SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU

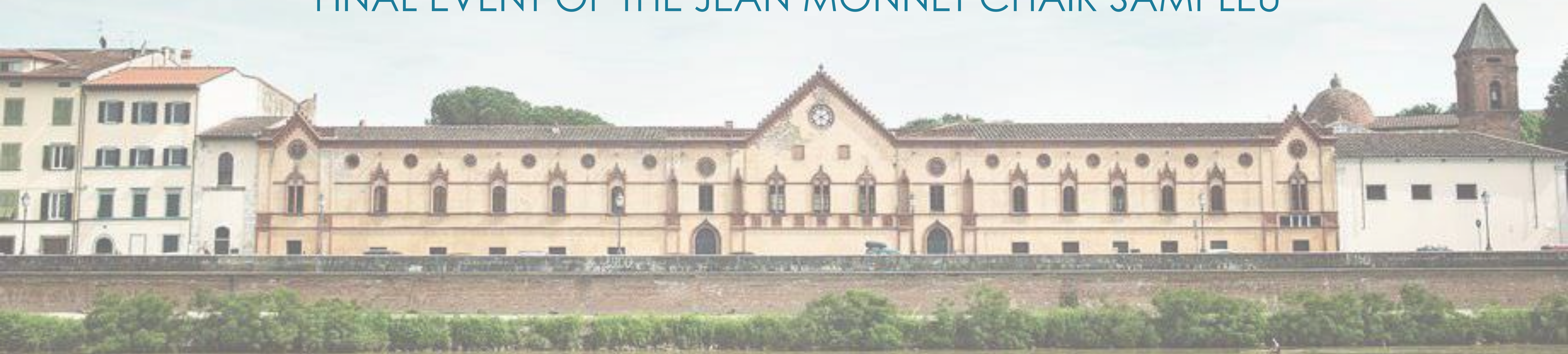
IN SUPREMÆ DIGNITATIS 1343 UNIVERSITÀ DI PISA

Jean Monnet Chair
European Commission

PISA, 8-10 MAY 2018

WORKSHOP "SMALL AREA METHODS AND LIVING CONDITIONS INDICATORS IN EUROPEAN POVERTY STUDIES IN THE ERA OF DATA DELUGE AND BIG DATA"

FINAL EVENT OF THE JEAN MONNET CHAIR SAMPLEU



Measuring accuracy of poverty measures at small area level

Table 4. Statistical quality of small area estimates – internal and external validation of the estimates?

Francesca Gagliardi

Pisa, May the 9th 2018

IN SUPREMÆ DIGNITATIS 1343

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU
Jean Monnet Chair

Importance of information on sampling errors - 1

- Data are subject to errors from diverse sources, information on sampling errors is of crucial importance in proper interpretation of the survey results

- Sampling error is only one component of the total error in survey estimates

↓

It is the lower (and more easily estimated) bound of the total error

↓

A survey will be useless if this component alone becomes too large for the survey results

Importance of information on sampling errors - 2

- The relative magnitude of sampling error increases as we move from estimates for the total population to estimates for individual subgroups (like small area) and comparison between subgroups.

- Information on the magnitude of sampling errors is essential in deciding the degree of detail with which the survey data may be meaningfully tabulated and analysed.

- Various practical methods and computer software have been developed for computing sampling errors

No justification for the continued failure to include information on sampling errors in the presentation of survey results.

Procedures for computing sampling errors:

(1) must take into account the actual, complex structure of the design

(2) should be flexible enough to be applicable to diverse designs

(3) should be suitable and convenient for large-scale application, and for producing results for diverse statistics and subclasses

(4) should be robust against departure of the design in practice from the ideal `model' assumed in the computation method

(5) should have desirable statistical properties such as small mean-squared error of the variance estimator

(6) should be economical in terms of the effort and cost involved

(7) suitable computer software should be available for application of the method

Jackknife Repeated Replication (JRR)

- It is a versatile and straightforward technique for variance estimation
- It is based on comparisons among replications generated through repeated re-sampling of the same parent sample
- Once the set of replications has been appropriately defined for any complex design, the same variance estimation algorithm can be applied to a statistic of any complexity.

The idea of replication techniques

- The basic requirement is that the full sample is composed of a number of subsamples or replications, each with the same design and reflecting complexity of the full sample, enumerated using the same procedures.

- A replication differs from the full sample only in size. But its own size should be large enough for it to reflect the structure of the full sample, and for any estimate based on a single replication to be close to the corresponding estimate based on the full sample.

- At the same time, the number of replications available should be large enough so that comparison among replications gives a stable estimate of the sampling variability in practice.

Formula of JRR

- z be a full-sample estimate of any complexity.
- i sample primary sampling unit (PSU)
- h stratum;
- $a \geq 2$ is the number of PSUs in stratum h.
- $z_{(hi)}$ estimate produced using the same procedure after eliminating primary unit i in stratum h and increasing the weight of the remaining $(a_h\text{-}1)$ units in the stratum by an appropriate factor g_h .
- $z_{(h)}$ be the simple average of the $z_{(hi)}$ over the a_h sample units in h.

$$\text{var}(z) = \Sigma_h \left[(1 - f_h) \cdot \Sigma_i g_{(hi)} \left(z_{(hi)} - z_{(h)} \right)^2 \right]$$

$$g_{(hi)} = w_h / (w_h - w_{hi}) \qquad w_h = \Sigma_i w_{hi}$$

$$w_{hi} = \Sigma_j w_{hij}$$ ultimate units j in primary selection units i

Practical aspects

In order to apply the JRR technique (and any other resampling technique) it is important to clarify two practical aspects:

- Explicit and implicit **stratification** and computational strata
- Computational **PSU** (Primary Selection Units)

In many practical situations some aspects of **sample structure need to be redefined** to make variance computation possible, efficient and stable.

Of course, any such redefinition is appropriate only if it does not introduce significant bias in variance estimation. The computational structure can differ from the actual sample structure because of various consideration.

Problems in computing variances

- <u>Sample structure availability</u> (*Strata, PSU, detailed description of sampling*)

  ⇨ *Alternative procedure*

- <u>Sample size</u>

Too small at regional level ⇨ Estimates with large variability

⇨ *Small Area procedures*

Cumulation □□ Internal validation

SAE □□ External validation

⇨ *Use of fuzzy measures of poverty*

Thank you for your attention!

IN SUPREMÆ DIGNITATIS 1343

Centro Camilo Dagum
Tuscan Interuniversity Centre
Advanced Statistics for Equitable and Sustainable Development

MAKSWELL
MAKing Sustainable development and WELL-being frameworks work for policy analysis

SMALL AREA METHODS FOR MONITORING OF POVERTY AND LIVING CONDITIONS IN EU
Jean Monnet Chair