



CROSS-NATIONAL  
DATA CENTER  
*in Luxembourg*

# The challenges of harmonising income data from middle- and low-income countries

World Bank-IARIW-TNBS Pre-conference Training Session  
Income and Wealth Measurement in Household Surveys  
10 November 2022, Arusha

Teresa Munzi

# Contents

- Overview of LIS and the LIS data
- Data harmonisation and dissemination
- Different welfare concepts
- Inclusion of middle/low-income countries
- An example: harmonisation of income data from Mali

# Contents

- **Overview of LIS and the LIS data**
- Data harmonisation and dissemination
- Different welfare concepts
- Inclusion of middle/low-income countries
- An example: harmonization of income data from Mali

# Luxembourg Income Study – the institution

- **Founded in 1983** by two US academics (Professors Tim Smeeding and Lee Rainwater) and a team of multi-disciplinary researchers out of the need to be able to **compare income distributions** across countries
- **Mission:** enable, facilitate, promote, and conduct cross-national comparative research on socio-economic outcomes and on the institutional factors that shape those outcomes
- Currently directed by Professor **Peter Lanjouw**, LIS' main strategy and priorities are set by a **Board** currently presided by Professor **François Bourguignon**
- LIS is supported in large part by the **Luxembourg government** and the **US NSF**, as well as by other national or supranational funders – supplemented by project-specific grants
- Located in **Luxembourg**, with two LIS satellite offices:
  - **US LIS Office** at the Stone Center on Inequality, CUNY, New York
  - **UK LIS Office** at the International Inequalities Institute, LSE, London



# Overview of the LIS data

## Deliverable

Three cross-national harmonised databases that allow international comparative research using micro-data:

- LIS (focus on **income**): 655 datasets – original database
- LWS (focus on **wealth**): 77 datasets – launched in 2007
- ERF-LIS (focus on income and consumption in **Arab countries**): 28 datasets – collaboration with ERF

## Scope

Initial focus on **high-income countries**, successively extended to **middle-** and more recently **low-income countries** (~60 countries)

## Time span

From the late 1960s to 2020, gradually becoming **annual**

## Geographical coverage

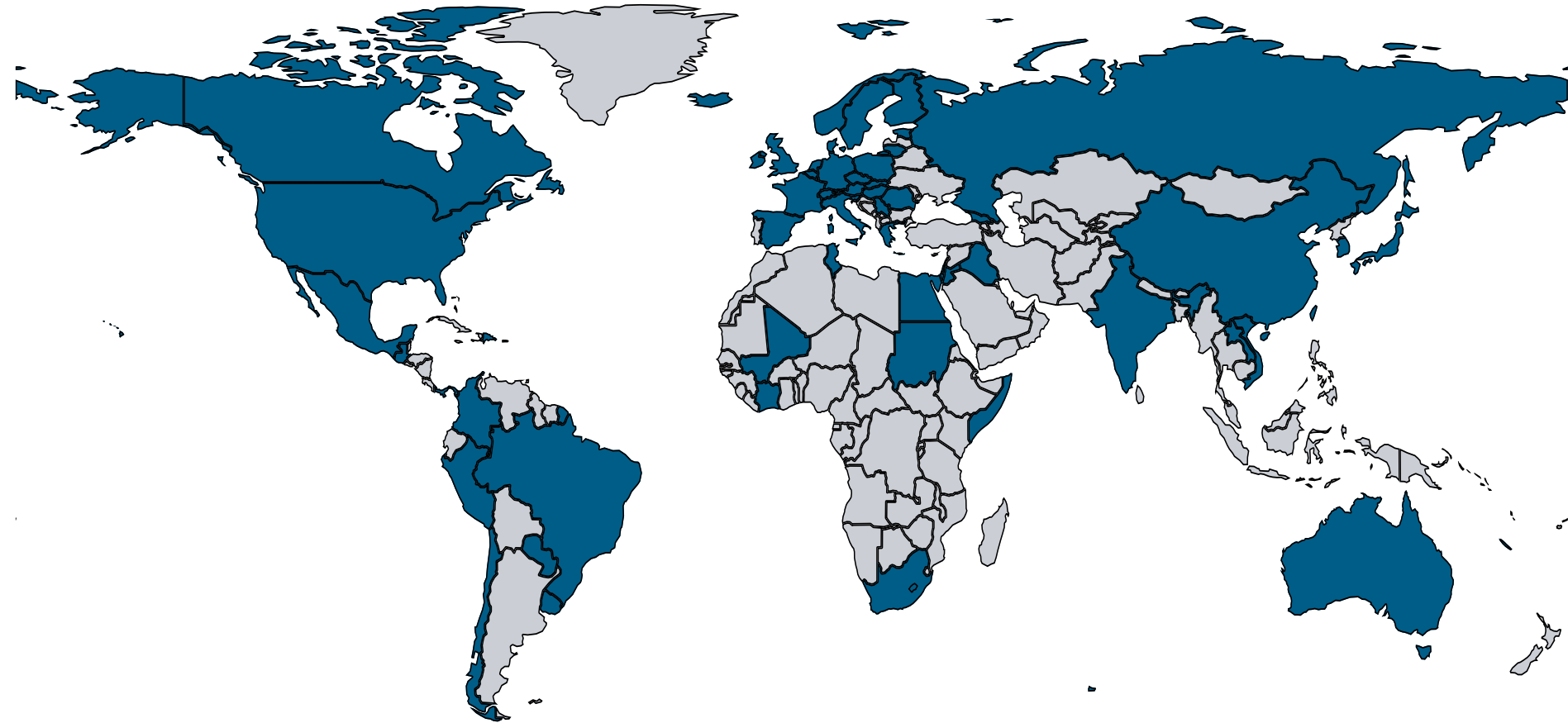
World-wide, but some regions are less covered (Africa, Middle-East, Caucasus and Central Asia...)

## Main contents

- household composition and characteristics
- socio-demographic characteristics of household members
- extensive set of labour market data
- detailed breakdown of household and individual income data
- household consumption data
- a detailed set of wealth and behavioural variables (LWS only)



# LIS Coverage



-  Not covered
-  Covered



# Contents

- Overview of LIS and the LIS data
- **Data harmonisation and dissemination**
- Different welfare concepts
- Inclusion of middle/low-income countries
- An example: harmonization of income data from Mali

# What we do at LIS

## Core activity: **Data work**

### Step 1. Data acquisition

We identify appropriate datasets (neutrality, completeness, representativeness)

We negotiate with each data provider

### Step 2. Data harmonisation

Common cross-national template

Data validation (consistency checks across countries, over-time and with external data)

Comprehensive documentation

### Step 3. Data dissemination

We create national-level indicators (*LIS Key Figures* and *DART*)

We provide harmonized microdata to researchers via remote execution (*LISSE*): an average of 1,000 researchers submit about 80,000 jobs remotely every year

## Other activities: **Research-promotion**

### Pedagogical activities

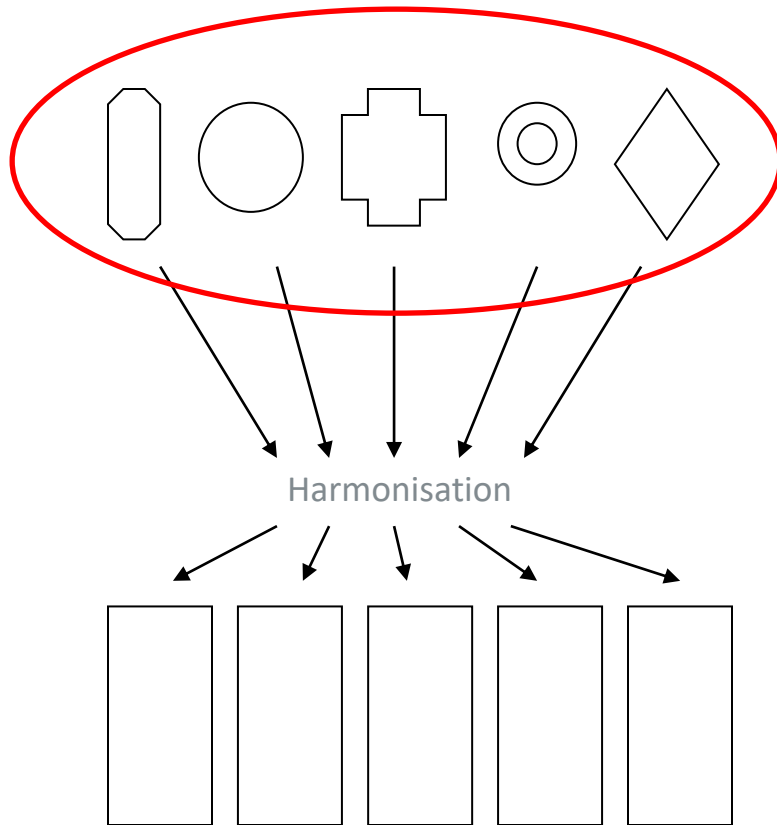
Annual training workshop, local workshops

### Research activities and support

User and research conferences, visiting scholar program, working paper series (N=920 papers), quarterly newsletter (N=23 issues), edited books



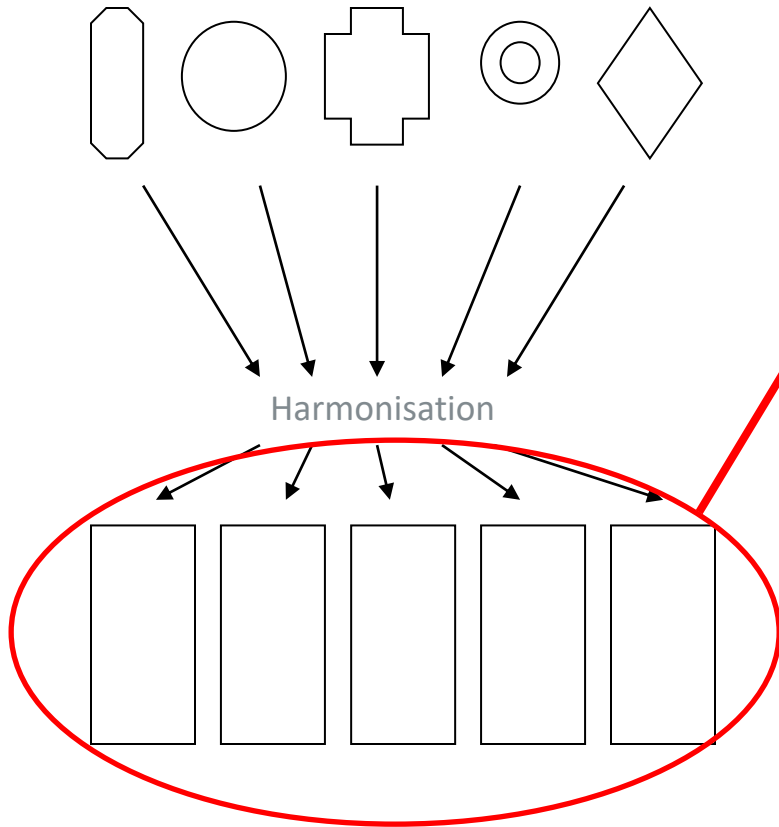
# Ex-post harmonisation at LIS



The origins of the LIS data

- ❑ Existing data sources:
  - *Survey data: income, household budget, living conditions, multipurpose, human development*
  - *Administrative records: tax records, employers records, social security records*
  - *Any mix of the above*
  
- ❑ Common denominator:
  - *microdata (household and individual level)*
  - *representative of the whole population*
  - *good quality income/wealth data*
  - *main demographic and (possibly) labour market information*

# Ex-post harmonisation at LIS



Final output:  
the LIS/LWS datasets (CCYY)

- ❑ Technical harmonisation: same file structure, same variables

LIS files

LIS Household File (H)			
HID	...	...	...
LIS Person File (P)			
HID	PID	...	...

LWS files

LWS Household File (H)			
HID	INUM	...	...
LWS Person File (P)			
HID	PID	INUM	...
LWS Replicate Weights File (R)			
HID	...	...	...

- ❑ Conceptual harmonisation
  - Based on the same definitions
  - Comparable concepts

 Operational comparability

 BUT: very challenging due to very different inputs!

# Challenges of harmonisation

Make comparable original data that are:

- ❑ from **various countries**
  - different institutional / societal setups
- ❑ **over time**
  - changes in institutions and original surveys
- ❑ **household / individual level data**
  - confidentiality issues
- ❑ from **various existing datasets**
  - output (or *ex-post*) harmonisation

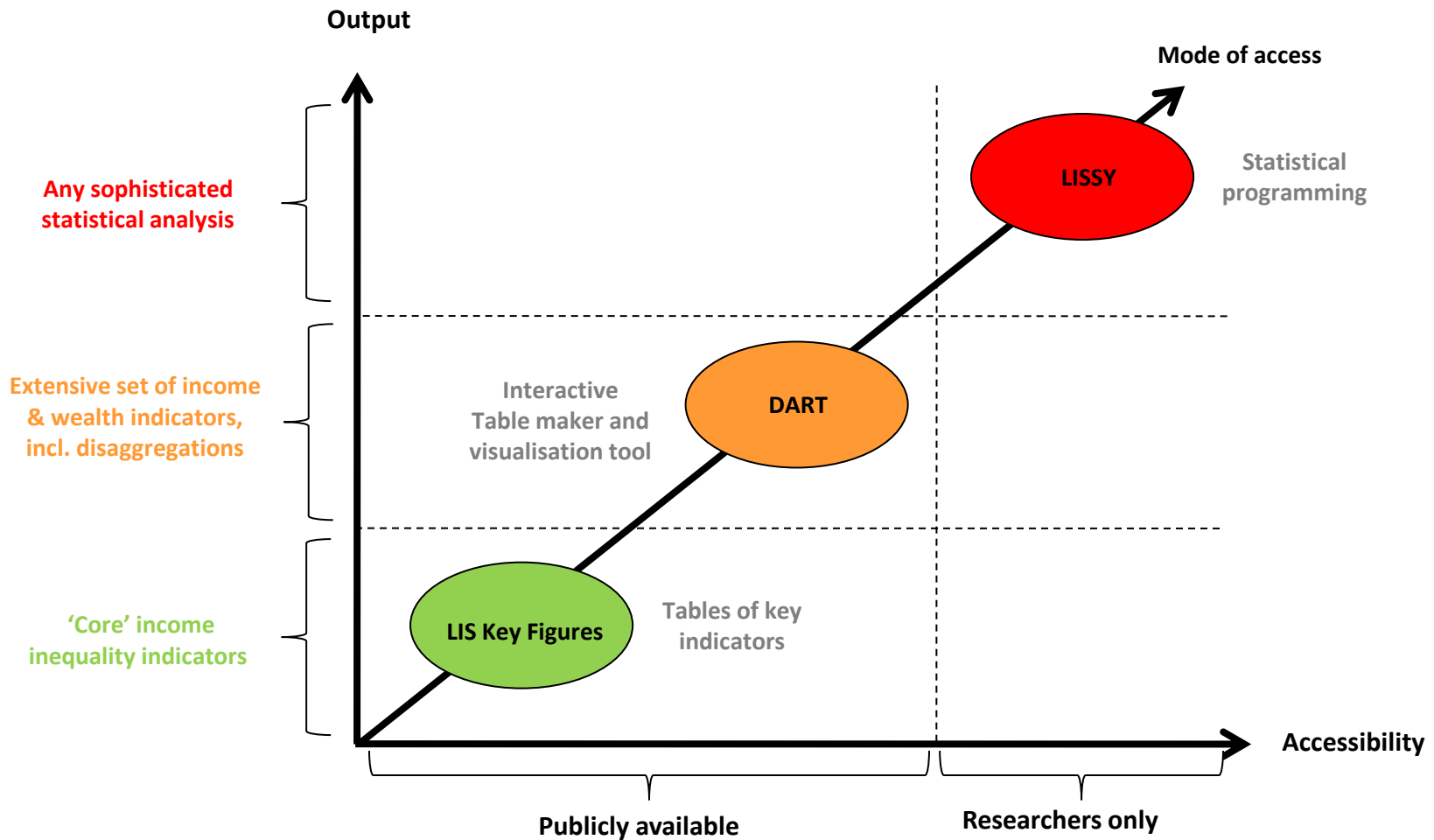
# Challenges of ex-post harmonisation

- ❑ Different types/purposes of **original collection instrument**
  - Survey versus administrative data (coverage and contents)
  - Cross-sections versus panels (sample selection)
- ❑ The **concepts** used in the original data collection are different
  - Different definitions (employment concept, total household disposable income concept, gross vs net incomes, assets vs net worth, etc.)
  - Different universes and reference periods
  - Country-specific categories (especially in education, social security benefits)
- ❑ The **level of detail** of information collected differs
  - Labor market (e.g.: LFS type of survey)
  - Incomes /wealth (detailed breakdown vs. overall questions)
- ❑ Different **statistical techniques**
  - Different sampling procedures (e.g. oversampling of the rich)
  - Weighting procedures (self-weighted, sampling weights, etc.)
  - Treatment of missing values, imputation methods

➔ **These challenges became all the more relevant when LIS started including middle-income countries**



# Data access



# Contents

- Overview of LIS and the LIS data
- Data harmonisation and dissemination
- **Different welfare concepts**
- Inclusion of middle/low-income countries
- An example: harmonization of income data from Mali

# LIS Household Disposable Income

		<b>CASH</b>	<b>NON CASH</b>
<b>LABOUR INCOME</b>	Dependent employment	<i>Wages, salaries, bonuses</i>	<i>In-kind earnings</i>
	Self-employment	<i>Profits and losses</i>	<i>Own consumption</i>
<b>+ CAPITAL INCOME</b>	Financial investment	<i>Interest and dividends</i>	-
	Real estate investment	<i>Rental income</i>	<i>Imputed rent</i>
<b>+ PENSIONS</b>	Non-contributory	<i>Universal and assistance</i>	-
	Contributory	<i>Work-related insurance</i>	-
	Private	<i>Occupational and individual</i>	-
<b>+ PUBLIC BENEFITS</b>	Family benefits	<i>Maternity, children</i>	
	Unemployment benefits	<i>Insurance and assistance</i>	<i>STIK</i>
	Sickness and disability	<i>Sick pay, work injury, disability</i>	
	Housing	<i>Housing, heating</i>	
	General	<i>Minimum income guarantee</i>	<i>In-kind assistance</i>
<b>+ PRIVATE TRANSFERS</b>	Private institutions	<i>Scholarships, charity</i>	<i>In-kind assistance</i>
	Inter-household	<i>Alimony, remittances</i>	<i>Gifts</i>
<b>= TOTAL GROSS INCOME</b>			
<b>- DEDUCTIONS</b>		<i>Income taxes</i>	
		<i>Social security contributions</i>	
<b>= HOUSEHOLD DISPOSABLE INCOME</b>			

# Consumption (!= Expenditure)

	<b>CASH</b> <u>(goods and services paid for)</u>	<b>NON CASH</b> <u>(goods and services not paid for)</u>
<b>CONSUMPTION</b>		
Food and non-alcoholic beverages		
Alcohol and tobacco		
Clothing and footwear		
Housing (rent and utilities)		
Housing equipment		- from own production
Health	Cash expenditures	- from employer
Transport		- from State
Communication		- from privates (gifts or charity)
Recreation and culture		
Education		
Restaurants and hotels		
Miscellaneous goods and services		
<b>= TOTAL CONSUMPTION</b>		

# Consumption and income overlap

CONSUMPTION	CASH	NON CASH		
	(goods and services paid for)	(goods and services not paid for)		
Food and non-alcoholic beverages	Cash expenditures	<ul style="list-style-type: none"> <li>- from own production</li> <li>- from employer</li> <li>- from State</li> <li>- from privates (gifts or charity)</li> </ul>		
Alcohol and tobacco				
Clothing and footwear				
Housing (rent and utilities)				
Housing equipment				
Health				
Transport				
Communication				
Recreation and culture				
Education				
Restaurants and hotels				
Miscellaneous goods and services				
<b>= TOTAL CONSUMPTION</b>				

		CASH	NON CASH
<b>LABOUR INCOME</b>	Dependent employment	<i>Wages, salaries, bonuses</i>	<i>In-kind earnings</i>
	Self-employment	<i>Profits and losses</i>	<i>Own consumption</i>
<b>+ CAPITAL INCOME</b>	Financial investment	<i>Interest and dividends</i>	-
	Real estate investment	<i>Rental income</i>	<i>Imputed rent</i>
<b>+ PENSIONS</b>	Non-contributory	<i>Universal and assistance</i>	-
	Contributory	<i>Work-related insurance</i>	-
	Private	<i>Occupational and individual</i>	-
<b>+ PUBLIC BENEFITS</b>	Family benefits	<i>Maternity, children</i>	<i>STIK</i>
	Unemployment benefits	<i>Insurance and assistance</i>	
	Sickness and disability	<i>Sick pay, work injury, disability</i>	
	Housing	<i>Housing, heating</i>	
<b>+ PRIVATE TRANSFERS</b>	General	<i>Minimum income guarantee</i>	<i>In-kind assistance</i>
	Private institutions	<i>Scholarships, charity</i>	<i>In-kind assistance</i>
	Inter-household	<i>Alimony, remittances</i>	<i>Gifts</i>
<b>= TOTAL GROSS INCOME</b>			
<b>- DEDUCTIONS</b>		<i>Income taxes</i>	
		<i>Social security contributions</i>	
<b>= HOUSEHOLD DISPOSABLE INCOME</b>			

# LIS Consumption/Expenditure

<b>CONSUMPTION EXPENDITURE</b>	Food and non-alcoholic beverages Alcohol and tobacco Clothing and footwear Housing (actual rent and utilities) Housing equipment Health Transport Communication Recreation and culture Education Restaurants and hotels Miscellaneous goods and services
<b>= TOTAL CONSUMPTION EXPENDITURE</b>	

Incl. items consumed and not paid for (own consumption, gifts, etc.) but excl. imputed rent

<b>IMPUTED RENT</b>	Imputed rent
---------------------	--------------

<b>NON CONSUMPTION EXPENDITURE</b>	Income taxes Social security contributions Property taxes Voluntary contributions Inter-household transfers paid Mortgage installment Installment for other loans
------------------------------------	---

# LIS Housing costs

<b>CONSUMPTION EXPENDITURE</b>	Food and non-alcoholic beverages Alcohol and tobacco Clothing and footwear <b>Housing (actual rent and utilities)</b> Housing equipment Health Transport Communication Recreation and culture Education Restaurants and hotels Miscellaneous goods and services
<b>= TOTAL CONSUMPTION EXPENDITURE</b>	

<b>IMPUTED RENT</b>	<b>Imputed rent</b>
---------------------	---------------------

<b>NON CONSUMPTION EXPENDITURE</b>	Income taxes Social security contributions Property taxes Voluntary contributions Inter-household transfers paid <b>Mortgage installment</b> Installment for other loans
------------------------------------	--

Housing costs

# The choice of welfare measures in LIS/LWS

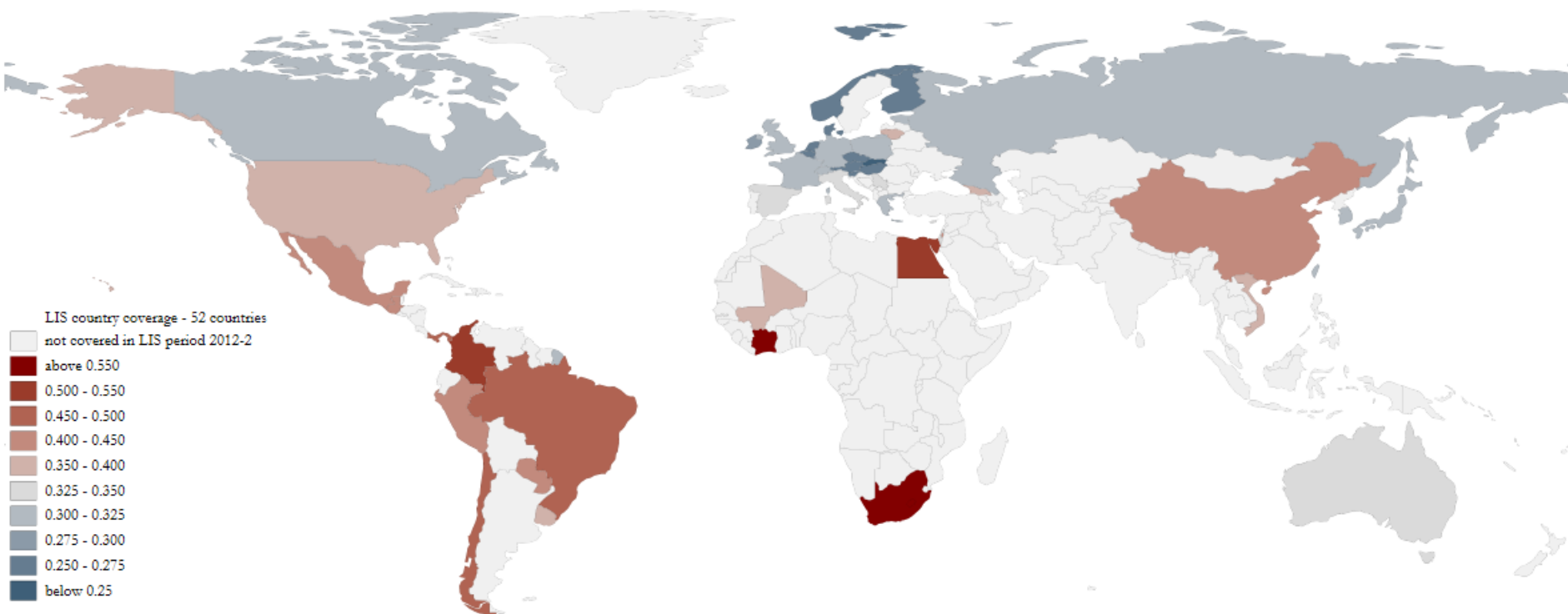
## Alternative measures based on incomes

- *Total income – before deduction of taxes and contributions*
- *Disposable household cash income*
- *Disposable household income (DHI)*
- *DHI\_extended (DHI + imputed rent)*
- *DHI\_reduced (DHI - monetary housing cost)*

Note, this choice depends on data availability / comparability / research question

# The choice of welfare measures in LIS: household disposable income

Luxembourg Income Study (LIS) Database - Gini coefficient, latest year available



Notes: data refer to disposable household income equalised by square root scale.

Wave XI data refer to years 2018-2020: AUS AUT CAN CHE COL DEU FRA GBR GEO IRL ISR LTU LUX MEX MLI NLD NOR PER POL PRY RUS SVK URY USA

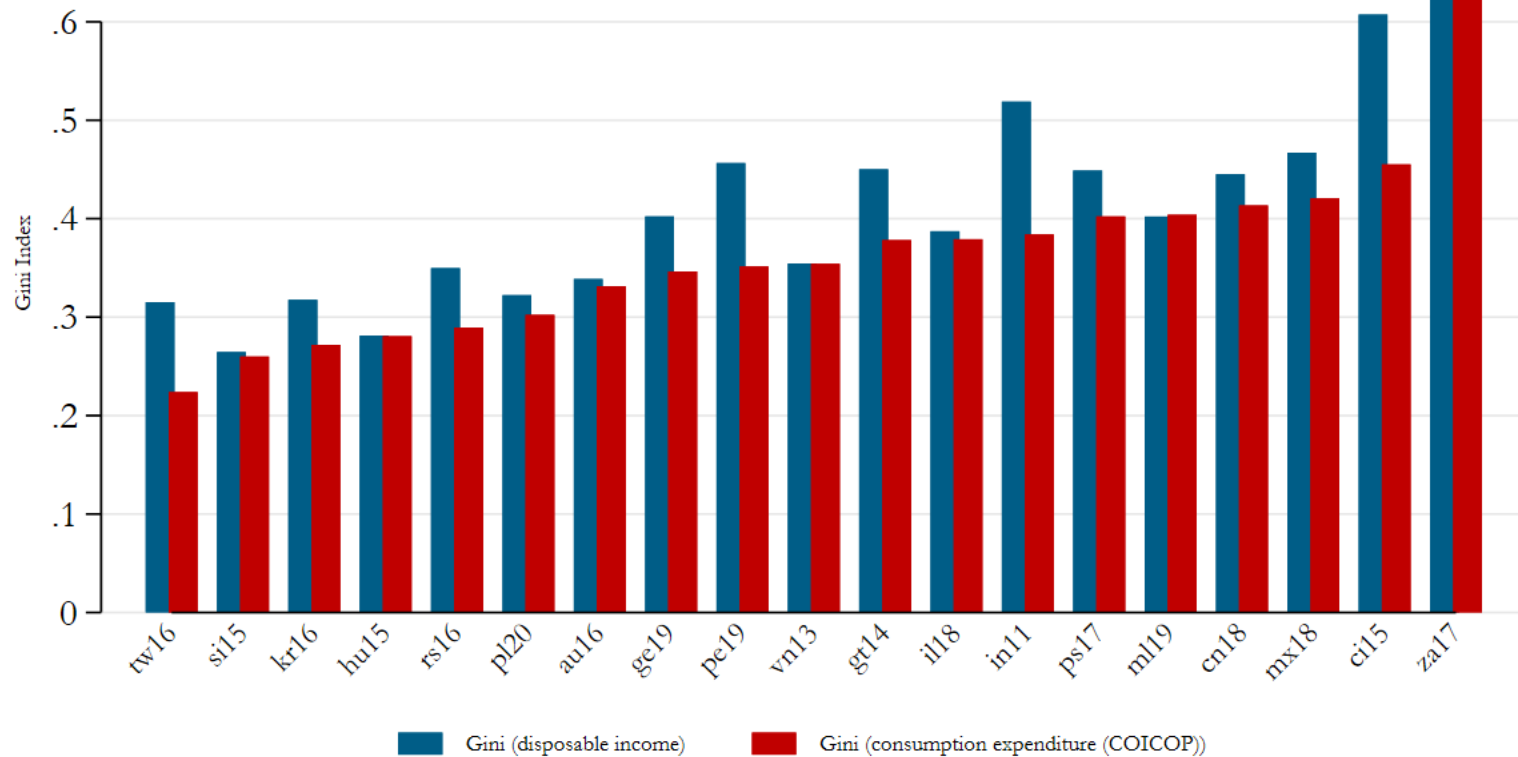
Wave X data refer to years 2015-2017: BEL BRA CHL CIV CZE DNK ESP EST FIN GRC HUN ITA KOR PAN PSE SRB SVN TWN ZAF

Wave IX data refer to years 2012-2014: CHN EGY GTM JPN VNM

Source: Luxembourg Income Study (LIS) Database, June 2022.

# The choice of welfare measures in LIS: disposable income vs. consumption expenditure

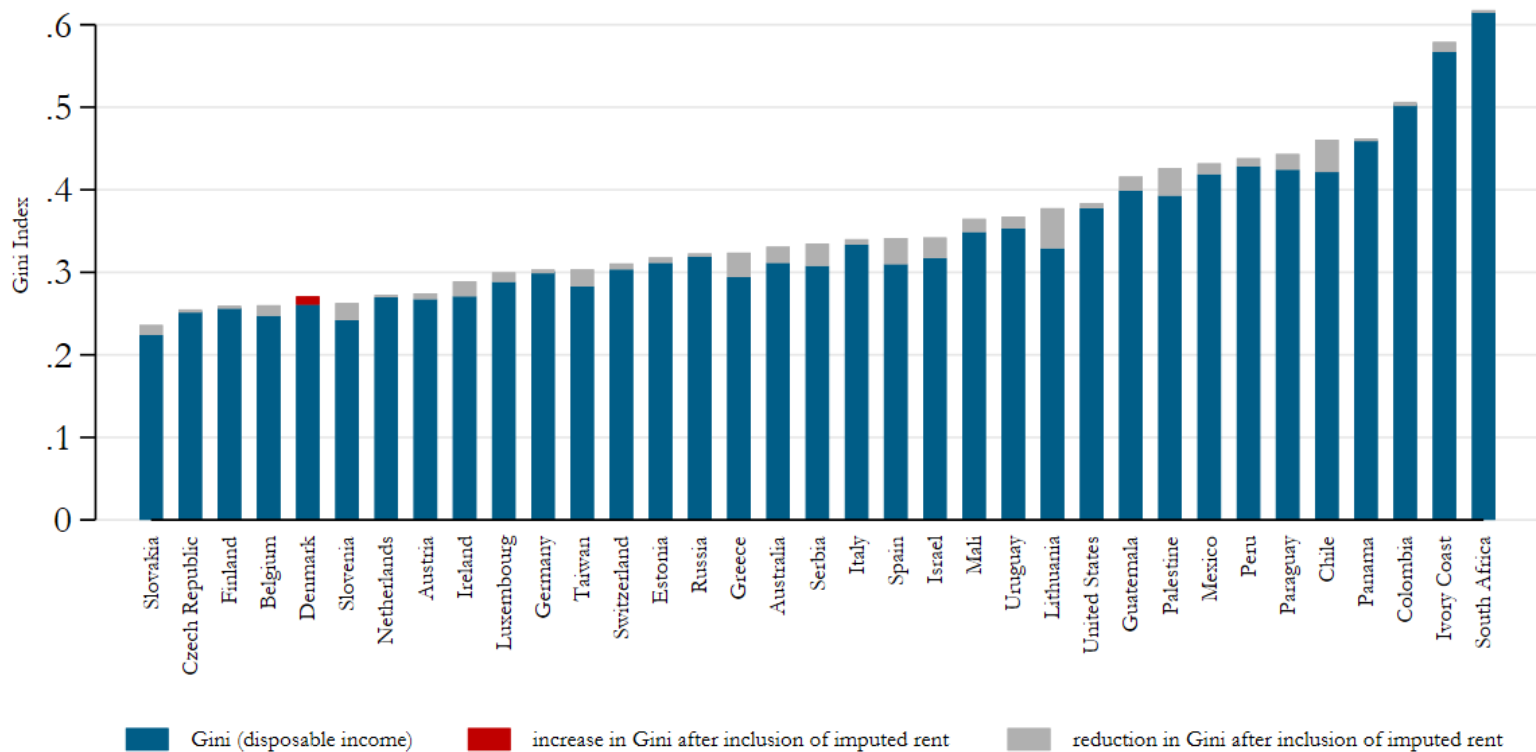
Luxembourg Income Study (LIS) Database - Gini Index LIS, latest year available  
disposable income vs. consumption expenditure (COICOP)



Notes: data refer to consumption expenditure equivalised by per capita scale.  
Source: Luxembourg Income Study (LIS) Database, Sep 2022.

# The choice of welfare measures in LIS: impact of adding imputed rents

Luxembourg Income Study (LIS) Database - Gini Index LIS, latest year available  
disposable income vs. disposable income including imputed rent



Notes: data are equalised by square root scale.

Wave XI data refer to years 2018-2020: AUS AUT CHE COL DEU IRL ISR LUX MEX MLI NLD PER PRY RUS SVK URY

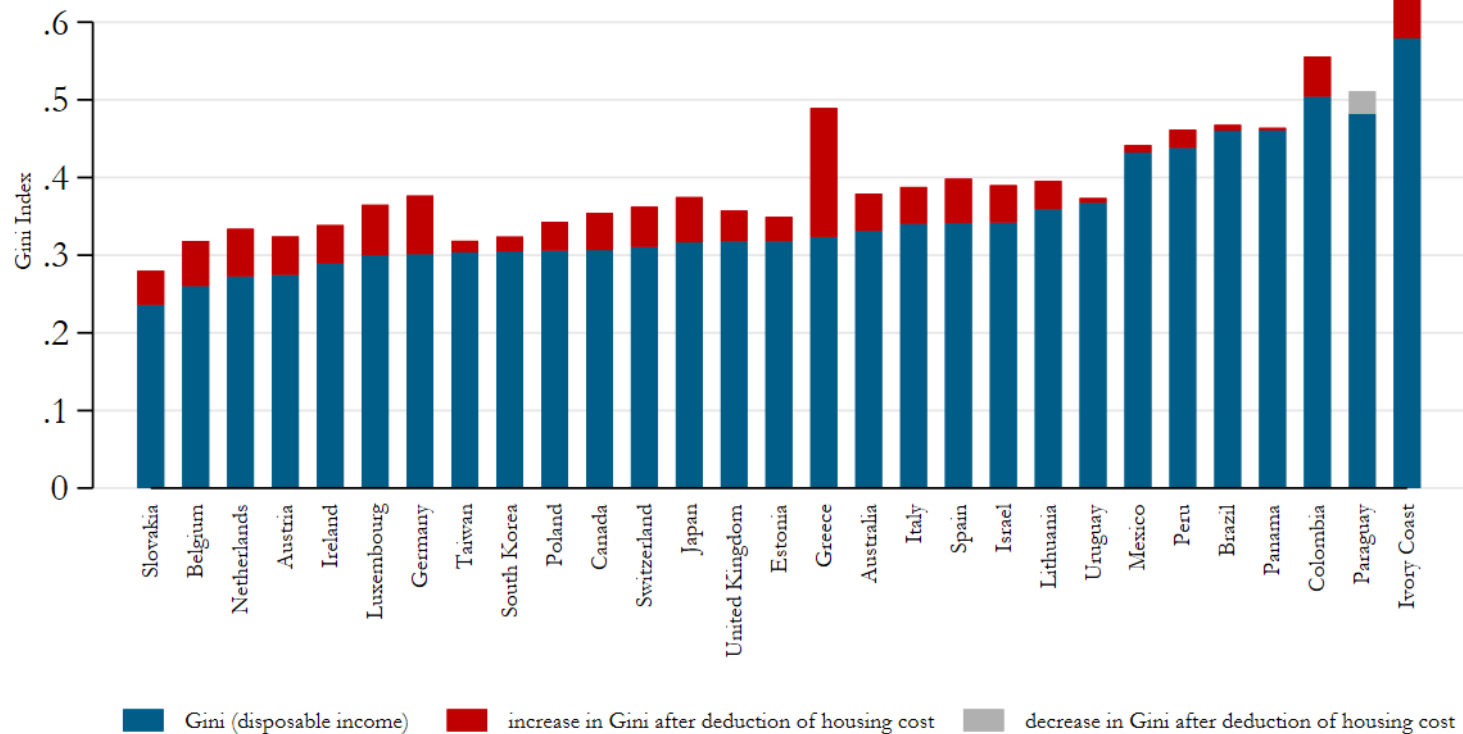
Wave X data refer to years 2015-2017: BEL CHL CIV CZE DNK ESP EST FIN GRC ITA PAN PSE SRB SVN TWN USA ZAF

Wave IX data refer to years 2012-2014: GTM LTU

Source: Luxembourg Income Study (LIS) Database, June 2022.

# The choice of welfare measures in LIS: impact of accounting for housing costs

Luxembourg Income Study (LIS) Database - Gini Index LIS, latest year available  
disposable income vs. disposable income after housing cost



Notes: data are equivalised by square root scale.

Wave XI data refer to years 2018-2020: AUS AUT CAN CHE COL DEU GBR IRL ISR LTU LUX MEX NLD PER POL SVK URY

Wave X data refer to years 2015-2017: BEL CIV ESP EST GRC ITA KOR PAN TWN

Wave IX data refer to years 2012-2014: BRA JPN PRY

Source: Luxembourg Income Study (LIS) Database, June 2022.



# Different income concepts: measure of redistribution

In order to measure the redistribution effect of the government interventions through public transfers and taxes, we need to consider two income concepts

Market income GINI

-

DHI GINI

=

## Market income

(Labour + Capital +  
Occupational and  
private pensions +  
Private transfers)

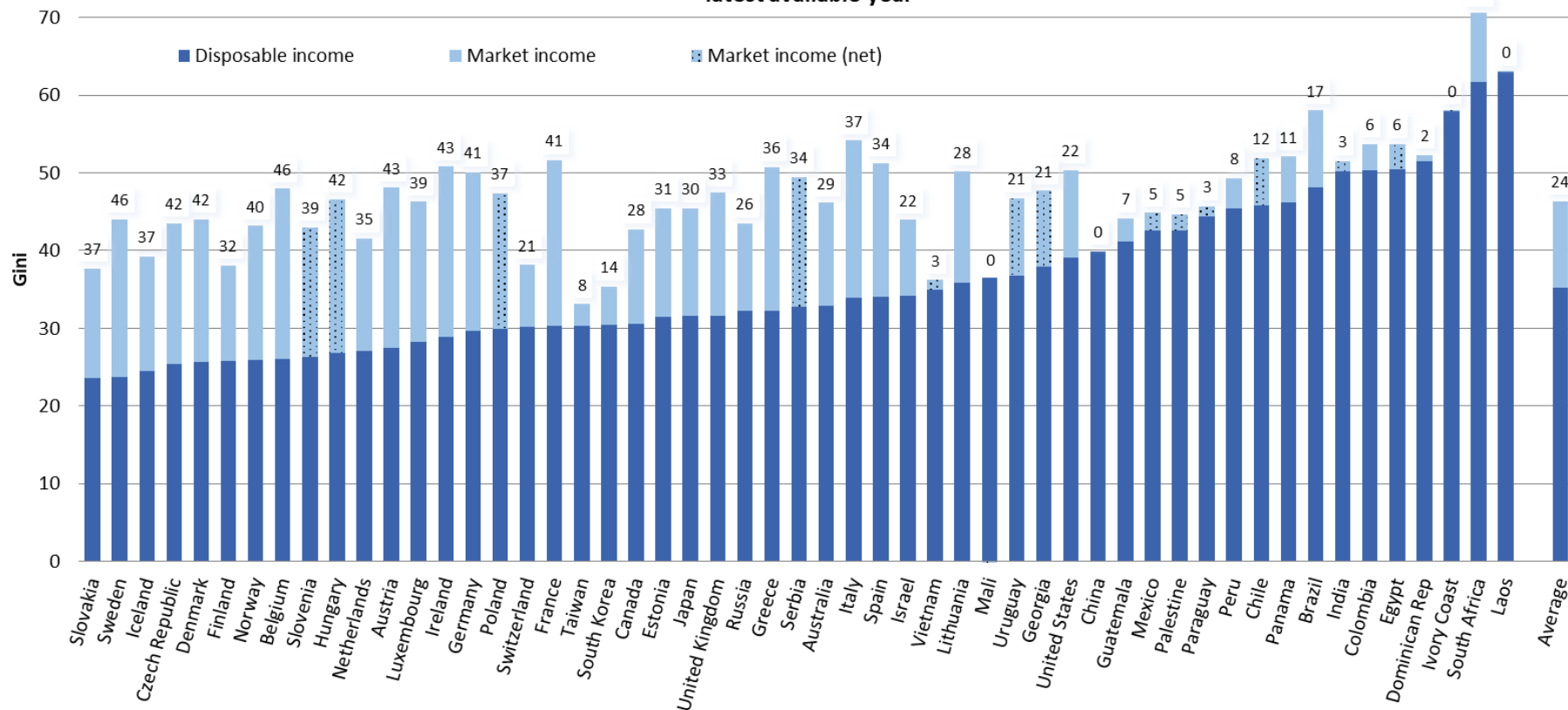
## Disposable Household Income (DHI)

(Market income +  
Public pensions +  
Other social security) –  
(Taxes + Contributions)

A measure of the  
redistributive impact  
of social security and  
direct taxation  
systems in a country

# Inequality reducing effect of redistribution of public transfers and taxes

Gini Index on Market Income and Disposable Income and percent reduction latest available year



Market income is defined as the sum of labour income, capital income, occupational and private pensions and private transfers. The number at the top of the bar represents the percent reduction of Gini after redistribution of public transfers and taxes. Countries denoted Market Income (net) are countries where only net incomes are available (so that redistribution includes transfers only).

Source: Luxembourg Income Study (LIS) Database, <http://www.lisdatacenter.org>. March 2022. Luxembourg: LIS.

# Different income concepts: a different type of redistribution

In order to measure the redistribution effect of the non formal interventions (imputed earnings through own consumption, remittances and other private transfers not regulated by the state), we need to consider two income concepts

Formal market income GINI

## Formal market income

(Cash labor earnings + Capital + Occupational and private pensions)

-

Market income GINI

## Market income

(Formal market income + Own consumption + Private transfers)

-

DHI Gini

## Disposable Household Income

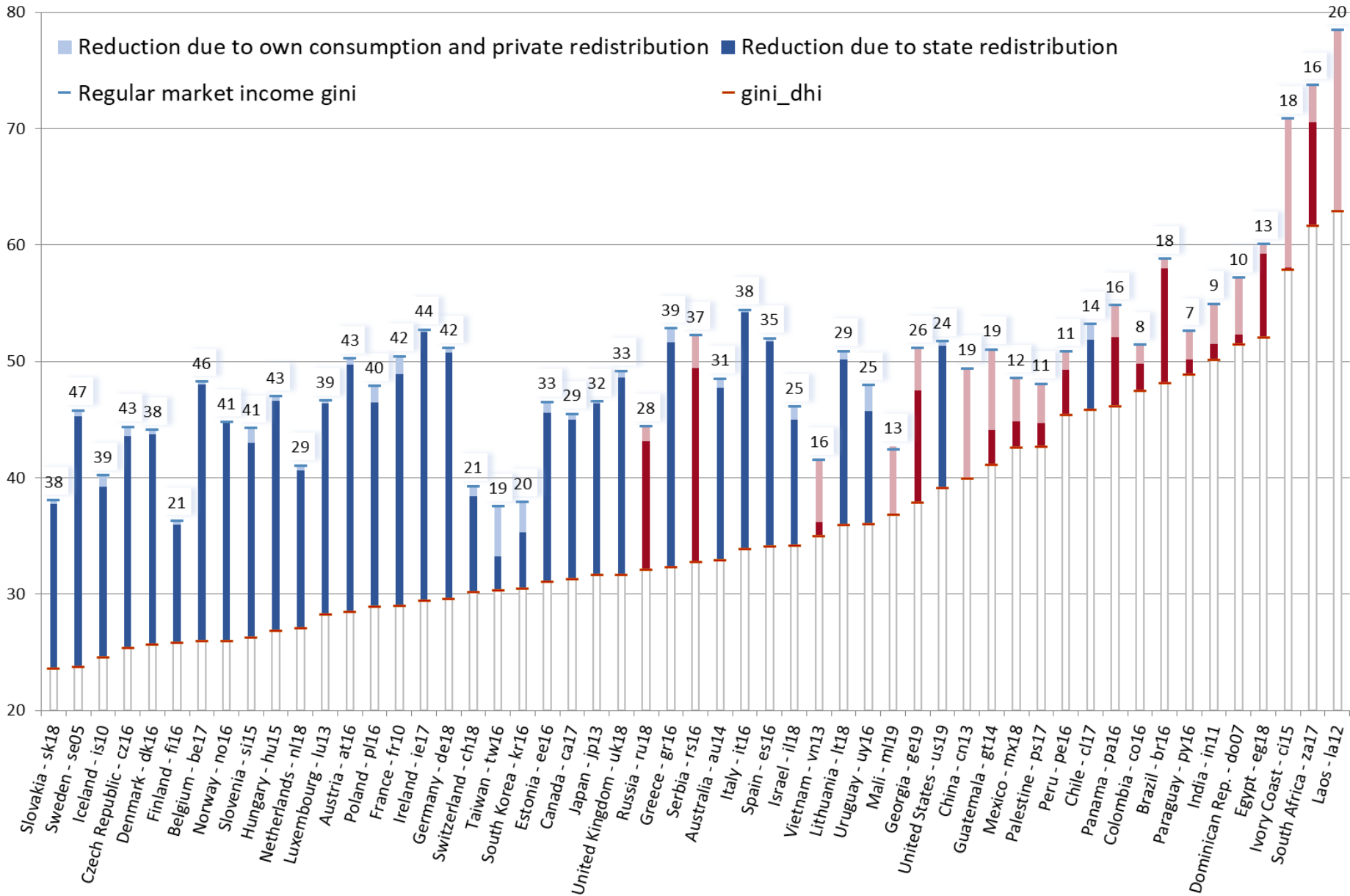
(Market income + Public pensions + Other social security) – (Taxes + Contributions)

A measure of the redistributive impact of non formal market income

A measure of the redistributive impact of social security and direct taxation systems in a country

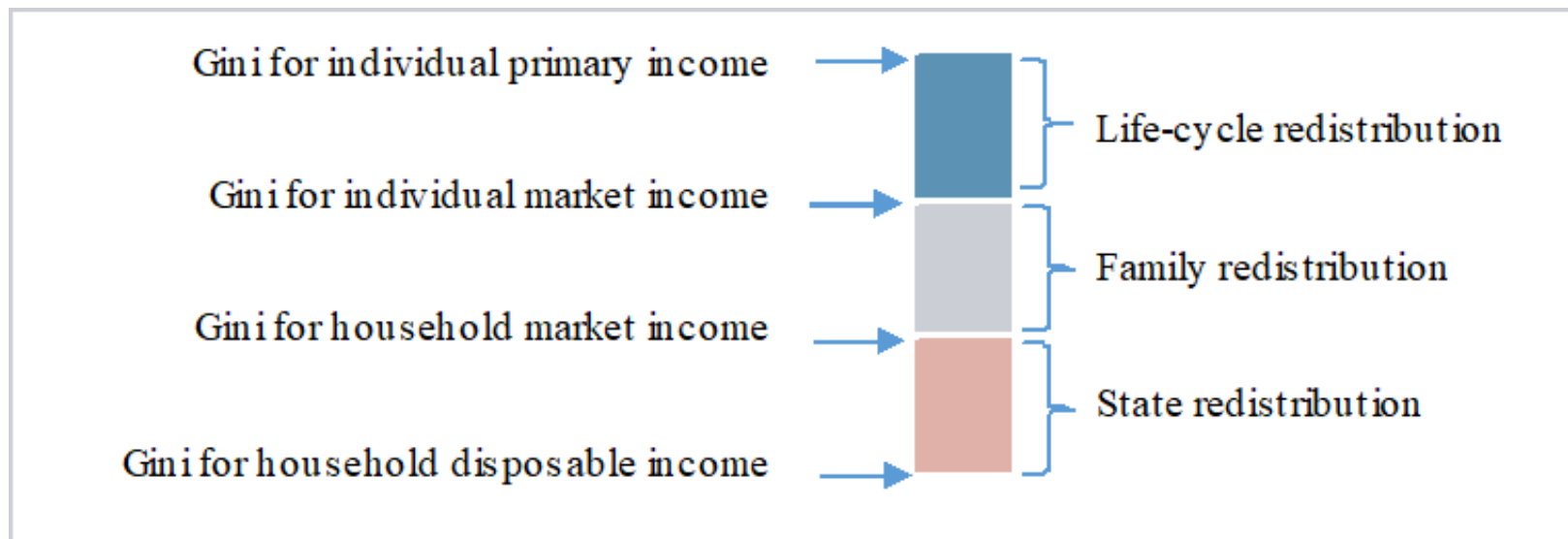


# Gini Index on Regular Market versus Disposable Income



Regular market income is defined as labour income (excluding own consumption) and capital income. The number at the top of the bar represents the percent reduction of Gini after own consumption and redistribution.

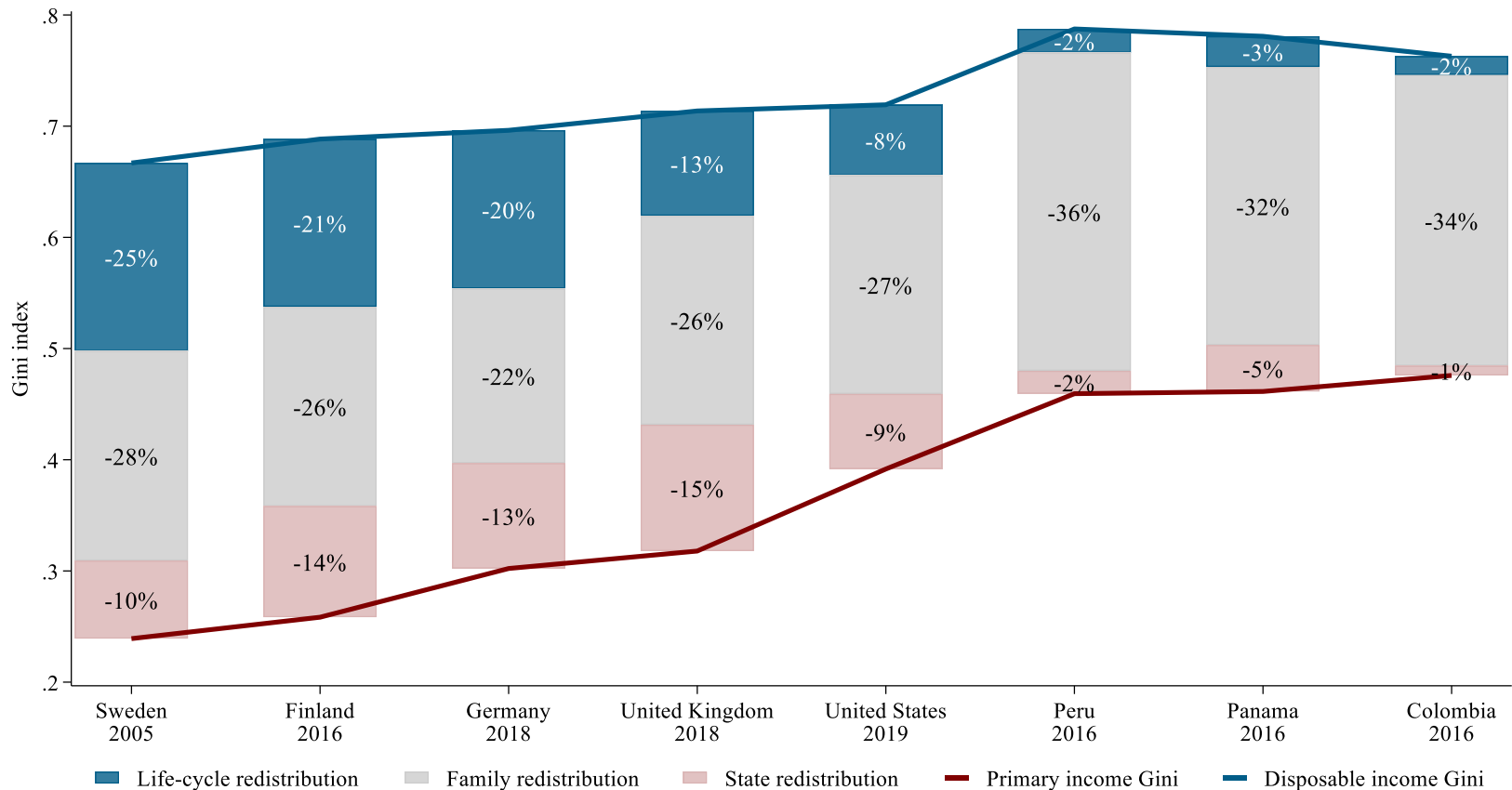
# Different redistribution channels



- **Redistribution over the life-cycle:** inter-temporal redistribution at individual level redistribution due reduction in income during activity (payment of social security contributions ) and increase in income during inactivity (receipt of individual insurance transfers)
- **Family redistribution:** effect of pooling resources within family and of inter-household transfers
- **State redistribution:** effect of including universal and assistance public benefits and deducing income taxes

# Different redistribution channels

Inequality reducing effect of intertemporal, family and state redistribution

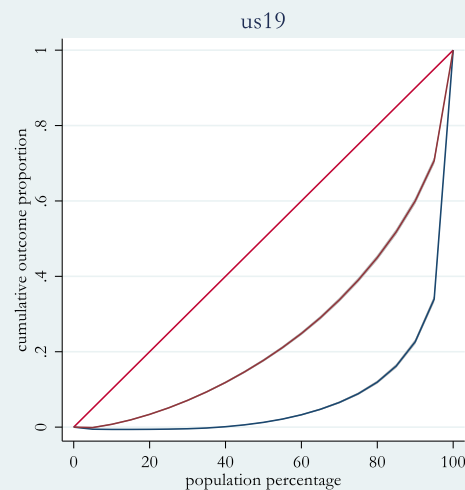
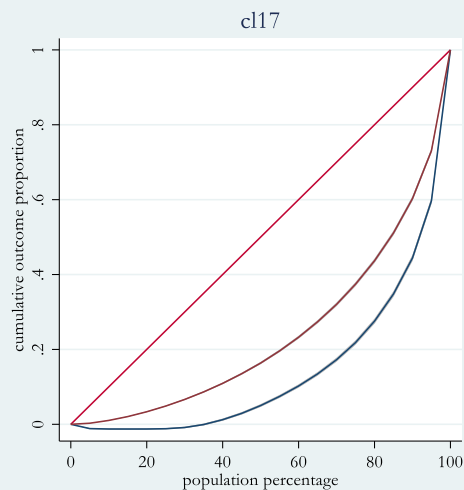
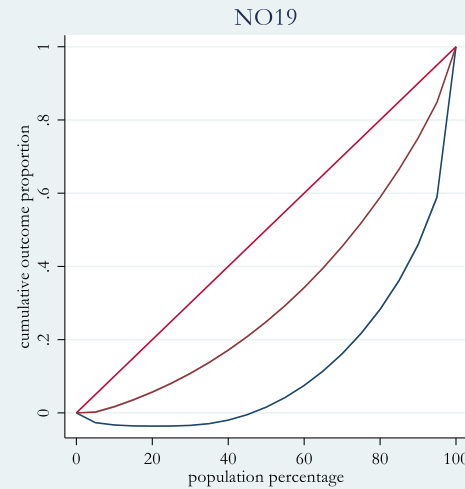
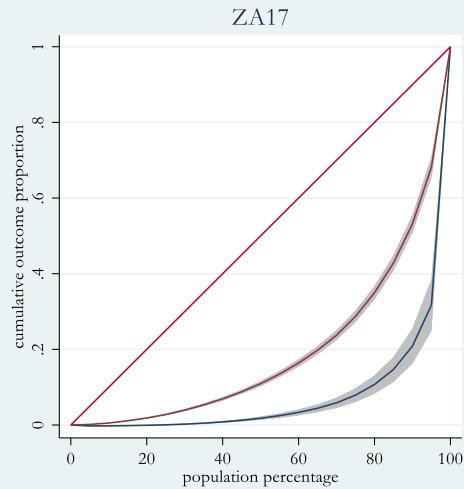


# LWS Household Net Worth

NON-FINANCIAL ASSETS	Real-estate Non-housing assets
+ FINANCIAL ASSETS	Deposit accounts and cash Financial investments Other non-pension financial assets
- TOTAL LIABILITIES	Real estate liabilities Non-housing liabilities
<b>= DISPOSABLE NET WORTH</b>	
+ LIFE INSURANCE AND VOLUNTARY INDIVIDUAL PENSIONS	
<b>= ADJUSTED DISPOSABLE NET WORTH</b>	
+ OTHER PENSIONS	Occupational pensions Social Security pension entitlements
<b>= TOTAL NET WORTH</b>	

# The choice of welfare measures in LIS/LWS: household net worth vs. disposable income

Lorenz curve for income and wealth – South Africa, Norway, Chile and USA



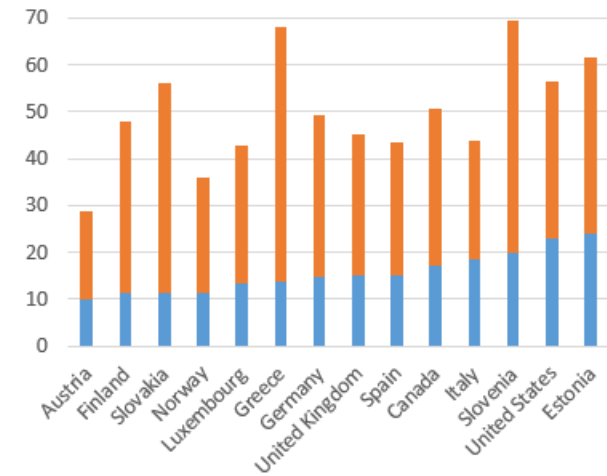
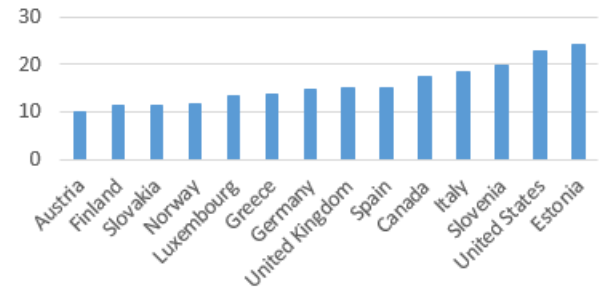
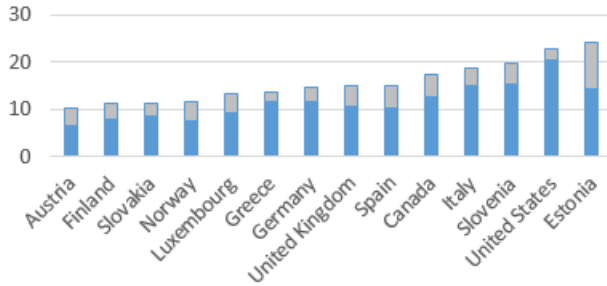
— disposable net worth

— disposable income



# The ranking of countries: Combining income and asset poverty

source: Luxembourg Wealth Study (LWS) Database

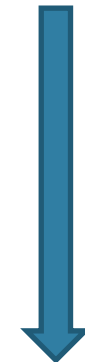


Income poor **AND** asset poor



reduced by 'vulnerable' group 2  
*income poor, but not asset poor*

Income poor

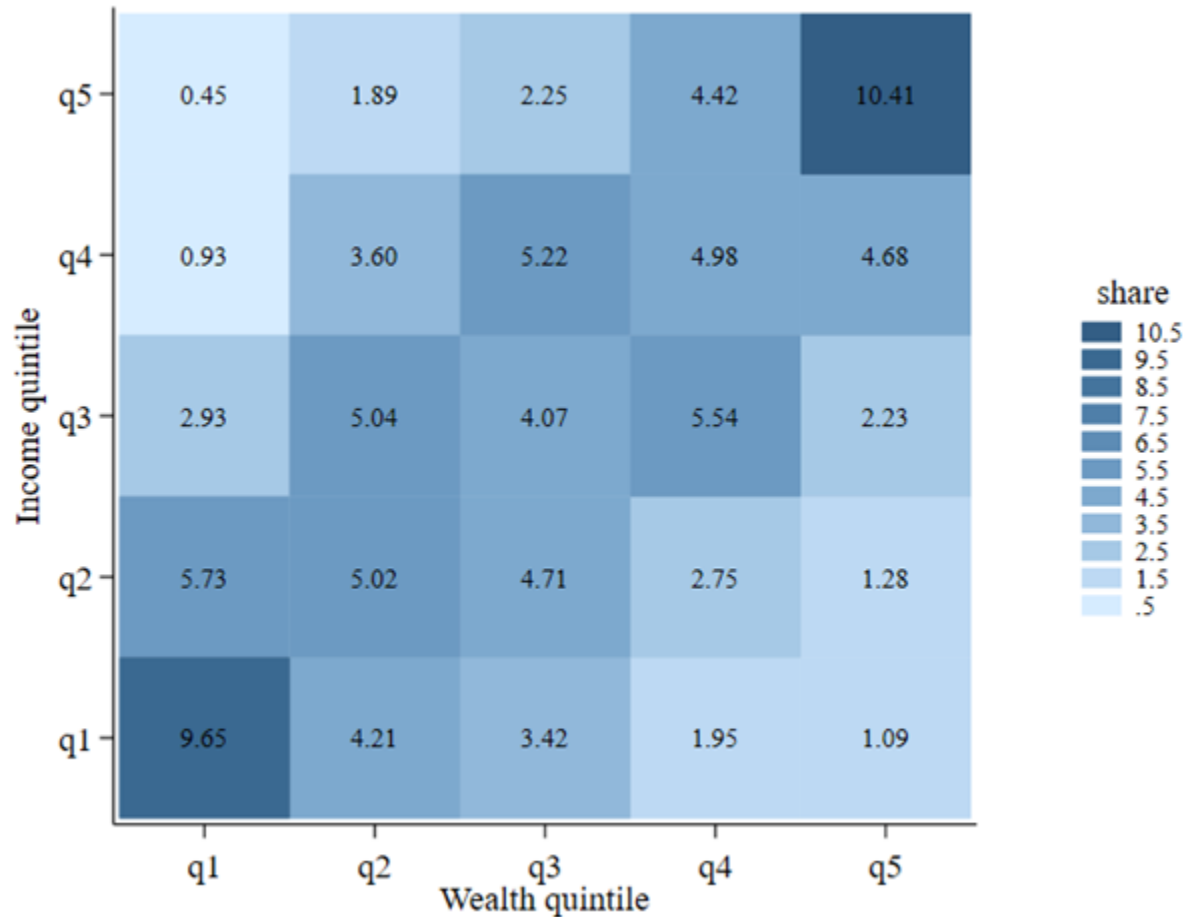


extended by 'vulnerable' group 1  
*asset poor, but not income poor*

Income poor **OR** asset poor

# The choice of welfare measures in LIS/LWS: household net worth vs. disposable income

Income/Wealth heatmap for Luxembourg 2018 (shares of the household population)



# Contents

- Overview of LIS and the LIS data
- Data harmonisation and dissemination
- Different welfare concepts
- **Inclusion of middle/low-income countries**
- An example: harmonization of income data from Mali

# Inclusion of middle/low-income countries in LIS

- **Historically:** focus on high-income countries
- **2007:** pilot project on the feasibility of including **upper-middle income countries** (5 Latin American countries from World Bank MECOVI)
- **2011: new template** adjusted to better accommodate middle-income countries
  - *inclusion of non-monetary income into DHI*
  - *adjustment of the concept of household member and addition of living arrangement variables*
  - *new topics: rural/urban indicator, farming activity indicator, type of dwelling, involvement in marginal/informal work, characteristics of second job*
- **Since 2018:** collaborations on **lower-middle- and low-income countries**
  - *World Bank: project to study feasibility of integrating LSMS data into LIS*
  - *ERF: harmonisation of data from 7 Arab countries*
  - *AFD: harmonisation of data from Côte d'Ivoire, Viet-Nam, Mali*
  - *STATEC/GOPA: harmonisation of data from Laos*
- **2021: new top-coding methodology** to better accommodate countries with very high inequality



# Middle- (and low-) income countries in LIS

High-income countries (N=33):			Upper-middle-income countries (N=16):		Lower-middle-income countries (N=8):	Low-income countries (N=3):
Australia	Hungary	Romania	Brazil	Jordan	(Bangladesh)	Mali
Austria	Iceland	Slovakia	China	Mexico	Côte D'Ivoire	Somalia*
Belgium	Ireland	Slovenia	Colombia	Paraguay	Egypt	Sudan
Canada	Israel	South Korea	Dominican Rep.	Peru	India	
Chile	Italy	Spain	Georgia	Russia	(Laos)	
Czech Rep.	Japan	Sweden	Guatemala	Serbia	Pakistan	
Denmark	Lithuania	Switzerland	Iraq	South Africa	Tunisia*	
Estonia	Luxembourg	Taiwan			Vietnam	
Finland	Netherlands	United Kingdom				
France	Norway	United States				
Germany	Panama	Uruguay				
Greece	Poland					

\* Consumption only (ERF-LIS dataset)

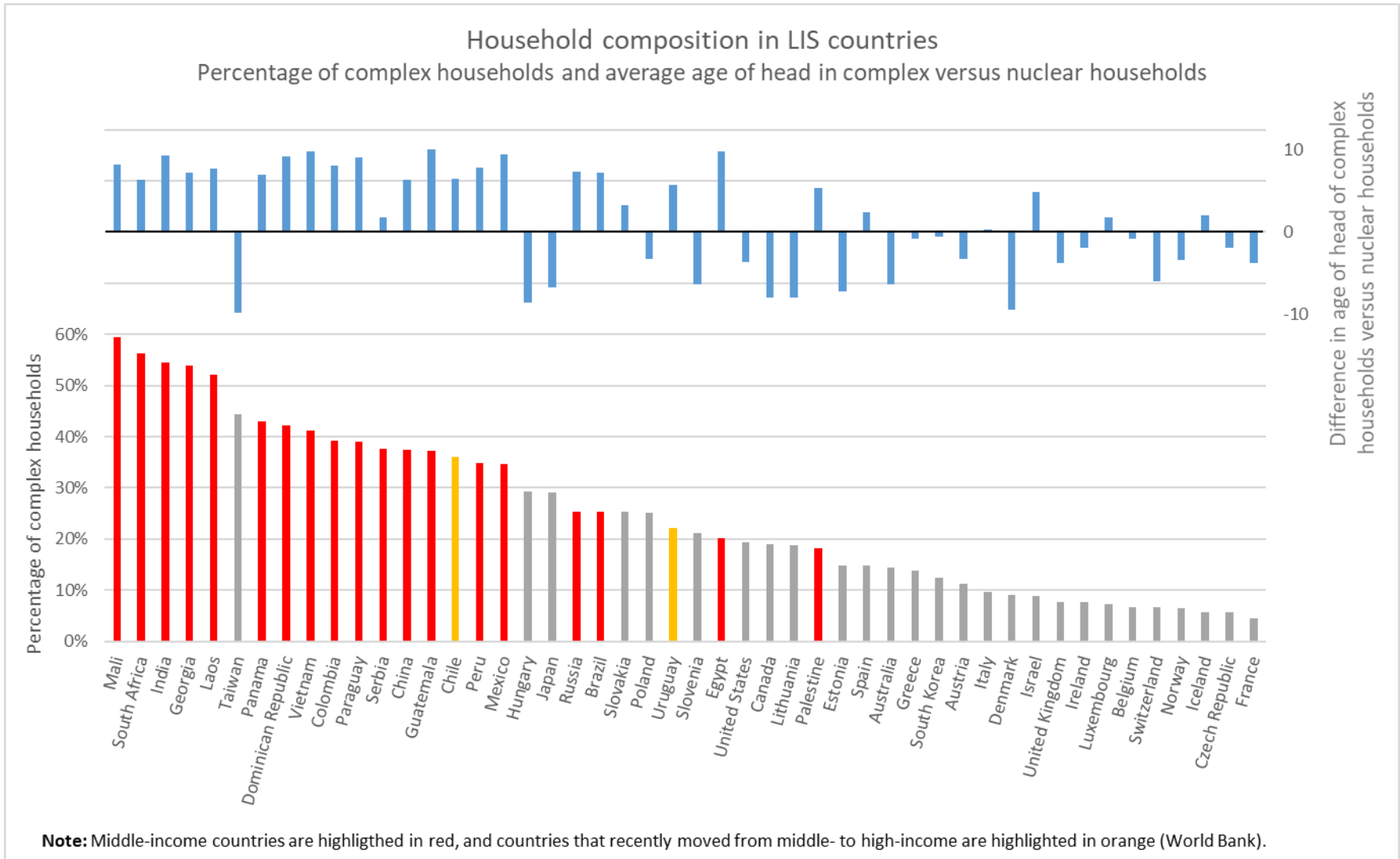
Country classification follows the 2022 World Bank Country classification by income



# Challenges of harmonising middle-income countries

- **Survey coverage and weights:** only certain regions/areas (e.g. urban, war zones), dissemination of partial sample, problems with corresponding weights
- **Household composition:** prevalence of complex household structures, identification of the household members
- **Labour market:** prevalence of informal/marginal employment
- **Income and wealth** (newly collected in many M/L income countries):
  - *Exhaustiveness (some incomes not collected at all, not collected for some years, not collected at the individual level or for some subgroups of the population)*
  - *Measurement (different methods, under-measurement, double counting, seasonality)*
  - *Incompleteness (observations with missing and zero values, hardly ever imputed)*
  - *Over-time (in)consistency (with big jumps from one year to the next)*
  - *Treatment of taxes and social security contributions*

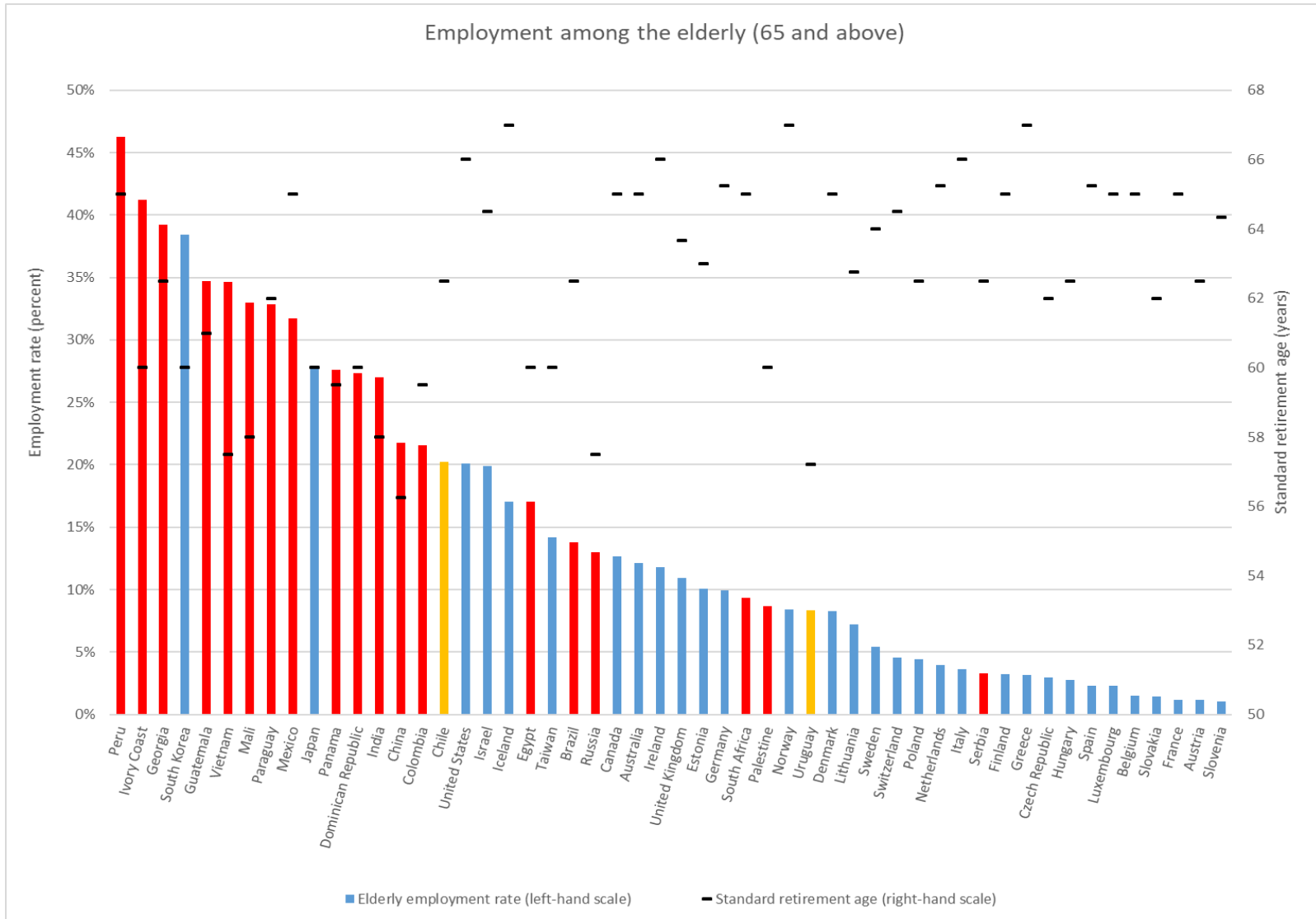
# Household composition: prevalence of complex household structures



## Issues:

- Difficult to identify household composition
- Analyses centered around the head/spouse may become biased

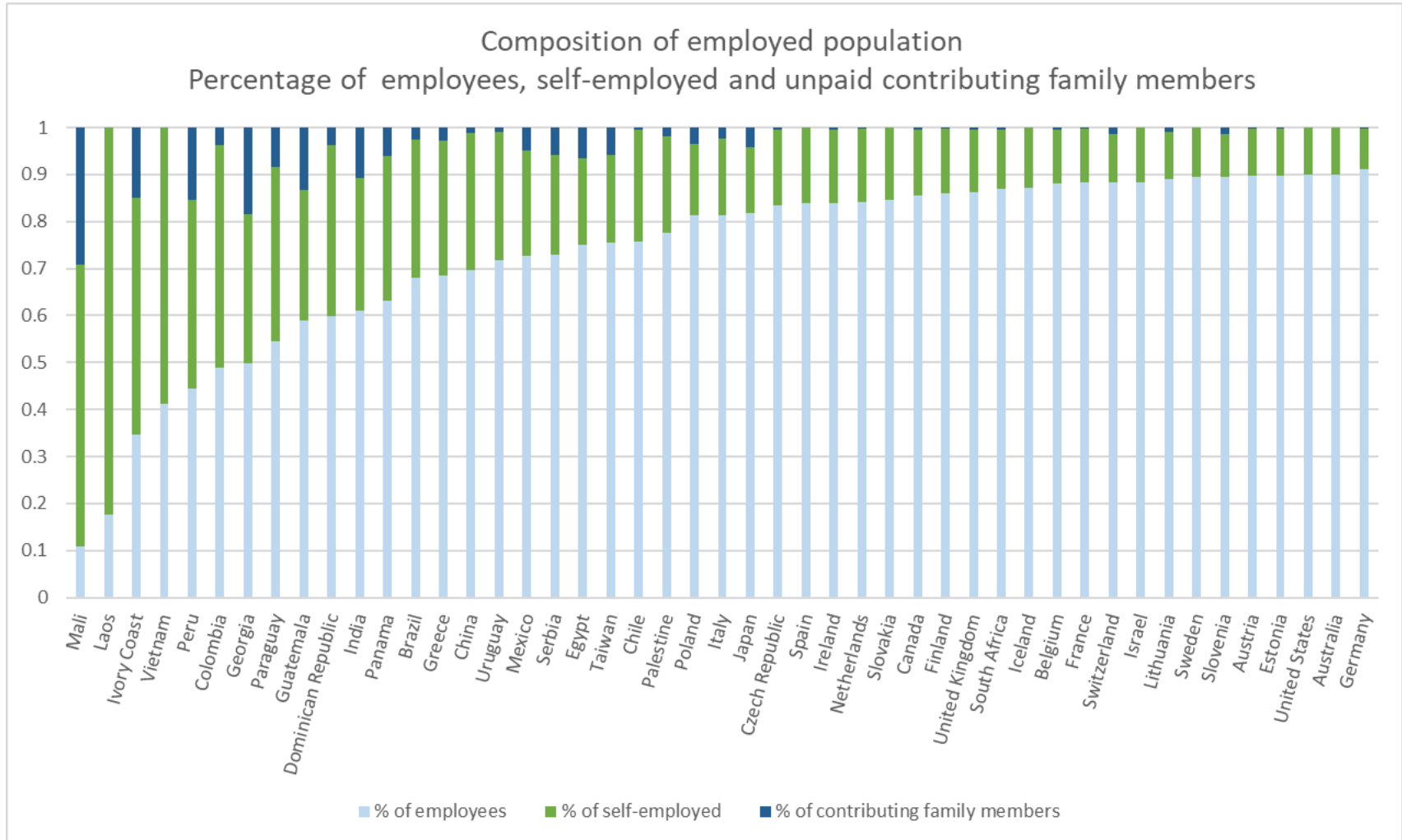
# Employment: prevalence of informal employment (elderly, child, unpaid)



## Issues:

- measurement of employment (artificially high)
- inconsistency between employment status and earnings

# Prevalence of self-employment



All middle- and low-income countries (with the notable exception of South Africa) are grouped on the left-hand side of the chart

# Treatment of non-cash incomes

- ❑ Relative importance of non-cash incomes (consumption from own production, in-kind individual goods received) w.r.t. high income countries
- ❑ LIS now includes those in total disposable income

## BUT

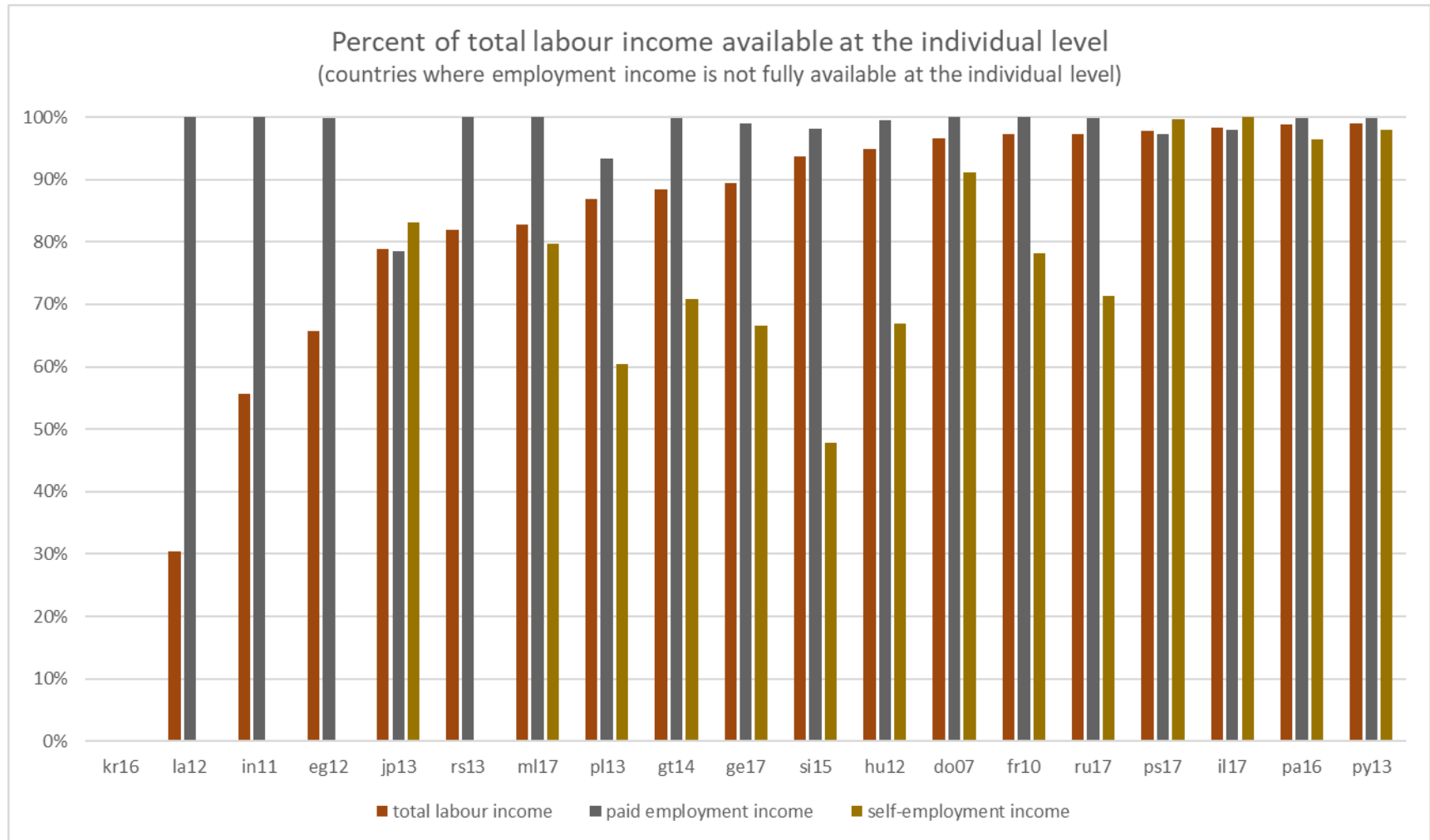
- ❑ Difficulty of **measurement**
- ❑ Not always available (**incomparability**)
- ❑ **Risk of double-counting** when in-kind incomes are collected in the consumption module and added to the incomes

# Measurement of self-employment incomes

Issue is of particular relevance in middle income countries as self-employment is much more common

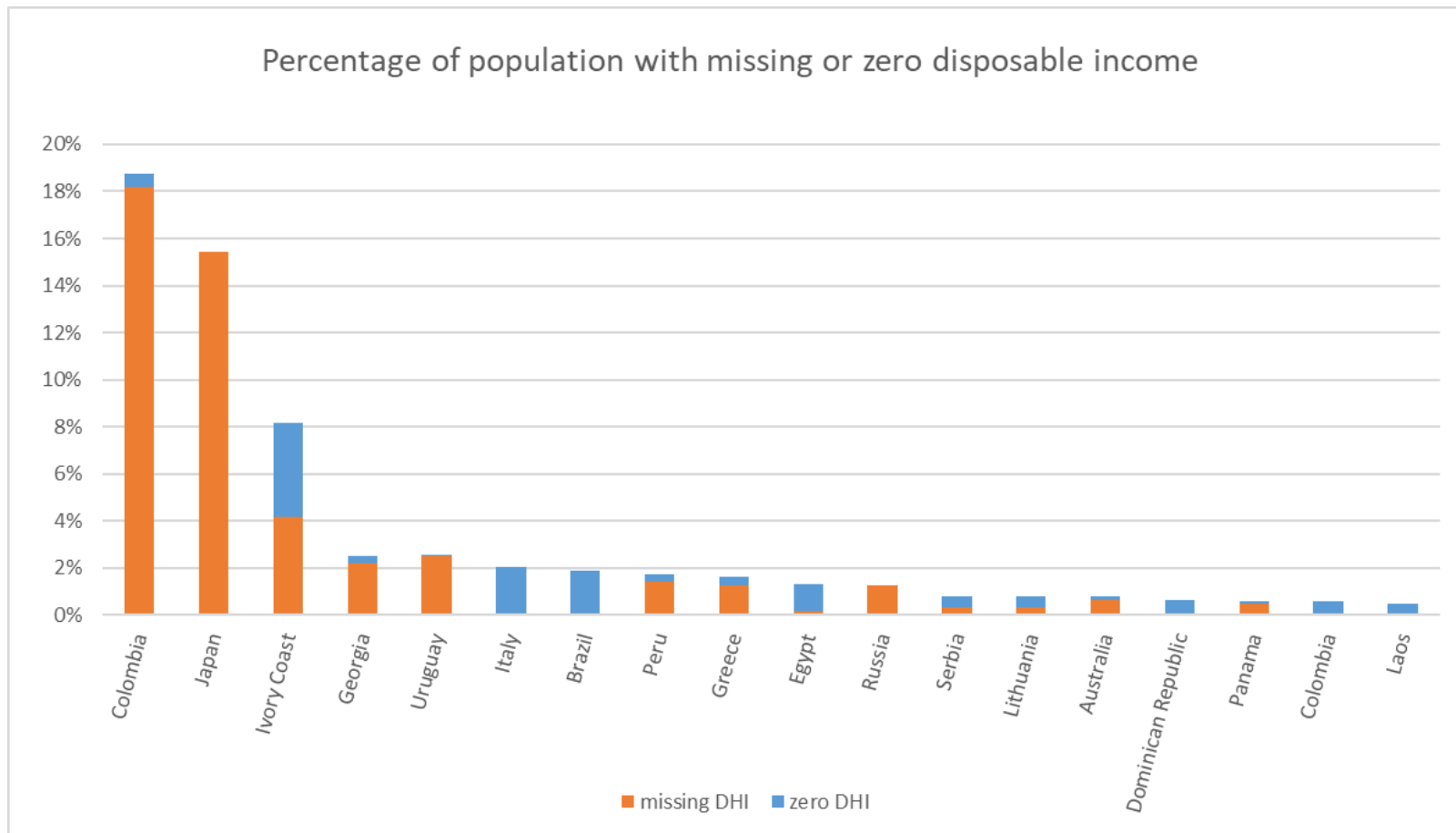
- ❑ **Underestimation**: self-employment incomes are much harder to capture in general, even more so in presence of informal/marginal employment
- ❑ **Risk of double counting**: when collected both at the individual level and in the household business sections of the questionnaire
- ❑ **Individual level self-employment income not measurable**: small portion collected at the individual level (possibly only the incomes of the liberal professions), while other self-employment incomes are collected at the household level (family businesses)

## Measurement of individual level self-employment income



**With the exception of France and Japan, these are all countries that are or have recently been middle-income countries!**

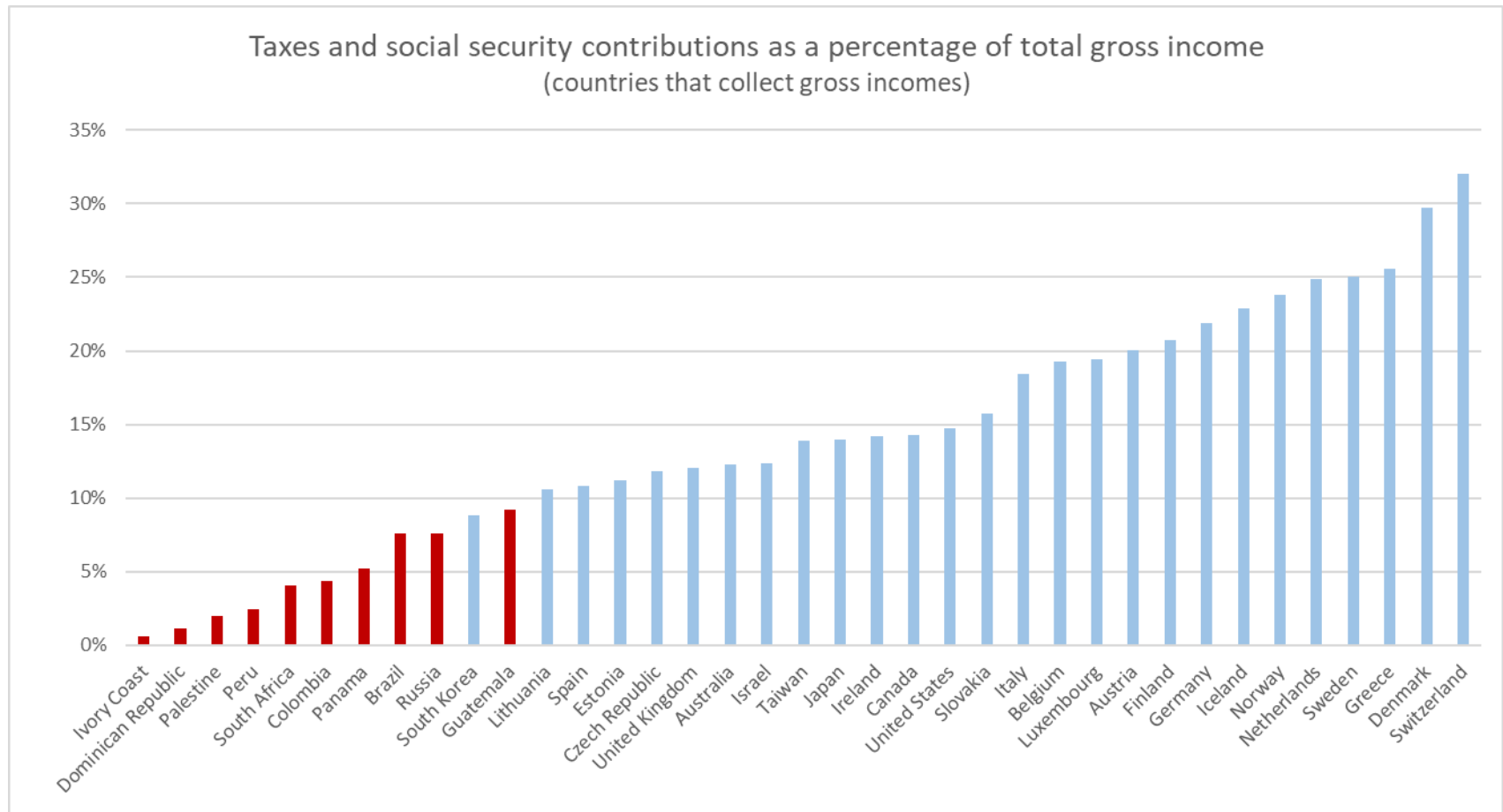
# Incomplete income data is much more prevalent



**Issue:** bias due to non-random distribution of households with no income information



# Measurement of direct taxes and social security contributions



## Issues:

- difficult to measure (typically net only datasets, or – even worse – gross without the indication of the taxes and contributions or undefined)
- higher reliance on indirect taxes (which are typically not as redistributing as the direct ones) biases downwards inequality when compared to countries that rely more on direct taxation

# Contents

- Overview of LIS and the LIS data
- Data harmonisation
- Different welfare concepts
- Inclusion of middle/low-income countries
- **An example: harmonisation of income data from Mali**



# Harmonisation of survey data from Mali

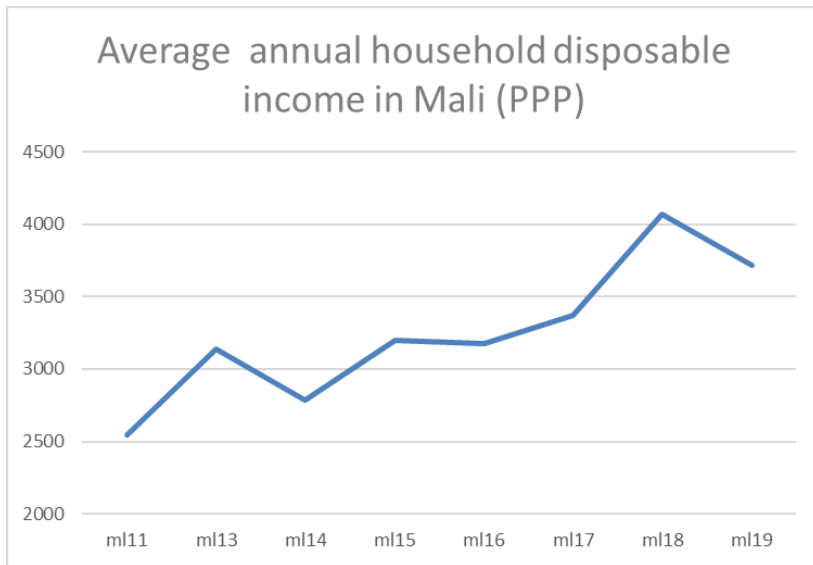
8 waves of the EMOP survey (2011 to 2019) – focused on consumption, including income module since 2011

Project made possible thanks to a collaboration with the AFD

## Challenges of the harmonization

- *Choice of the sample and associate weight*
- *Creation of an exhaustive measure of total disposable income:*
  - *some income sources not available in all years*
  - *many employed persons with zero labour incomes*
  - *many missing income data*
  - *inter-households transfers collected in two different modules*

# Mali harmonised income data

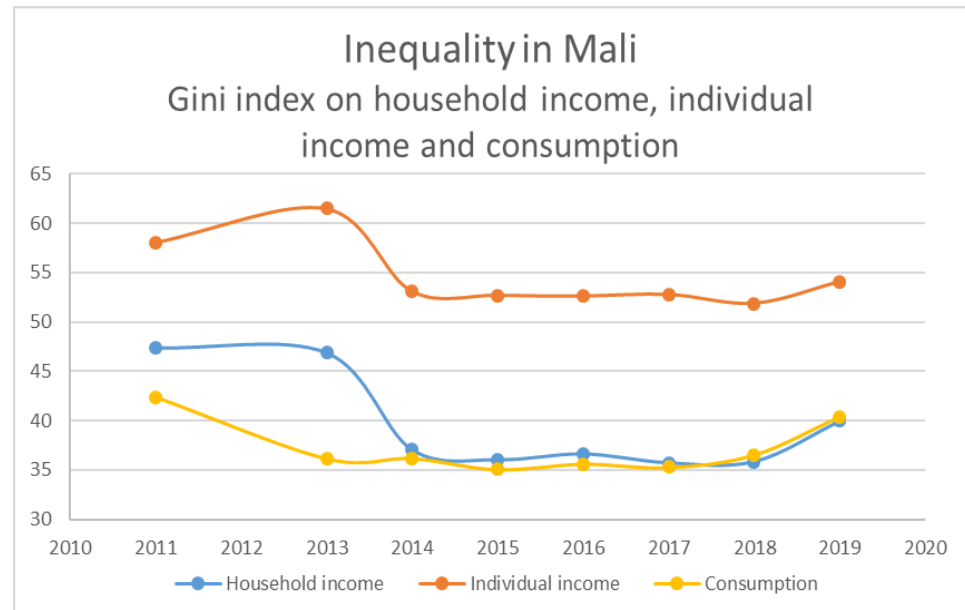


Possible explanations for jumps in 2013 and 2018

- in 2013 smaller coverage due to civil unrest
- in 2018...?

Gini index:

- relatively low level
- income Gini very close to consumption Gini
- very different levels in income of first two waves



# Conclusive remarks

- ❑ Very careful when comparing low- and middle-income countries to high-income countries  
→ Need to adjust the conceptual framework
- ❑ (Ex-post) harmonisation does not imply full comparability  
→ Comparability of underlying data is very important
- ❑ Income-, wealth- and consumption-based estimates often differ greatly (because they measure related but different facets of the same phenomena)
- ❑ Importance of complementing the 3 when creating a measure of well-being