povlines(natline)
                                                                   ///
           year(year) missing urban(urban)
                                                                   ///
                                                                   ///
           age(age) male(male) hhhead(head)
           edu(educat4) married(married) school(school)
                                                                   ///
           services(imp wat rec imp san rec electricity)
                                                                   ///
                                                                   ///
           assets(tv car cellphone computer fridge)
           hhsize(hsize) hhid(hhid) pid(pid)
                                                                   ///
                                                                   ///
           industrycat4(industrycat4)
           industrycat10(industrycat10)
                                                                   ///
           lstatus(nowork)empstat(empstat) relationharm(relationharm)
```