Session 3
Multisource Analyses for Official Statistics

Discussion

M. Giovanna Ranalli[1]

[1]Dip. Scienze Politiche, Università degli Studi di Perugia

Fourth Workshop on Methodologies for Official Statistics
ISTAT — Rome, 1-2 December 2025

Outline

1. Overview of the two papers
2. Innovations in the Quality of EU-SILC Survey Estimates
   - D. Chianella, M. D'Alò, S. Gerosa, M. Terribili, C. De Vitiis — Istat, Italy
3. A Constrained Optimization Approach for adjusting Misclassified Proxy Variables
   - F. Solari, V. Quondamstefano, R. Filippini, Gaia Rocchetti — Istat, Italy
4. A final note...

The two papers

Similarities

- Administrative data
- Calibration

Differences

- Increase efficiency vs. editing (reducing bias)
- Use of modeling

Outline

1. Overview of the two papers
2. Innovations in the Quality of EU-SILC Survey Estimates
   - D. Chianella, M. D'Alò, S. Gerosa, M. Terribili, C. De Vitiis — Istat, Italy
3. A Constrained Optimization Approach for adjusting Misclassified Proxy Variables
   - F. Solari, V. Quondamstefano, R. Filippini, Gaia Rocchetti — Istat, Italy
4. A final note...

Overview of the paper

- Regulation 2019/1700 $\rightarrow$ new and demanding precision requirements for key EU-SILC indicators (e.g. AROPE for NUTS2, APROP for national level over 4 years).
- Large nonresponse rates: $\downarrow$ sample size and $\uparrow$ sampling weight variability $\rightarrow$ effects on cross-sectional and longitudinal estimates.
- Availability of new data sources and their integration into the system of statistical registers

$\Rightarrow$ A revision in the procedures for (1) non-response adjustment and (2) calibration used in the EU-SILC survey

- Advisory Board

Revision for weight adjustment

Non-response adjustment

- From individual propensities (Logistic 0) to groups based on quantiles
- Use of CART
- ? Model selection
  - auxiliary variables
  - number of groups for Logistic and `maxdepth` parameter for CART
  - bias/variance trad-off

Calibration adjustment

- From two steps to one step: essential constraints required by European regulations + efficiency-related constraints
- ? Constraints selection – "analyzing the reduction in coefficients of variation (CVs) and the stability of the estimates"

BDR? Can BDR be used for model/constraint selection?

Revision for weight adjustment

Non-response adjustment

- From individual propensities (Logistic 0) to groups based on quantiles
- Use of CART
- ? Model selection
  - auxiliary variables
  - number of groups for Logistic and `maxdepth` parameter for CART
  - bias/variance trad-off

Calibration adjustment

- From two steps to one step: essential constraints required by European regulations + efficiency-related constraints
- ? Constraints selection – "analyzing the reduction in coefficients of variation (CVs) and the stability of the estimates"

BDR? Can BDR be used for model/constraint selection?

Revision for weight adjustment

Non-response adjustment

- From individual propensities (Logistic 0) to groups based on quantiles
- Use of CART
- ? Model selection
  - auxiliary variables
  - number of groups for Logistic and `maxdepth` parameter for CART
  - bias/variance trad-off

Calibration adjustment

- From two steps to one step: essential constraints required by European regulations + efficiency-related constraints
- ? Constraints selection – "analyzing the reduction in coefficients of variation (CVs) and the stability of the estimates"

BDR? Can BDR be used for model/constraint selection?

Revision for weight adjustment

Non-response adjustment

- From individual propensities (Logistic 0) to groups based on quantiles
- Use of CART
- ? Model selection
  - auxiliary variables
  - number of groups for Logistic and `maxdepth` parameter for CART
  - bias/variance trad-off

Calibration adjustment

- From two steps to one step: essential constraints required by European regulations + efficiency-related constraints
- ? Constraints selection – "analyzing the reduction in coefficients of variation (CVs) and the stability of the estimates"

BDR? Can BDR be used for model/constraint selection?

Revision for weight adjustment

Non-response adjustment

- From individual propensities (Logistic 0) to groups based on quantiles
- Use of CART
- ? Model selection
  - auxiliary variables
  - number of groups for Logistic and `maxdepth` parameter for CART
  - bias/variance trad-off

Calibration adjustment

- From two steps to one step: essential constraints required by European regulations + efficiency-related constraints
- ? Constraints selection – "analyzing the reduction in coefficients of variation (CVs) and the stability of the estimates"

BDR? Can BDR be used for model/constraint selection?

Strategy

- Design + Estimation + SAE

Outline

1. Overview of the two papers
2. Innovations in the Quality of EU-SILC Survey Estimates
   - D. Chianella, M. D'Alò, S. Gerosa, M. Terribili, C. De Vitiis — Istat, Italy
3. A Constrained Optimization Approach for adjusting Misclassified Proxy Variables
   - F. Solari, V. Quondamstefano, R. Filippini, Gaia Rocchetti — Istat, Italy
4. A final note...

Overview of the paper

- Target variable y is observed on S, proxy x is observed for P
- Proxy has misclassification and structural bias $\rightarrow$ produce adjusted proxy $\tilde{x}$ for the entire population, and not only within the reference sample
- The paper proposes a deterministic alternative to weighting, stochastic adjustment, or model-based replacement. It is based on two phases:
  1. Two-step calibration from empirical confusion matrix $\hat{A}$ to adjusted matrix B
  2. Transition from class j of x to class k of $\tilde{x}$ – two strategies (1) random selection of units or (2) model-assisted reassignment
- Application to educational attainment – 2011 Population Census x, and 2023 sample survey of the Permanent Population Census y.

Overview of the paper

- Target variable y is observed on S, proxy x is observed for P
- Proxy has misclassification and structural bias $\rightarrow$ produce adjusted proxy $\tilde{x}$ for the entire population, and not only within the reference sample
- The paper proposes a deterministic alternative to weighting, stochastic adjustment, or model-based replacement. It is based on two phases:
  1. Two-step calibration from empirical confusion matrix $\hat{A}$ to adjusted matrix B
  2. Transition from class j of x to class k of $\tilde{x}$ – two strategies (1) random selection of units or (2) model-assisted reassignment
- Application to educational attainment – 2011 Population Census x, and 2023 sample survey of the Permanent Population Census y.

Questions

1st phase

1. $\hat{A} \rightarrow \tilde{A}$ via Calibration of $\hat{N}_x$ on N_x AND $\hat{N}_y$ on $\hat{N}_y$
2. $\tilde{A} \rightarrow B$ via Calibration of N_x on $\hat{N}_y$ AND $\hat{N}_y$ on $\hat{N}_y$

? Why two steps?

$\hat{A} \rightarrow B$ via Calibration of $\hat{N}_x$ on $\hat{N}_y$ directly AND $\hat{N}_y$ on $\hat{N}_y$

2nd phase

? No explicit model is used, "only" random sampling to select units to be edited. Choice of sampling method → non-informative vs. informative editing

? Moving away from "integer" transition, using "fractional" transitions

Questions

1st phase

1. $\hat{A} \rightarrow \tilde{A}$ via Calibration of $\hat{N}_x$ on N_x AND $\hat{N}_y$ on $\hat{N}_y$
2. $\tilde{A} \rightarrow B$ via Calibration of N_x on $\hat{N}_y$ AND $\hat{N}_y$ on $\hat{N}_y$

? Why two steps?
$\hat{A} \rightarrow B$ via Calibration of $\hat{N}_x$ on $\hat{N}_y$ directly AND $\hat{N}_y$ on $\hat{N}_y$

2nd phase

? No explicit model is used, "only" random sampling to select units to be edited. Choice of sampling method $\rightarrow$ non-informative vs. informative editing

? Moving away from "integer" transition, using "fractional" transitions

Questions

1st phase

1. $\hat{A} \rightarrow \tilde{A}$ via Calibration of $\hat{N}_x$ on N_x AND $\hat{N}_y$ on $\hat{N}_y$
2. $\tilde{A} \rightarrow B$ via Calibration of N_x on $\hat{N}_y$ AND $\hat{N}_y$ on $\hat{N}_y$

? Why two steps?
$\hat{A} \rightarrow B$ via Calibration of $\hat{N}_x$ on $\hat{N}_y$ directly AND $\hat{N}_y$ on $\hat{N}_y$

2nd phase

? No explicit model is used, "only" random sampling to select units to be edited. Choice of sampling method $\rightarrow$ non-informative vs. informative editing

? Moving away from "integer" transition, using "fractional" transitions

Questions

1st phase

1. $\hat{A} \rightarrow \tilde{A}$ via Calibration of $\hat{N}_x$ on N_x AND $\hat{N}_y$ on $\hat{N}_y$
2. $\tilde{A} \rightarrow B$ via Calibration of N_x on $\hat{N}_y$ AND $\hat{N}_y$ on $\hat{N}_y$

? Why two steps?
$\hat{A} \rightarrow B$ via Calibration of $\hat{N}_x$ on $\hat{N}_y$ directly AND $\hat{N}_y$ on $\hat{N}_y$

2nd phase

? No explicit model is used, "only" random sampling to select units to be edited. Choice of sampling method $\rightarrow$ non-informative vs. informative editing

? Moving away from "integer" transition, using "fractional" transitions

Title of Session 3: Multisource Analyses for Official Statistics

Title of Session 3: Multisource Analyses for Official Statistics

Subtitle: Calibration – Die hard