Jean Monnet Chair

Analysis of European Data by Small Area Methods

Lecture 1 bis: At risk of poverty rate, the survey data, the small area statistics framework and the small a area estimation problem

http://sampleu.ec.unipi.it

InGRID
Integrating expertise in inclusive growth

Poverty indicators

- There are many approaches to poverty indicators

- Here we refer to Laeken indicators:

At risk of poverty rate (Head Count Ratio)

1. At-risk-of-poverty rate

1.1 At-risk-of-poverty rate (after social transfers)

1.1.1 Definition

The share of persons with an equivalised total net income below 60% national median income.

Source : European Community Household Panel (ECHP)

1.1.2 Algorithm

1.1.2.1 Calculation of equivalised income

The total net income of each household is calculated by adding together the income received by all the members of the household from all sources.
For each person, the 'equivalised total net income (EQ_INC)' is calculated as its household total net income divided by equivalised household size according to the modified OECD scale (which gives a weight of 1.0 to the first adult, 0.5 to other persons aged 14 or over who are living in the household and 0.3 to each child aged less than 14).
Consequently, each person in the same household receives the same 'equivalised total net income'.
The population consists of all the persons living in private households of a country. The term person therefore includes all the members of the households, whether they are adults or children.
Persons with missing 'equivalised total net income' are excluded from the calculations (ie. people with missing household income or households with missing composition details).

1.1.2.2 Calculation of the 'at-risk-of-poverty threshold'

Firstly, persons have to be sorted according to their 'equivalised total net income' (sorting order: lowest to highest value).

Secondly, the median is calculated as the equivalised income of the household person for whom the cumulative sum of personal weights is less than or equal to 50% of the total sum of weights.

In other words, persons in the same household are located together, on the same side of the median.

Thirdly, the 'at-risk-of-poverty threshold' is calculated as 60% of the national median.

$$\text{At risk of poverty threshold} = 60\% * EQ\_INC_i \Big|_{i = \text{person for whom the cumulated sum of weights} = 0.5 * \text{total sum of weights}}$$

		2004	2009	2010	2011
Italy		26,2	24,9	25	28,1
	Nord-Ovest	16,7	15,7	16,2	17,8
Piemonte		18,5	16,8	18,2	21,8
Valle d'Aosta/Vallée d'Aoste		14,7	12,1	14,5	13,6
Liguria		19,4	15,8	16,4	19,4
Lombardia		15,4	15,3	15,2	15,9
Nord-Est		14,6	14,4	14,2	15,7
Provincia Autonoma di Bolzano/Bozen		15,2	11,9	10,6	11,1
Provincia Autonoma di Trento		11,4	10,1	10,8	14,4
Veneto		16,4	14,6	16,1	16,2
Friuli-Venezia Giulia		14,9	16,7	14,6	16,3
Emilia-Romagna		12,9	14,3	12,7	15,7
	Centro (IT)	20,1	18,5	19,9	22,3
Toscana		14,5	14,9	17,5	20
Umbria		18,4	17,1	17,7	20,7
Marche		18,4	16,3	18,4	21,9
Lazio		24,7	21,9	22,4	24,1
	Sud	41,7	39,7	39,7	44,7
Abruzzo		21,4	27,2	27,4	34,3
Molise		27,7	33,7	32	33,8
Campania		45,5	44	45,1	48,3
Puglia		42,4	36,4	36,3	42
Basilicata		35,2	41,7	35,2	47,6
Calabria		46,7	42,9	41,5	47,1
	Isole	44,4	43,8	42,2	48,9
Sicilia		49,9	48,2	47,4	54,3
Sardegna		27,6	30,5	26,2	32,4

EUSilc Sample Survey - 1

The EU-Statistics on Income and Living Conditions (EU-SILC) instrument is the EU reference source for comparative statistics on income distribution and social inclusion at the European level.

EUSilc Sample Survey - 2

It provides two types of annual data for 27 European Union countries, Croatia, Iceland, Norway, Switzerland and Turkey:

- Cross-sectional data pertaining to a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions, and
- Longitudinal data pertaining to individual-level changes over time, observed periodically over a four year period.

EUSilc Sample Survey - 3

- EU-SILC does not rely on a common questionnaire or a survey but on the idea of a "framework".
- The latter defines the harmonised lists of target primary (annual) and secondary (every four years or less frequently) variables to be transmitted to Eurostat;
- common guidelines and procedures; common concepts (household and income) and classifications aimed at maximising comparability of the information produced.

EUSilc Sample Survey - 4

- EU-SILC provides “statistically sound” estimates at Country level, not always at NUT2 level (2020 goal).
- EU-SILC does not provide statistically sound estimates at level of the Local Administrative Units 1 and 2;
- What does “statistically sound” mean?

“Statistically sound estimate” 1

In descriptive statistics: the coefficient of variation (CV) is the ratio of the standard deviation to the value of the mean

Coefficient of Variation = (Standard Deviation/ mean) * 100.

For example, the expression “The standard deviation is 15% of the mean” is a coefficient of variation.

“Statistically sound estimate” 2

In descriptive statistics:

the CV is particularly useful when you want to compare variability of two different groups or populations.

For example: Income in Pop A has CV=15%, Income in Pop B has CV=30%...the distribution of income in Pop B has more dispersion (is more variable)

“Statistically sound estimate” 3

In Statistical Inference: the coefficient of variation (CV) is the ratio of the standard error of an estimate to the value of the estimate

Coefficient of Variation = (Standard Error / Estimate) * 100.

For example, the expression “The standard error is 15% of the estimate” is a coefficient of variation.

“Statistically sound estimate” 3

In Statistical Inference:

For example: estimator A has CV=15%, estimator B has CV=30%...the sampling distribution of estimator B has more dispersion (is more variable) and the estimator B is less efficient than A

“Statistically sound estimate” 4

In sample survey (Inference)

The CV is particularly useful when you want to assess the accuracy (efficiency + unbiasdness) of the results of a survey (estimate):

The MSE (Mean Squared Error) is equal to Variance + Bias^2

MSE(estimator) = Variance(estimator)+bias(estimator)^2

Coefficient of Variation = square root(MSE(estimate)) /(Estimate) * 100.

For example, the expression “The sqrt(MSE) is 15% of the estimate” is a coefficient of variation and it is a measure of the accuracy of the estimate

“Statistically sound estimate” 5

It means accurate, with a low CV.

When I say low it means that its value should not exceed the 20-30% of the value of the estimate itself.

Many Official Statistical Agencies do not publish estimates with CV higher than 20%

Estimation: scope and purpose

Estimation (or **estimating**) is the process of finding an **estimate**,

Estimation is often done by survey sampling, projecting the values of the estimator on the sample on to larger population.

In statistics, an estimator is the formal name for the rule by which an estimate is calculated from data, and estimation theory deals with finding estimates with “good” properties.

What is estimation for domains and small areas?

- *Estimation for domains*, or *domain estimation* for short, refers to the estimation of population quantities, such as:
  - **Totals**
  - **Means**
  - **Proportions**
  - **Medians, Quantiles, Percentiles…**

  for the desired population subgroups called **domains** (small or large)

Special case - SAE

- *Small area estimation,* SAE

  - **Estimation for domains whose sample size is small or very small (even zero)**

  - **Alternative definition (Partha Lahiri):**

    Small area = Domain of interest, for which the sample size is not adequate to produce reliable direct estimates

More general framework: Small area statistics

- Production of statistics for domains and small areas based on sample survey data and/or population data from registers
- **SAE approach**
  - Production of statistics using sample survey data and auxiliary data & models to improve accuracy of estimates
- **Alternative approach**
  - Production of statistics based solely on population data from statistical registers

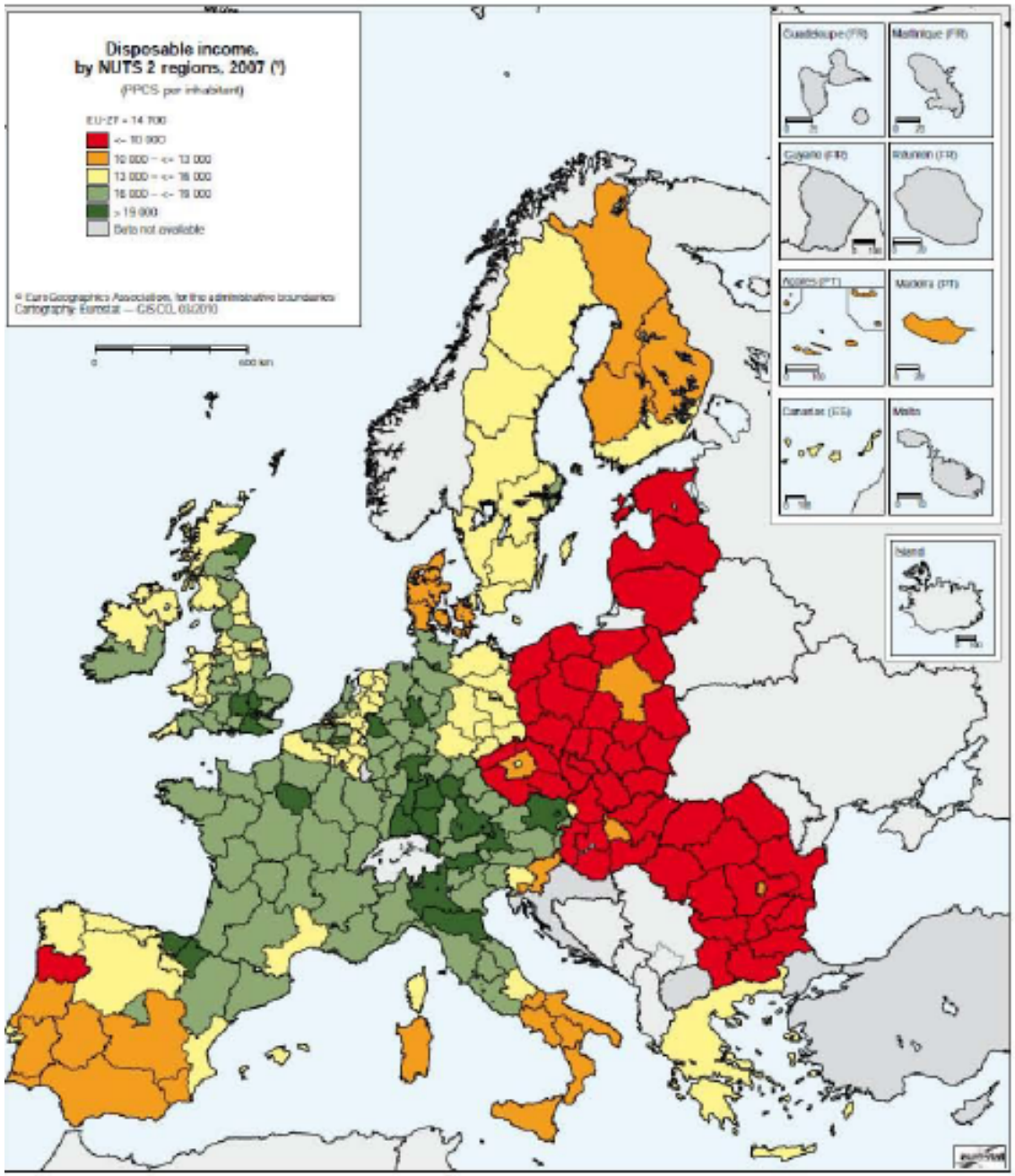
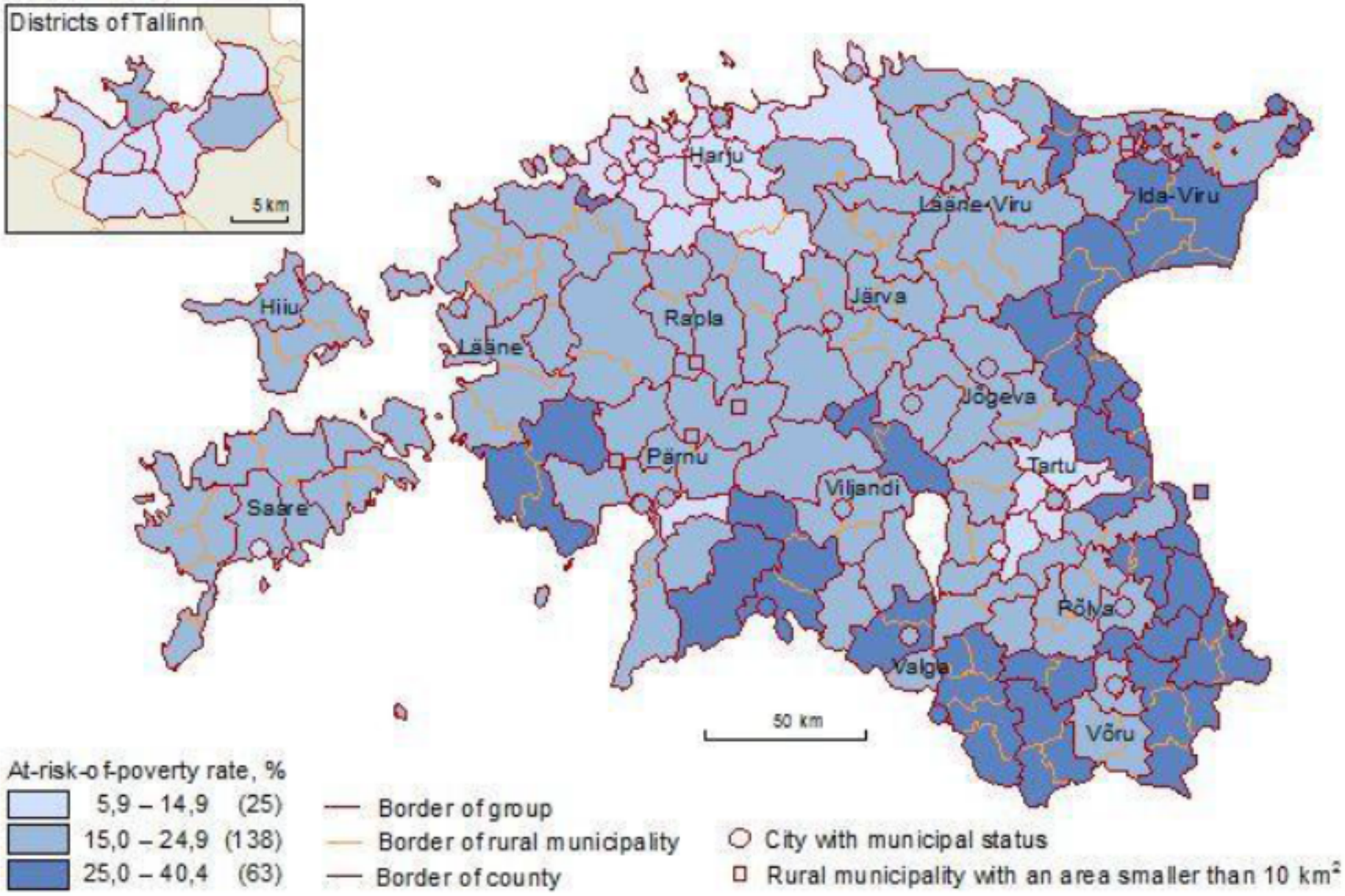


Figure 1. Disposable income by NUTS 2 regions in 2007 in the European Union

Source: Eurostat Regional Yearbook 2010, p.93, Section on Household Accounts. Information about the metadata is available at http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/reg_ecohh_esms.htm

Poverty map: Estonia

World Bank 2014 – Regional poverty rates based on SILC data



Typical estimation task - 1

- Specify and identify the domains of interest
  - **Breakdown of population into sub-populations**
  - **The number of domains can be large**
- Specify study variable *y*
- Specify target parameters for the domains
  - **Totals**
  - **Means**
  - **Ratios**
  - **Percentiles, Medians, …**

Typical estimation task - 2

- Specify data sources
  - **Sample survey data**
  - **Auxiliary data: Census, Admin. Registers, Statistical registers**
- Specify estimators of domain parameters
- Specify variance and MSE estimators
- Computation, graphical illustration
- Quality assurance
- Publication

Examples

- Estimation of regional number of ILO unemployed by sex and age group, based on Labour Force Survey LFS
- Estimation of median household disposable income by municipality, based on sample survey data such as EU SILC
- Estimation of regional poverty indicators, such as regional poverty rate, based on sample survey data such as EU SILC