INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

The way to NUTS 2 regional estimates in the Portuguese EU-SILC survey

Eduarda Góis
(Statistics Portugal, Social Statistics Department)

SCORUS 2016 Lisbon Conference

Main points

- Motivation
- Regional features of the national SILC sample design
- The risk of poverty by NUTS 2 using SILC
- AROPE as a target
- New sample plan and the new sample size in 2015
- 2014-2016 income based estimates
- Concluding remarks

The regionalisation target is being implemented in the context of a grant agreement between Eurostat and Statistics Portugal.

Motivation

All people in the EU can realize their full potential whatever the region they live in

- The Europe 2020 Strategy establishes, besides other objectives, the importance of promoting social inclusion and combating poverty both at national and regional grounds
- Monitoring the implementation of regional funds with increased use of EU-SILC data over the period 2014-2020
- A criteria on poverty and social exclusion for the allocation of funds will be included as from 2020, to be decided in 2018

Motivation

- **Full development of the regional dimension in SILC as from 2016 data collection, at least at NUTS 2** or some aggregation of NUTS 2 level (possibly NUTS 1 if detailed enough)
- **Interest goes to the indicator on poverty or social exclusion, as well as its source indicators** (severe material deprivation and each of the MD items, low work intensity and at-risk-of-poverty rate)

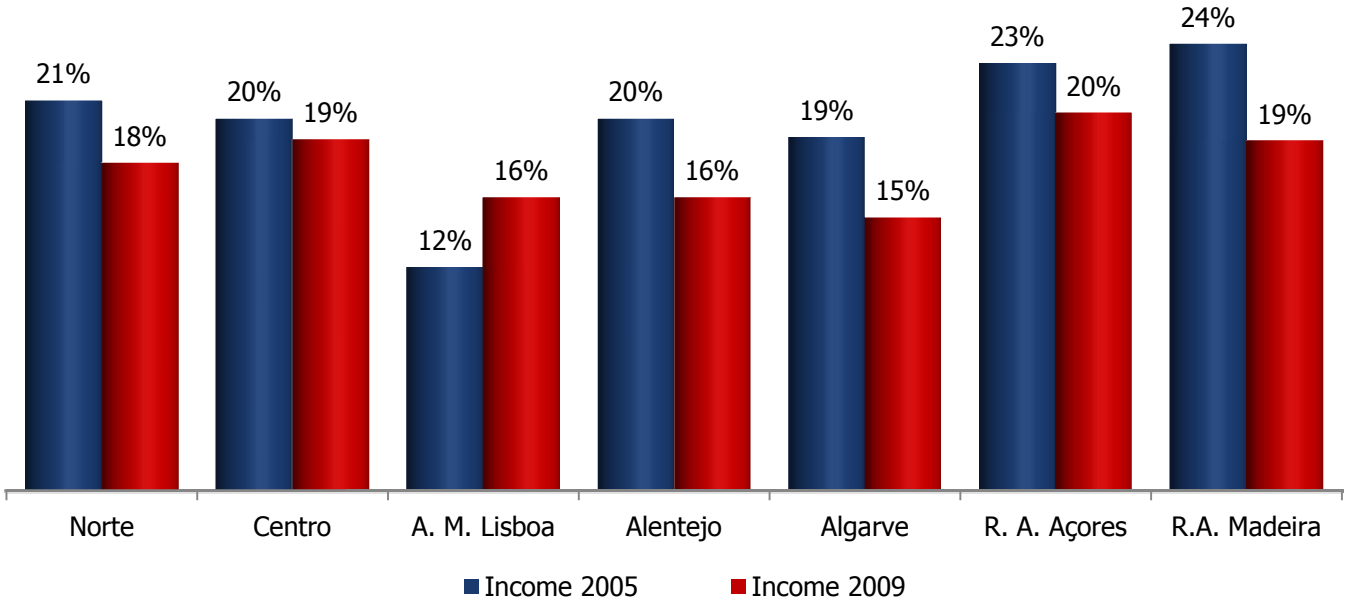
A problem: national sample designed to be country representative

Indicators that refer to smaller sub-populations are not likely to be currently reliable at the regional level due to the sample size

Motivation

- Portugal has been developing **HBS** approximately every five years since 1967, including both the expenditure and the income components at a detailed level and using **NUTS 2 representative samples**
- This allows for the calculation of reliable NUTS 2 income distribution estimates, however only each 5 years

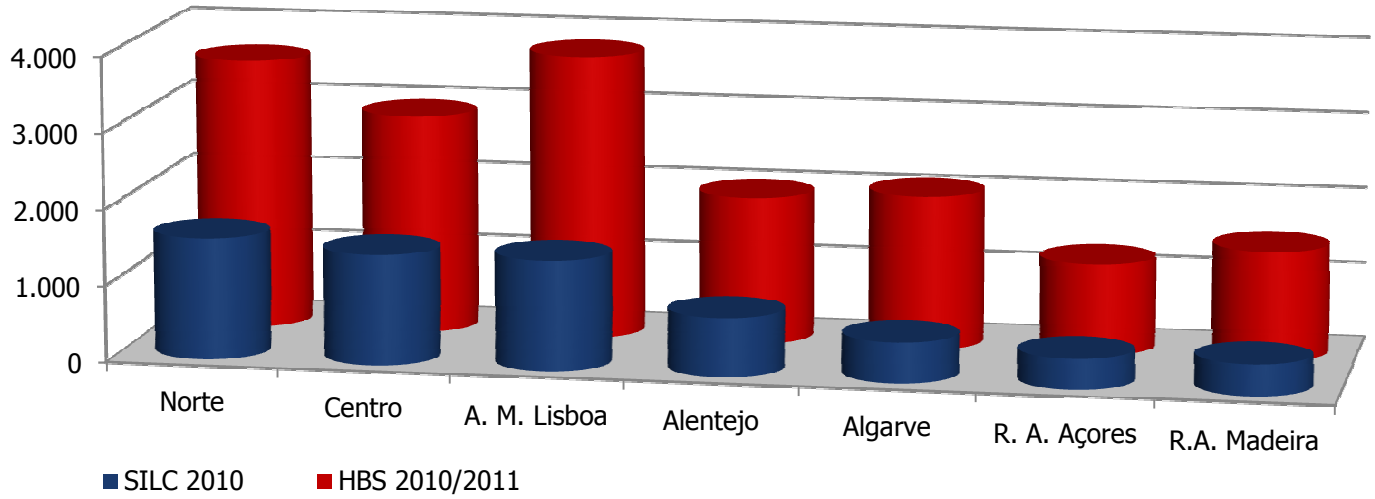
At-risk-of poverty rate (net monetary income), HBS



Motivation

- So, the implementation of **EU-SILC** in Portugal as from 2004 strictly followed the framework regulation in what regards a **nationally representative** probability sample
- For instance, the **2010 SILC sample size was less than 40% of the corresponding HBS sample size (less than 30% for smaller regions)**

Sample size HBS vs. SILC in 2010



The national SILC sample design

Two **sampling frames, selected using census information and constituted by private dwellings of usual residence and excluding collective households and institutions**

The Master Sample 2001 (MS), used in SILC between 2004 and 2015, was designed and selected using the information of the Census 2001

The new sampling frame, in use since 2013, was selected from the National Dwellings Register (NDR) which in turn uses information collected in the 2011 Census

Both **sampling frames are NUTS 3 stratified one-stage cluster samples**.

The national SILC sample design

Sample is made of four independent sub-samples (panels) where **each one follows a stratified two-stage cluster sampling design**

2004 to 2012: primary sampling units (PSU) were areas of the MS

2013 to 2015: a gradual transition between the two sampling frames is being implemented - the new rotations are selected using NDR while the "old" rotations are the ones previously selected using the MS.

2016 all rotations are NDR based rotations.

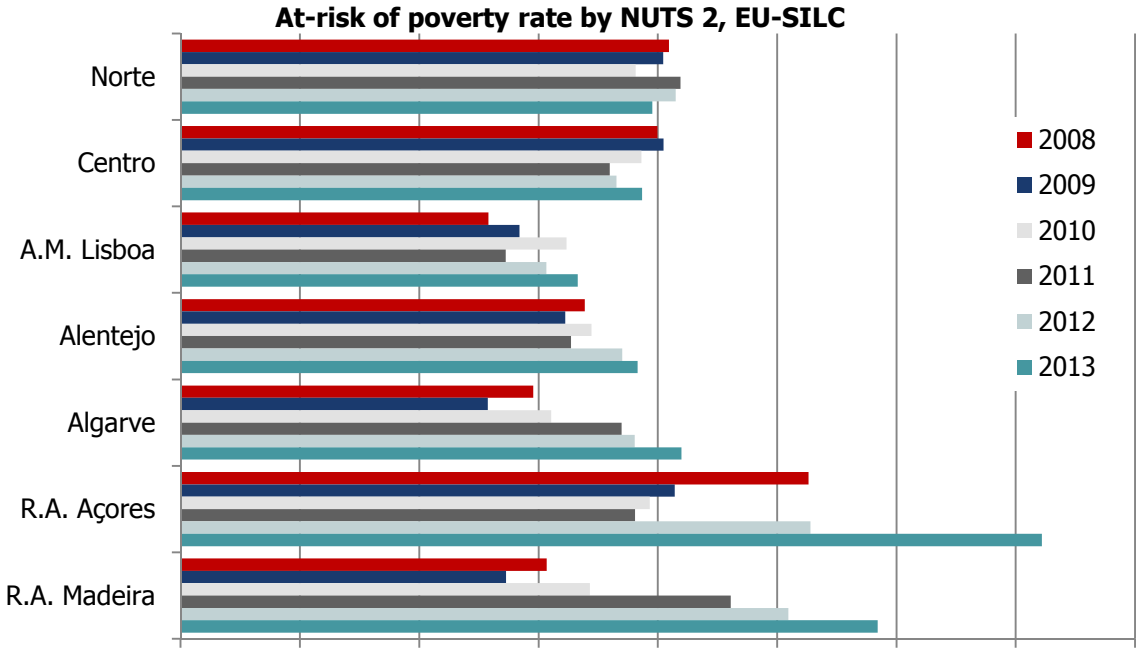
	Year			
Sampling frame	2013	2014	2015	2016
	MS	MS	MS	NDR
	MS	MS	NDR	NDR
	MS	NDR	NDR	NDR
	NDR	NDR	NDR	NDR

The national SILC sample design

In both cases the **PSU are stratified by a regional criterion (NUTS 2)**.

However when selecting the dwellings in each stratum, representativeness is not ensured at strata level (NUTS 2) but only at country level

When experimenting the calculation of indicators by NUTS 2 the estimates tend to be unreliable and erratic, especially for smaller regions



The calculation can be done but unreliable data for smaller regions!

The risk of poverty by NUTS 2 using SILC

Standard errors for the at-risk of poverty rates by NUTS 2, 2008-2013

NUTS II	2008	2009	2010	2011	2012	2013
Norte	1.7	1.7	1.4	1.4	1.5	1.4
Centro	1.6	1.9	1.7	1.4	1.3	1.3
Lisboa	2.1	1.8	1.7	1.5	1.7	1.6
Alentejo	1.9	1.9	2.1	1.8	1.9	1.9
Algarve	1.8	2.3	2.0	2.1	2.4	2.2
R.A. Açores	3.7	3.7	3.9	2.9	3.0	4.0
R.A. Madeira	3.6	2.4	2.6	3.3	3.9	3.5
Portugal	**0.8**	**0.8**	**0.7**	**0.6**	**0.6**	**0.6**

3 groups of regions according to **reliability**:

- **Less than 2 p.p. – 4 larger regions** (Norte, Centro, Área Metropolitana de Lisboa and Alentejo)
- **Between 2 and 2.5 pp. - Algarve**
- **Between 3 and 4 pp. - Regiões Autónomas** dos Açores e da Madeira

The risk of poverty by NUTS 2 using SILC

Standard errors for the at-risk of poverty rates by NUTS 2, 2008-2013

NUTS II	2008	2009	2010	2011	2012	2013
Norte	1.7	1.7	1.4	1.4	1.5	1.4
Centro	1.6	1.9	1.7	1.4	1.3	1.3
Lisboa	2.1	1.8	1.7	1.5	1.7	1.6
Alentejo	1.9	1.9	2.1	1.8	1.9	1.9
Algarve	1.8	2.3	2.0	2.1	2.4	2.2
R.A. Açores	3.7	3.7	3.9	2.9	3.0	4.0
R.A. Madeira	3.6	2.4	2.6	3.3	3.9	3.5
Portugal	**0.8**	**0.8**	**0.7**	**0.6**	**0.6**	**0.6**

The smaller regions show a tendency to higher changes in the risk of poverty

With large standard errors it is not clear whether this is due to a **real change** in regional income distribution vis-à-vis the national income median

or because of an **insufficient sample size**

AROPE as a target

"to have a 95% confidence interval for the AROPE indicator at 24% of ±2-2.5 percentage points

This would mean a value of the X in the formula and consequently a sample size of around 1100 (2,5pp) to 1800 (2pp) households in each region."

- Standard error for AROPE and NUTS 2 is not available but at country level we know the SE for AROP is 0.6 and for AROPE is 0.9
- Correlation between AROP and AROPE is ≈ 80%
- Expecting larger SE for AROPE than for AROP but not exceeding 5 p.p.
- **For Norte, Centro and AMLisboa sample sizes are above 1100 households and SE for AROP are less than 2 p.p.**
- **For the autonomous regions sample sizes are less than 500 households and SE for AROP are over 2 p.p.**

The new sample plan

- an **enlarged sample size** was developed taking into account a 95% confidence interval for the **AROPE** indicator at 24% of **±2-2.5 percentage points**
- to be implemented in the context of a 4-year rotational pattern, on the assumption that the best solution for the **adaptation to the new sample size is a partial increase by new rotation**
- a **design effect of 1.6** was considered, derived from the national at-risk of poverty rate
- the **average regional response rates between 2008 and 2012** in the national SILC survey were considered

The new sample plan

- the new sample size requires the **duplication of the size of the national representative sample** in the year of full implementation, 2018
- the complete regional **sample sizes will vary between 2 352 in the Região Autónoma dos Açores and 3360 in the AM Lisboa**
- the increase of sample size occurs in every region, however more strongly in **Alentejo (it doubles), Algarve (triples), Açores (almost 5 times the original size) and Madeira (more than 5 times)**

Sample size by NUTS 2 (standard error of 2.5 p.p.)

NUTS 2	National representativeness before the action	National representativeness in 2018	Increase of sample size
Norte	1,944	2,592	1.3
Centro	2,040	2,856	1.4
A.M. Lisboa	2,400	3,360	1.4
Alentejo	1,344	2,688	2.0
Algarve	864	2,592	3.0
R.A. Açores	504	2,352	4.7
R.A. Madeira	540	2,880	5.3
Portugal	**9,636**	**19,320**	**2.0**

The new sample plan

- **Increase** of the sample size is **gradual** , **2421 dwellings per new rotation between 2015 and 2018**
- In the first year of implementation, **2015**, the size of the new rotation was increased in **2421 dwellings**
- Process is cumulative in the 2nd, 3rd and 4th years of implementation: a difference of 4842 dwellings in 2016, 7263 in 2017, and 9684 dwellings in 2018

Sample size per year of implementation and NUTS 2

NUTS 2	Representativeness		Annual difference vs. before the action				Sample size per year of implementation			
	National	Regional	2015	2016	2017	2018	2015	2016	2017	2018
Norte	1,944	2,592	162	324	486	648	2,106	2,268	2,430	2,592
Centro	2,040	2,856	204	408	612	816	2,244	2,448	2,652	2,856
A.M. Lisboa	2,400	3,360	240	480	720	960	2,640	2,880	3,120	3,360
Alentejo	1,344	2,688	336	672	1,008	1,344	1,680	2,016	2,352	2,688
Algarve	864	2,592	432	864	1,296	1,728	1,296	1,728	2,160	2,592
R.A. Açores	504	2,352	462	924	1,386	1,848	966	1,428	1,890	2,352
R.A. Madeira	540	2,880	585	1,170	1,755	2,340	1,125	1,710	2,295	2,880
Portugal	**9,636**	**19,320**	**2,421**	**4,842**	**7,263**	**9,684**	**12,057**	**14,478**	**16,899**	**19,320**

The new sample size in 2015

- Overall the implemented sample size in 2015 (which includes the application of tracing rules for previous rotations) ensures the planned surplus for the year, in particular for the smaller regions
- In the Algarve the new sample size is 1.7 times the previous one, while it more than doubled for the Regiões Autónomas dos Açores and Madeira

NUTS 2	Sample			Expected difference in 2015
	2014	2015	Surplus	
Norte	1,686	1,817	131	162
Centro	1,657	1,886	229	204
A.M. Lisboa	1,725	2,075	350	240
Alentejo	1,078	1,440	362	336
Algarve	688	1,156	468	432
R.A. Açores	414	895	481	462
R.A. Madeira	458	1,054	596	585
Portugal	**7,706**	**10,323**	**2,617**	**2,421**

The new sample size in 2015

- The number of valid responses (8740) ensures at least 1100 households for the larger regions (Norte, Centro, AM Lisboa and Alentejo)
- For the smaller ones there is a considerable improvement, but a minimum of 1100 households is not achieved

NUTS 2	2015	
	Sample	Valid responses
Norte	1,817	1,616
Centro	1,886	1,683
A.M. Lisboa	2,075	1,560
Alentejo	1,440	1,254
Algarve	1,156	985
R.A. Açores	895	731
R.A. Madeira	1,054	911
Portugal	**10,323**	**8,740**

2014-2016 income based estimates

- Plan of action will not be fully implemented before 2018 (income 2017)
- However **regional estimates are needed for 2015, 2016 and 2017** (income 2014, 2015 and 2016)
- **Several options were considered namely "the use of aggregations, averages, small area estimates or model estimates"**

Use of aggregations

- The regional estimates will be used in particular for the allocation of funds in the context of the Social Cohesion policy
- People living in the smaller regions, namely in the autonomous regions, tend to face a risk of poverty above the average for the country
- On the other hand, different regional governments are associated to different decision and monitoring procedures
- Consequently, aggregating the smaller regions can lead to insufficient results

2014-2016 income based estimates

Small area estimation

- An area based SAE using HBS data was considered
- A disadvantage is that information available does not cover the material deprivation feature and the use of SAE would be restricted to AROP
- A second model would be needed to estimate the regional AROPE

Keep to direct estimation

- Use the collected data, taking advantage of the gradual increase of the sample and the gradual decrease of SE per region
- The use of HBS 2015 (2014 income) in the first year is a possibility

Concluding remarks

- The action plan was designed to ensure the availability of reliable regional AROPE estimates in 2018 (2017 income)
- By now only the first year of implementation was concluded
- New results based on half of the increase in sample size will be available in December 2016, when a new evaluation will be completed and new conclusions will be drawn about the increase of sample size
- Several options are still open for the calculation of regional intermediate estimates

Thank you for your attention!

Questions?

eduarda.gois@ine.pt