

# WORLD BANK STANDARDIZED DATABASE FOR EASTERN EUROPE AND CENTRAL ASIA – ECAPOV DATABASE

WHY, WHAT AND HOW

João Pedro Azevedo  
Global Lead

Minh Nguyen  
Economist

# What is ECAPOV?

- The ECAPOV (ECA Poverty) is an ex-post harmonization effort for Eastern Europe and Central Asia (ECA) countries based on available household budget surveys (HBS) and Living Standards Measurement Surveys (LSMS).
- There are multiple versions/vintages depending on the exercise and the period of analysis.
- ECAPOV has been founded around 2000 in ECA PREM, was sponsored by the ECA Chief Economists Office.

# Modules in the ECAPOV database

- It has several modules (added some recently)
  - Individual characteristics (module 2)
  - Consumption/welfare module (module 3)
  - Utilities module (module 4)
  - Social Protection and social assistance modules (module 6)
  - Income module (module 7)
  - Assets and services (module 9)
  - National welfare aggregate module (module 10 - new)

# History



# Coverage of the ECAPOV database

*Currently available for....*

## MODULES & VARIABLES

**232** Expenditure

**85** Income

**222** Household Characteristics

**29** countries.

**236** surveys.

**18** years  
(1995-2014).

# What is the purpose?

- Comparable data series within and across countries
- Ensure regional comparability data and statistics
- Enable scalability on the process of data harmonization
- Enable greater efficiency in the use of statistics for regional research and policy dialogue
- Share the knowledge between teams and units
- Provide easy access to the raw/original and harmonized data through the ***datalibweb*** system (Stata-based ado command)

# What are difficulties?

*Components of the harmonization are not strictly comparable in the ECA region because...*

**Countries have different systems:** i.e. Years of required primary education range from 4 to 6.

**Surveys are not kept unchanged through time:** i.e. In MDA 2006 the HBS was substantially modified. Main changes came from sampling frame, sampling areas, and questionnaires.

**Surveys do not have the same modules:** i.e. HBS reports expenditure on health but not prevalence. Only 9 HBS surveys had a health section.

**Questions are not asked in the same way:** i.e. In BIH 2010 the reference period for expenditure during holidays or vacations had changed from “last month” to “12 months.”

**Questions do not have the same categories:** i.e. In BIH 2010 the labor status question had a detailed category non-existing in previous years. Employment had been divided into full-time and part-time employment and unemployment into previously unemployed and seeking first time employment.

# What are the compromises?

## **Prioritize country–time comparison more than cross-country comparison**

The first priority is comparability within all years of available data for each country.



## **Harmonized variables might be different from the NSO definitions.**

This means that indicators could be different and can presents a challenge for discussing policy implication, for example when the poverty rates are different. Transparency is our main remedy.

## **Harmonization is done with available surveys**

Most available surveys in ECA region are HBS surveys design to gather information about expenditure, not consumption.



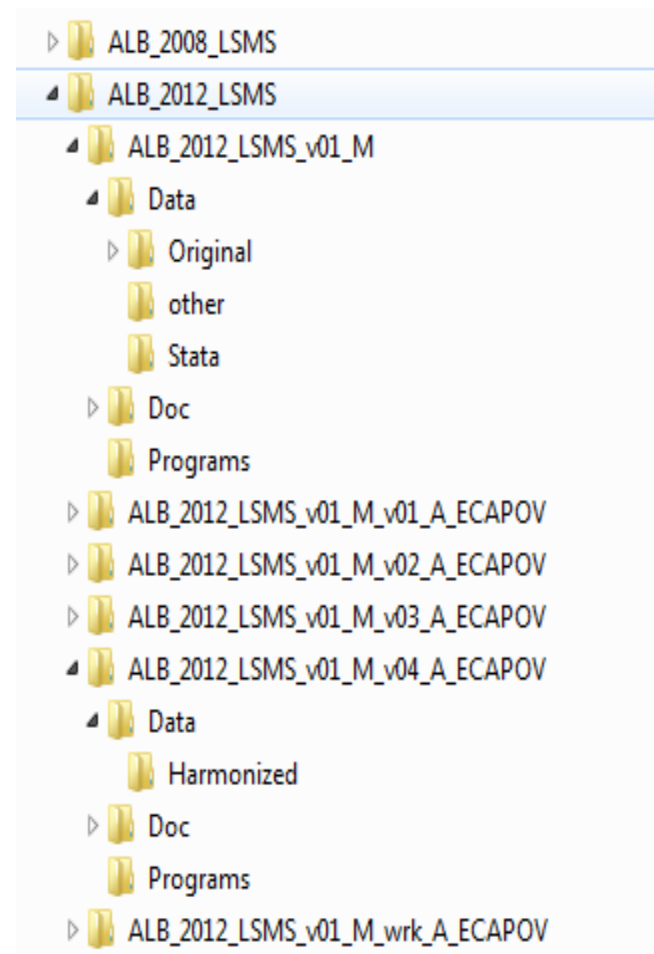
# Who is our audience?

ECAPOV data is meant **to complement** and **not substitute** national official poverty estimates.

WHO?	WHAT?	FORMATING?	DATA		TOOLS	ANALYTICS
			Microdata	Indicators		
TTL, STC, ETC	Tailored	Self or other Teams	X	X	X	
TTL	Tailored	ECATSD	X	X	X	X
TTL, STC, ETC, External Clients	Off-the-self	Self		X		X

## How do we structure the data?

- The data and programs (scripts) were organized in the hierarchical structures, with the following structures:
  - CCC – Countrycode
  - YYYY – Survey year
  - SSSS – acronym of the survey (i.e. HBS, LSMS, HIES)
  - vNN\_M – version of master/original data
  - vMM\_A – version of harmonized/adaptation data
  - i – ECAPOV modules (i=2,3...,10)



## Module 3: Consumption - Set of variables

- Household level data
- General Survey Information
- General HH Characteristics
- HH Consumption Information
- Consumption aggregates

# Welfare: income vs. consumption

To examine poverty and inequality, one needs a measure of **material well-being**. Ideally, this measure should correspond as closely as possible to the way a person experiences his or her standard of living.

**Income** is often considered to be the preferred measure because it is an indicator of the “potential” to enhance welfare. However, **measuring income suffers from several defects**, both in theory and in practice:

- Can be highly volatile
- Income underreporting is common in transition countries (respondents not willing to fully disclose illegal or semi-legal income sources)
- Inability to separate out what is “household” income and what is “business income” in households engaged in informal activities

Given the problems with income-based measurement, and the practices of countries in the Region to measure poverty, ECAPOV relies on measuring welfare with **consumption**.

# Comparable Consumption Aggregate

ECAPOV deals with issues of comparability with the following methodologies that are applied consistently across the region:

- Estimates of flow of services of durables, or durable purchases or rents are not used since data availability limits the application of this approach to all countries.
- Higher consumption should indicate a higher level of well-being. For health expenditures, this correspondence is questionable.
- Own food production and reciprocal exchange with other households and institutions are taken into account by using the **median local prices** to impute the value of **in-kind food consumption** and **households' own estimations** of the value of gifts and transfers in-kind for food and nonfood items.
- To limit the effect of extreme values at either end of the distribution, the data are “bottom-coded” at **1 per cent of per capita mean real consumption** and “top-coded” at **10 times the median of household consumption**. The final data set excludes all records with zero consumption.

# Price Deflation and Equivalence Scale

## Deflation

- Across Time - ECAPOV uses **Quarterly CPI (IMF)** indexes to compute real values across time.
  - This measure ignores the differential impact of price increases on the poor and non-poor. Unfortunately, No price indexes for low-income groups are routinely available in the Region.
- Across Areas/Countries - **The Paasche price index** is used to account for regional price differences
  - This index involves not only the prices faced by a household in relation to the reference prices but also its expenditure pattern
  - The spatial price deflator is based entirely on differences in food prices given data limitations.

## Equivalence

- **Per capita household consumption** is used to attribute the share of a household's resources to individuals within the household

# Set of variables

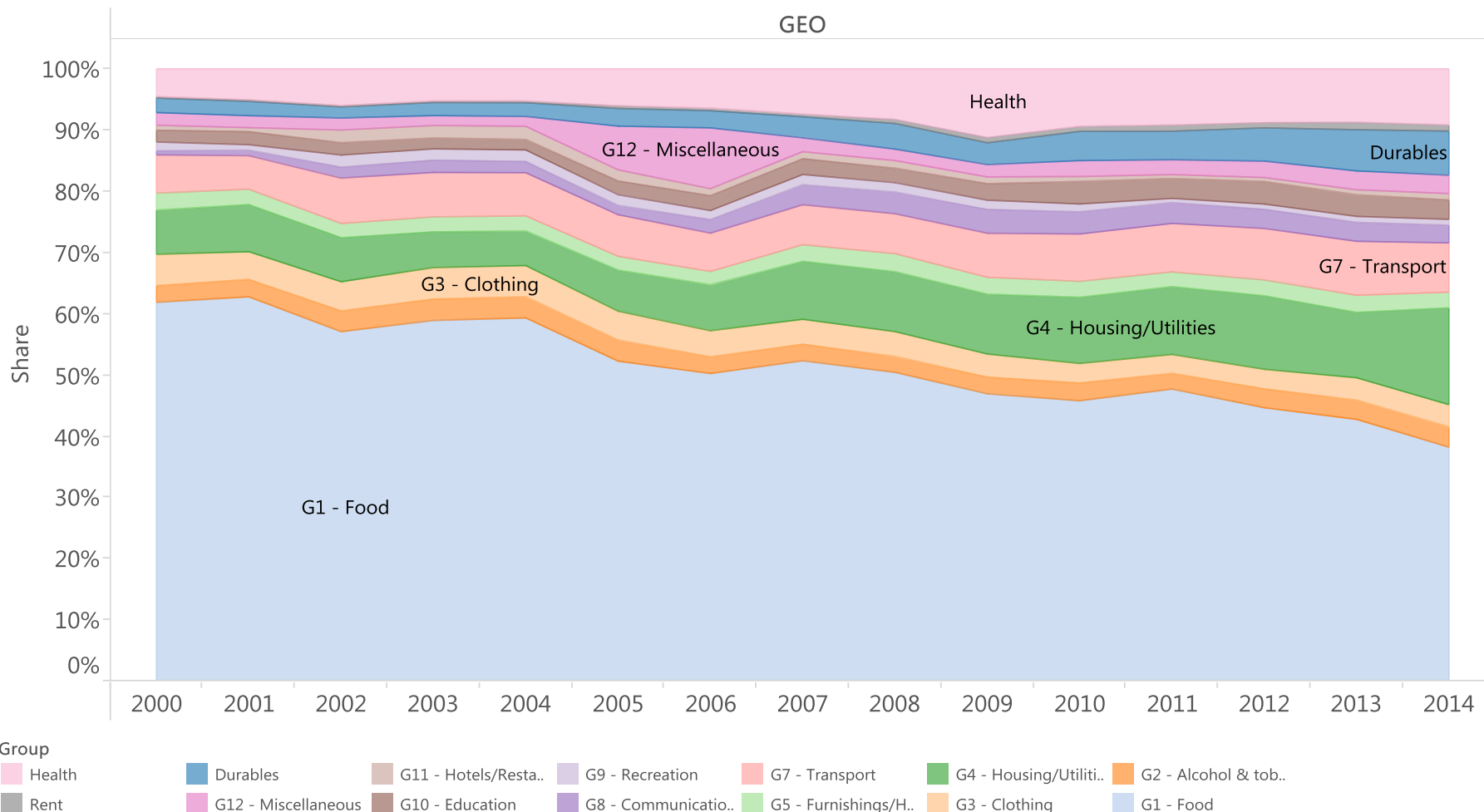
- General HH Characteristics
  - hhid (we preserve original hhid)
  - hhsize
  - weight
  - popw
  - region
  - urban
  - rdef
- Consumption aggregates
  - gall (excluding health, rent, durables)
  - **gallT** (including health, rent, durables) **← preferred**
  - gallT2 (shared prosperity)
- HH Consumption Information
  - g1pc: Food
  - g2pc: Alcohol and tobacco
  - g3pc: Clothing
  - g4pc: Housing/Utilities
  - g5pc: Furnishings/HH Equipment
  - health
  - g7pc: Transport
  - g8pc: Communications
  - g9pc: Recreation
  - g10pc: Education
  - g11pc: Hotels/Restaurants
  - g12pc: Miscellaneous
  - rent
  - durables

# Share of 1-digit COICOP groups overtime

Code

GEO

Consumption share (%) of 1-digit COICOP groups overtime





## Module 2: Individual characteristics

- Individual level data
- General Survey Information
- Core Individual Characteristics
- Other info (not necessarily standardized across survey)

# Module 2: Individual characteristics

- Core Individual Characteristics

hhid

hhsiz

pid

age

female 1-female, 0-male

relthead 1-head, 2-spouse, 3-children, 4-parents, 5-other relative, 6-other non-relative

marstat 1-Married, 2-Single, 3-Living Together, 4-Divorced/Separated, 5-Widow/er

edlev 1-incomplete 5-12, 2-general secondary, 3-special secondary, 4-tertiary, 5-none or incomplete primary

lfstat 1-employee, 2-self employed, 3-unemployed, 4-retired, 5-student, 6-OLF

- Other info (not necessarily standardized across survey)

occup Army, Managers, Professionals, Technicians, Clerks, Service and sales workers, Skilled agricultural, forestry and fishery workers, Craft and related trades workers

sector private, public, mixed

## Module 2: Individual characteristics (other variables)

Other variables also included

- Sector of work activity (4 or 10 categories)
- Type of contract (temporary, permanent, other)
- Form of employment (full-time, part-time, other)
- More detailed education level (complete vs. incomplete)

## Module 4: Utilities

- This module contains information on energy, water and urban services at the household level
- It corresponds to the COICOP (2-3 digit level items)
- It includes non-purchased items such as received in-kind or self-collected items
- It is recently added, thus cover not all countries and years


# Module 6: Social Protection and Social Assistance

- Individual level data
- General Survey Information
- Core Variables



M10 Old age pension  
M11 Disability benefits  
M12 Health and life insurance  
M13 Survivor pension  
M14 Maternity allowance  
M15 Other social insurance  
M16 Other pension  
M20 Social assistance  
M21 Child and family benefits  
M22 Scholarship  
M23 Housing and utilities  
M25 Disability - social assistance  
M26 Other benefits  
M27 Privileges  
M30 Unemployment insurance  
M40 Subsidies and veterans  
M50 Money from relatives/friends in country  
M51 Remittances from outside

# Module 7: Income

- Individual level data
- General Survey Information
- Core Variables 

totinc	1. annual total income excludes 4,5,11,12
oinc	2. annual other income
wage	3. annual wage income 4+5
hired_wage	4. annual employees wage
self_wage	5. annual self employed wage
SA	6. annual social assistance
remit	7. annual remittances
pension	8. annual pension
child_benefit	9. annual child and family benefits
unemp	10. annual unemployment benefit
public	11. annual wage from public sector
private	12. annual wage from private sector
property	13. annual income from propertyl,K,L
agr_inc	14. annual agricultural income

## Module 9: Assets and Services

- Individual level data
- Contains access on assets and services
  - Basic opportunities include access to education (primary, secondary, and post-secondary);
  - Basic infrastructure (central heating, gas, piped water, sewerage, toilet, electricity, refrigerator, phone, and personal computer).

# Module 10: National welfare aggregate module

- Household level data
- Contain national welfare aggregates, which the countries produced, available in
  - One variable includes the total welfare
  - When possible, include 12 COICOP group variables
  - Poverty lines
  - Weights



# ECAPOV: Challenges

## Coding of expenditure (COICOP)

*Standardization and norms are very important.*

## Durables

*HBS surveys only gather information about the durables acquired in the reference period. Neither information about the stock of durables nor information about how long households keep their durables is reported. So far, the harmonized welfare aggregate included the total amount paid for the durables acquired in the reference period.*

## Rent Imputation

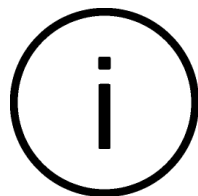
*Most surveys do not have a self-reported rent question (implicit rent, and when they do, they do not always ask this question in the same way). Not all HBS gather household characteristics that can be used in a hedonic model. Another challenge is that not all countries have a developed dwelling rental market, which makes the imputation more difficult (i.e. in KSV 2011, only 23 households pay a rent).*

## Health expenditure

*A decision whether or not to include health expenditure has to be made. The elasticity of the expenditure in health with respect to the total expenditure differs among countries.*

# ECAPOV: ACCESS

Harmonized data can be accessed online through the internal World Bank's Microdata Library.



## Documentation

Metadata documented  
using DDI standard



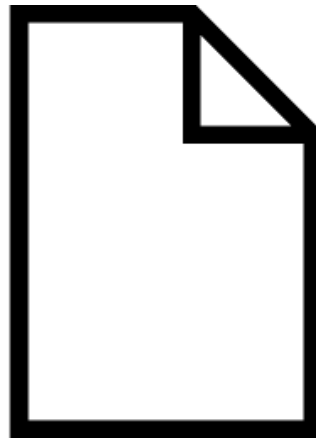
## Access Control

All users must submit  
request to access  
microdata (terms of use of  
original data producer are  
kept).

*The WB has been providing Technical Assistance and transferring this technology to our countries.*

# ECAPOV: TERMS OF USE

ECAPOV is available for poverty, inequality and shared prosperity monitoring in the region and follows the **NSO guidelines for terms of use**.



## ECAPOV: ANONYMITY

**All of our data is anonymous.** Terms of use explicitly forbid any attempt to de-anonymize data.



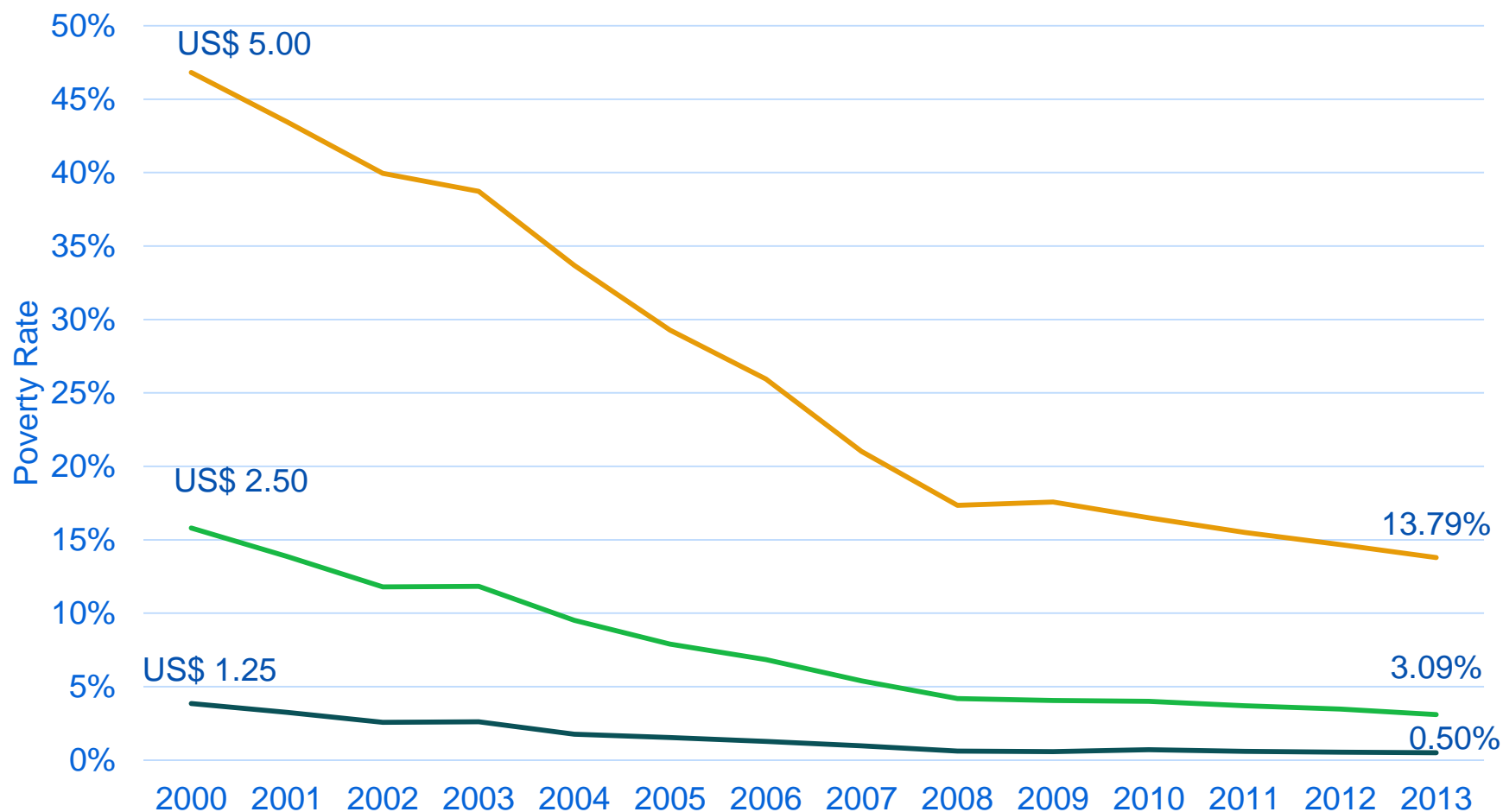
*The World Bank has been providing Technical Assistance on microdata anonymization.*

# **ECAPOV users can monitor trends and levels and make regional wide analysis**

World Development Indicators (PovCalNet)  
Shared Prosperity  
Poverty Reduction/Projections  
Economic groups  
Decomposition  
Profiling

# Poverty Reduction

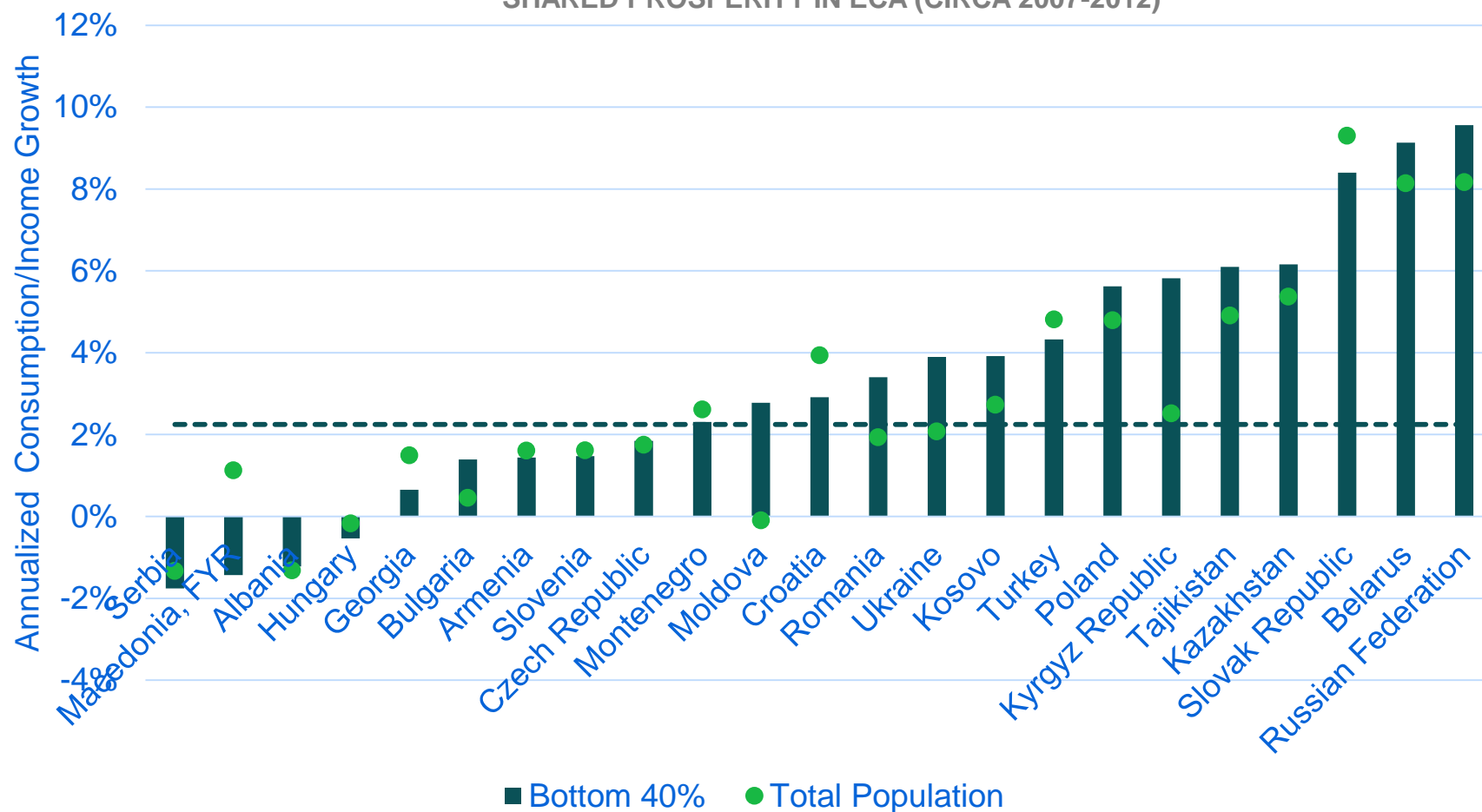
POVERTY RATE IN ECA 2000-2013



Source: ECATSD calculations using ECAPOV and EUSILC data  
Note: Regional trends are population weighted and include estimates

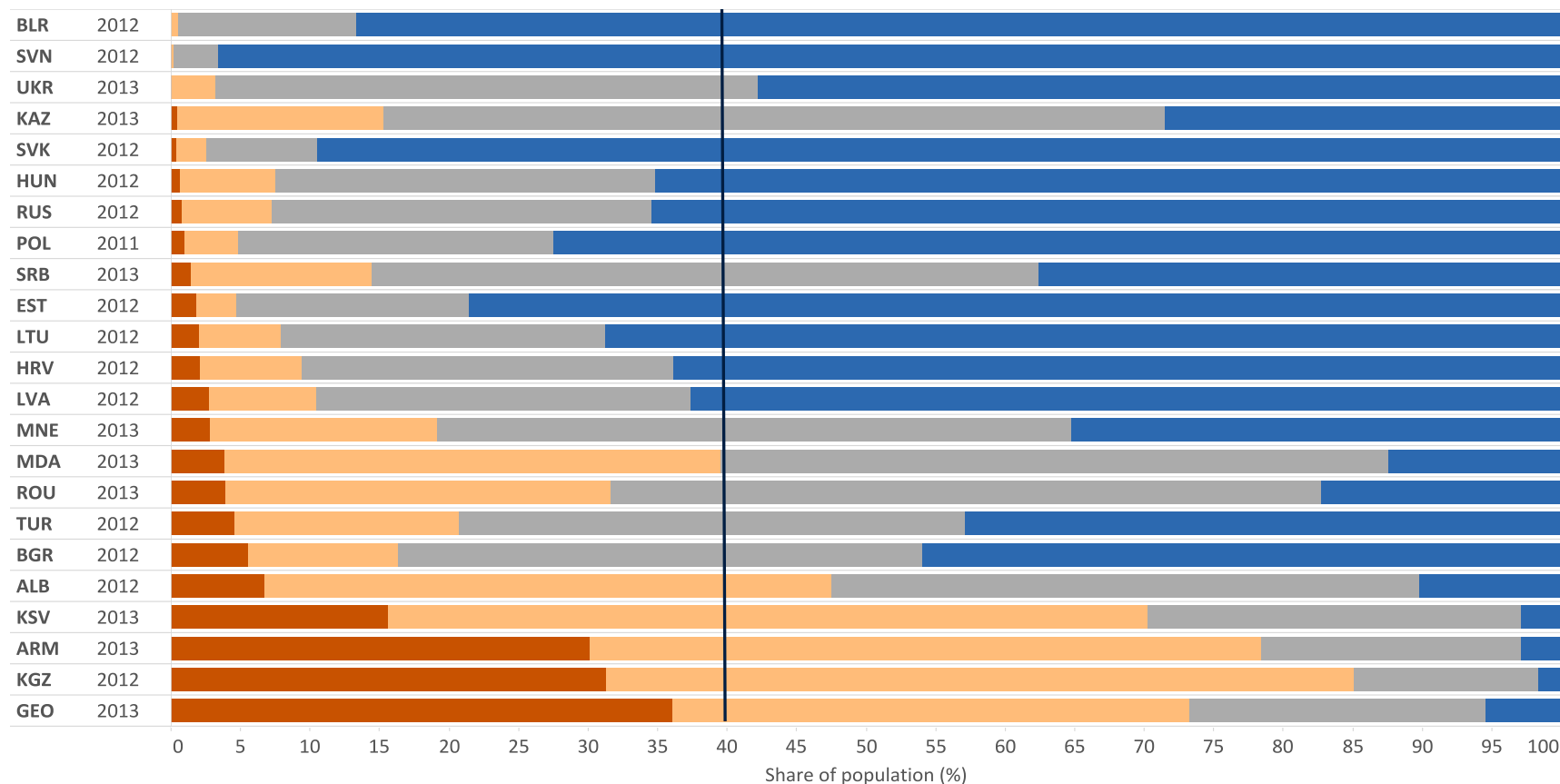
# Shared Prosperity

SHARED PROSPERITY IN ECA (CIRCA 2007-2012)



Source: ECATSD calculations using ECAPOV and EUSILC data

# ... and in many cases the vulnerable group accounts for a significant share of the B40



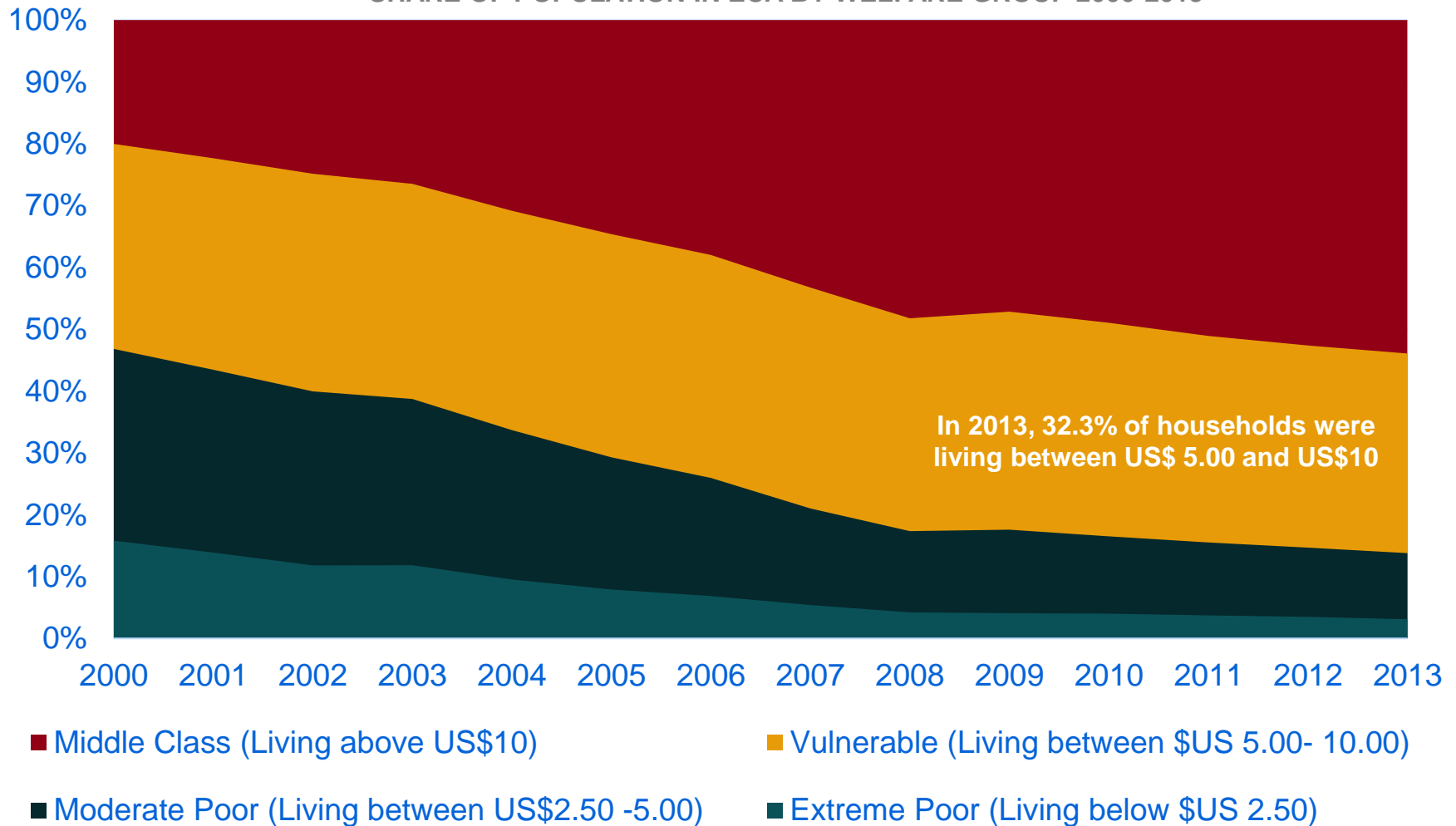
## Economic Group

- Middle class (above US\$10 PPP)
- Vulnerable (between US\$5-10 PPP)
- Poor (between US\$2.50-5 PPP)
- Extreme poor (below US\$2.50 PPP)



# Economic Groups

SHARE OF POPULATION IN ECA BY WELFARE GROUP 2000-2013



# REFERENCES

- Deaton, A., & Zaidi, S. (2002). *Guidelines for constructing consumption aggregates for welfare analysis*. Living Standard Measurement Survey Working Paper No. 135.
- Ravallion, M. (1998). "Poverty Lines in Theory and Practice." LSMS Working Paper No. 133, World Bank, Washington, DC.
- Ravallion, M. (1992). "Poverty Comparisons: A Guide to Concepts and Methods." LSMS Working Paper No. 88, World Bank, Washington, DC.
- Ravallion, M. (2008). "Poverty Lines." The New Palgrave Dictionary of Economics. Second Edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan. The New Palgrave Dictionary of Economics Online. Palgrave Macmillan. 17 September 2013
- Ravallion, M. (2012). "Poor or Just Feeling Poor" World Bank Policy Research Working Paper 5968.

# THANK YOU

JOÃO PEDRO AZEVEDO

[jazevedo@worldbank.org](mailto:jazevedo@worldbank.org)

MINH CONG NGUYEN

[mnguyen3@worldbank.org](mailto:mnguyen3@worldbank.org)