

# Weighting and estimation methods: description in the Memobust handbook

Loredana di Consiglio, **Fabrizio Solari**

2013 European Establishment Statistics Workshop

Norimberg, 10 September 2013

# MEMOBUST

- Project MEMOBUST (MEthodology for MOdern BUusiness STatistics)
- Main goals:
  - identification of best practices and the development of common methodologies and guidelines supporting the production of business statistics aiming at fostering efficiency and integration of processes.
  - replacement of a methodological handbook for the production of business statistics. This is intended to be an update of the existing “Handbook on design and implementation of business statistics” (edited by Willeboordse in 1998).

# Handbook on Business Statistics

- Modules are written for
  - statisticians/managers responsible for running and producing business statistics
  - methodologists
- Two types of modules are used:
  - theme modules (of general nature; broad readership)
  - method module (more technical; especially for methodologists)

# Weighting and Estimation (theme module)

Method modules contained in the theme module “Weighting and estimation” can be roughly grouped in the following topics:

- Calibration
- Robust estimation in presence of outliers
- Preliminary estimates
- Small area estimation
- Estimation with administrative data

# Weighting and Estimation (theme module)

## XIX. Weighting and Estimation

- 0. Main theme module
- 1. Design of estimation
- 2.c. Calibration
- 2.d. GREG
- 4.a. Outlier treatment (robust estimation)
- 4.b. Winsorisation
- 4.c. Weight trimming
- 5. Model-based estimation
- 7.a. Estimation for short term statistics
- 7.b. Preliminary estimation with design-based methods
- 7.c. Preliminary estimation with model-based methods
- 8.a. Small area estimation
- 8.b. Synthetic estimator
- 8.c. Sample size dependant and composite estimator
- 8.e. EBLUP area level for small area estimation
- 8.f. EBLUP unit level for small area estimation
- 8.g. Small area estimation methods for time series data
- 9. Estimation with administrative data

# Calibration estimation

Design weights equal the inverse of the inclusion probability and can be thought as the number of units of the population each unit in the sample is representative of.

Hence, a simple method to obtain estimates is to use these design weights to inflate the sample observations. Design weights can also be adjusted to consider non-response.

Design weights can be modified to take into account auxiliary information (Sändal *et al.*, 1992) by applying calibration estimation.

## Weighting adjustment for non-response

The principle of weighting is also applied to account for unit non-response. Design weights can be adjusted also to consider non-response in order to reduce the possible bias of resulting estimates.

For example, the sample can be partitioned into sub-groups of units where the response rates are assumed to be constant, and where it can be assumed that non-respondents behave similarly to respondents.

Equivalently, the method is based on the assumption that the non-response depends on auxiliary variables defining a partition of the population, but conditionally on these variables it is independent of the target variable (i.e. non-response is MAR, see Little and Rubin, 2002).

# Robust estimation in presence of outliers

The effect of outliers on estimation can be significant, since in such situations estimators don't retain their properties in terms of bias or efficiency. Outliers treatment at estimation stage (robust estimation) aims at reducing the effect on variance of outliers, also controlling the possible bias of the estimator.

The module presents some nonstandard robust estimation techniques:

- modification of the Greg estimator as proposed by Chambers, Falvey, Hedlin and Kokic (2001),
- Winsor estimator (see Mackin and Preston, 2002),
- local regression (Kim, Breidt and Opsomer, 2001)



# Robust estimation in presence of outliers

## Modification of the Greg estimator

The modification proposed by Chambers, Falvey, Hedlin and Kopic (2001) refers to GREG estimators assuming heteroscedasticity.

They propose to reduce the proportion of disjunctive observations by their substitution or by post-stratification.



# Robust estimation in presence of outliers

## Winsor estimation

Units drawn into the sample, for which the variable takes a value beyond specified border points, are changed (Kokic and Bell, 1994, and Chambers, 1996).

The biggest problem in Winsor estimation is the designation of adequate border points, which are crucial in indicating outliers.

# Preliminary estimation

Preliminary (or provisional) estimates are the estimates that are computed using the statistical information available on the basis of the sample that is observed at time of first release of the estimates.

The main problem that has to be faced off in preliminary estimation context concerns the possible self-selection of early respondents, since self-selection can lead to biased estimates.

# Preliminary estimation

Preliminary estimation methods may be classified in function of the stage on which the preliminary method is applied:

1. at the sampling design stage, by selecting a preliminary subsample of the final sample;
2. at the estimation stage, in the following ways:
  - a) by means of imputation techniques of missing data, that are applied to non respondent units;
  - b) by means of weighting adjustment;
  - c) by applying direct and indirect estimators and/or time series of preliminary and final estimates

## Small area estimation

The aim of small area estimation methods is to produce reliable estimates for unplanned domains for which direct estimates cannot be considered reliable (in some cases direct estimator cannot be even computed when no sampling units are observed for some specific domain).

The main idea underlying small area techniques is to increase their effective sample size, (see Rao, 2003).

An improvement in the efficiency of the estimates can be achieved by assuming, implicitly or explicitly, a relationship linking together sampling units in the small area of interest and sampling units in the small areas which behave similarly to the small area of interest.

13 Furthermore, an increase in efficiency can be obtained using information coming from exploiting the sample for the same areas at different times.

---

# Small area estimation

SAE methods described in the modules can be divided in:

- design based methods:
  - Sample-size dependent estimator (Drew et al., 1982), the James-Stein estimator (see Rao, 2003).
- model based methods:
  - Model based methods based on Linear Mixed Models (see Rao, 2003).

A specific module is dedicated to small area estimation methods for time series data.

## Estimation with administrative data

Two situations can be distinguished when producing estimates in an administrative data based system with the largest enterprises being surveyed.

Situation A: the available administrative data provide good coverage when the estimates have to be made.

Situation B: no or only few administrative data are available for the STS estimates.

The module provides the technical descriptions for producing estimates for situations A and B respectively.

Draft results from MEMOBUST project are available at

[www.cro-portal.eu](http://www.cro-portal.eu)

(registration needed)



Thanks for your attention!!!