Estimates of Poverty Indicators Using Small Area Method

Phuong Hoang Vu, Zobaer Ahmed, Sanghyun Jeon, Khanh Lam Duy, Nathaline Onek Aparo, Ishimwe Aline

Department of Economics and Management, Università di Pisa

Introduction:

Research on living conditions conducted by official statistics allows for measurement of poverty. However, because of the methodology and costs, results of the this surveys are published on a very general level of spatial aggregation - the country and the region. Information for a more detailed sections are not available due to the small sample size, which leads to large mean square errors of obtained estimates and hence low reliability of these ratings. In order to obtain estimates at lower level of aggregation than it is published, small area estimation methods are used.

Data and Source:

We consider as target territory the Tuscany region, Italy. In this region the Local Labour Systems (LLSs) are the 57 areas of interests for many socioeconomic indictors.

- Unit level data (dataset Census.Rdata; SurveyA.Rdata)
- Population U (Tuscany regions) with N = 1,388,252 (households)
- Number of Domains: D=57, Domain sample size n_d is random Survey A with sample size n=1426 (households)

Method:

In this working, we compute SAE estimates of mean household equivalised income:

The mean household equivalised income using

1. Horvitz-Thompson estimates
2. GREG estimator and
3. EBLUP unit level estimator.

Horvitz-Thompson (HT) Estimator

$$\hat{t}_{dHT} = \sum_{k \in S_d} a_k y_k$$

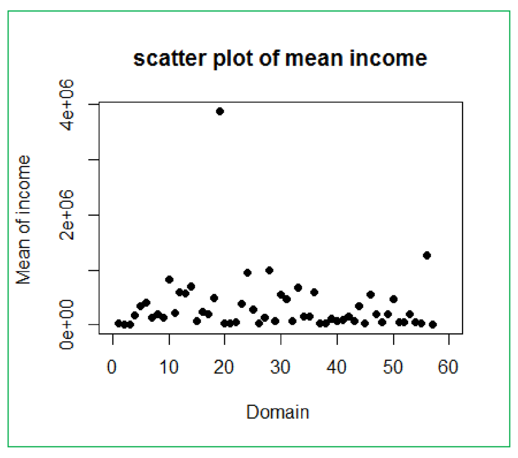
$$\hat{V}\left(\hat{t}_{dHT}\right) = \frac{1}{n_d(n_d-1)} \sum_{k \in S_d} \left(n_d a_k y_k - \hat{t}_{dHT}\right)$$

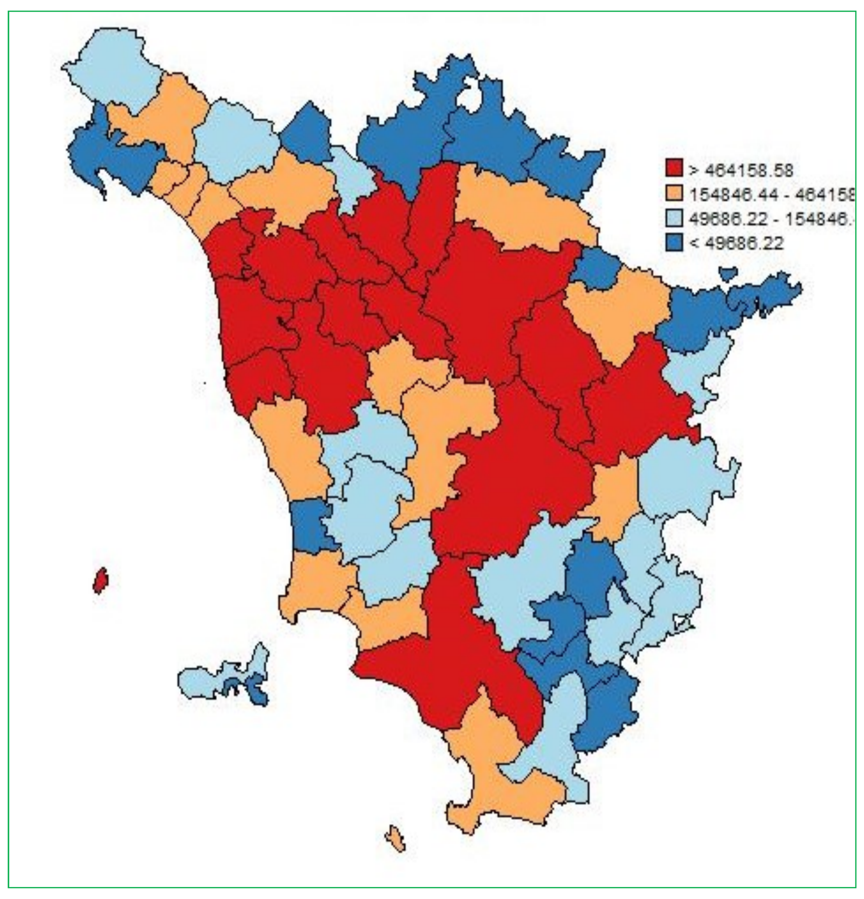
Direct estimator of domain

$$\hat{y}_{dHT} = \hat{t}_{dHT} / N_d$$

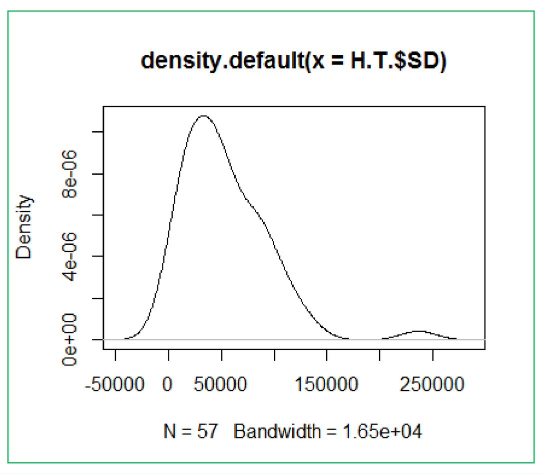
$$\hat{V}_u\left(\hat{y}_{dHT}\right) = \hat{V}_u\left(\hat{t}_{dHT}\right) / N_d^2$$

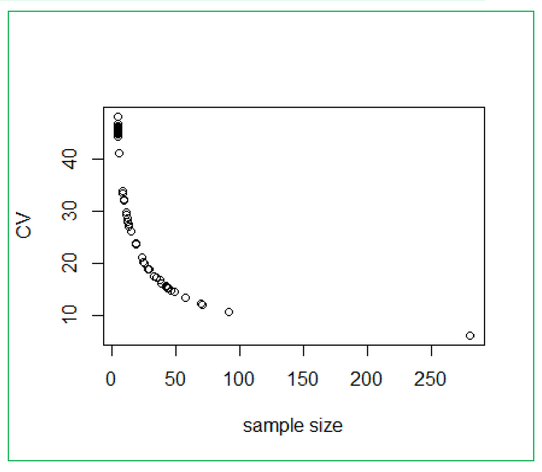
0 %	25 %	50 %	75 %	100 %
3712.683	49686.22	154846.44	464158.58	3884803.35





Domain	Sample size	Direct	SD	CV
191	5	29242.91	13098.89	44.79338
210	5	3712.68	1648.35	44.39783
214	5	11489.21	5189.33	45.16701
234	13	174399.75	50038.64	28.69192
235	26	344637.79	68747.50	19.94775
236	30	398703.99	74718.10	18.74024





Indirect GREG for Unplanned Domain

Assisting linear fixed-effects model:

$$y_k = x_k' \beta + \varepsilon_k \quad var\left(\varepsilon_k\right) = \sigma^2, k \in U$$

$\beta = \left(\beta_1, \ldots, \beta_j\right)$ is common for all domain

β is estimated from the sample by weighted least squared by weight $a_k = 1/\pi_k$

$$\hat{\beta} = \left(\sum_{k \in s} a_k x_k x_k'\right)^{-1} \sum_{k \in s} a_k x_k y_k$$

Using β to find the fitted value of income of the whole population

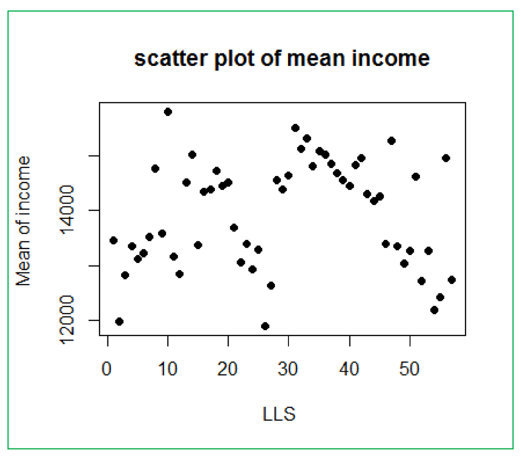
$$\hat{y}_k = x_k' \hat{\beta}, k \in U$$

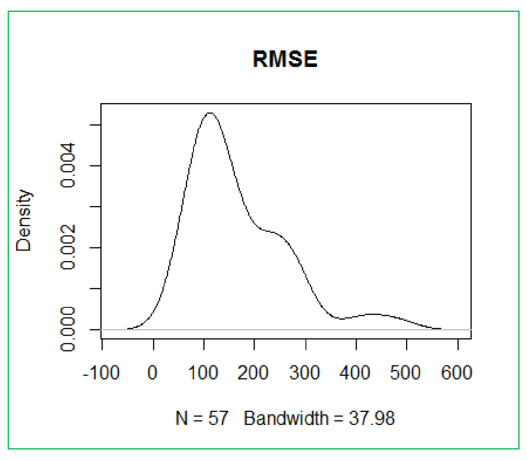
Then the Direct GREG estimator for each domain could be calculated as: $\hat{t}_{dGREG} = \sum_{k \in U^d} \hat{y}_k + \sum_{k \in s_d} a_k\left(y_k - \hat{y}_k\right) = \sum_{k \in U^d} \hat{y}_k + \sum_{k \in s_d} a_k e_k$

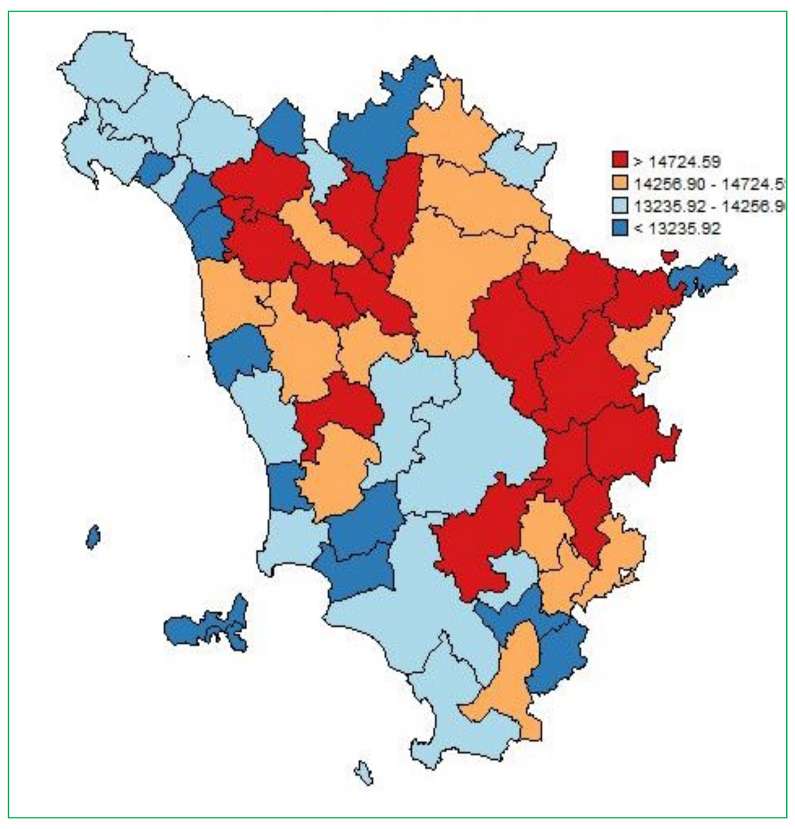
Approximate variance estimator of indirect GREG for unplanned domains by using extended residuals:

$$\hat{V}_u\left(t_{dGREG}\right) = \frac{n}{n-1} \sum_{k \in s} \left(a_k e_{dk} - \frac{\hat{t}_{dHTe}}{n}\right)^2$$

0 %	25 %	50 %	75 %	100 %
11901.11	13235.92	14256.90	14724.59	15801.76







EBLUP unit level estimator

$$y_{ij} = x_{ij}^T \beta + d_{ij} u_i + \varepsilon_{ij},\ j = 1, \ldots, n_i$$

β the vector of random coefficients

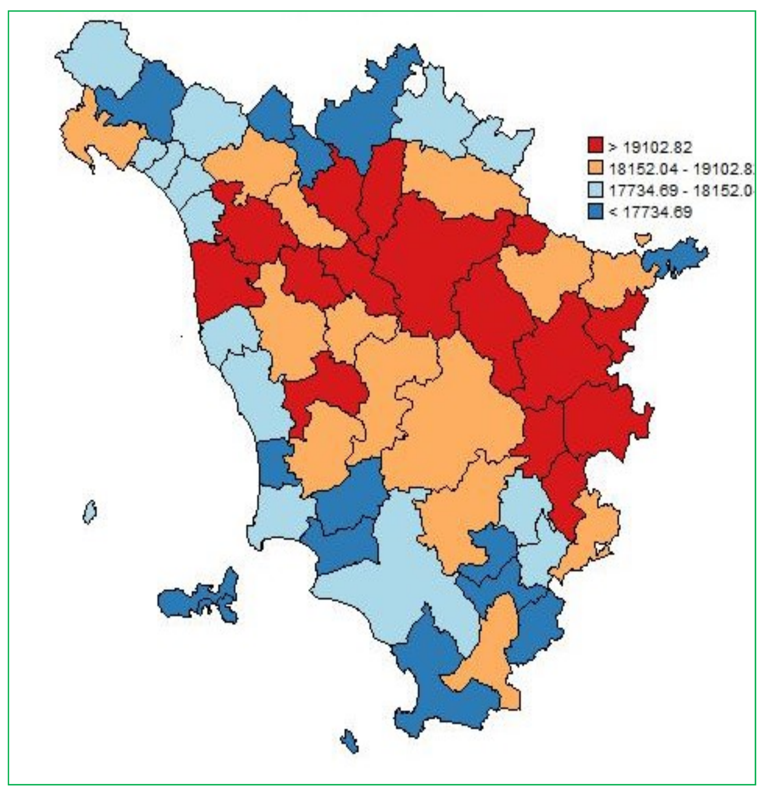
u denotes a random area effect that characterizes differences in the conditional distribution of y given x between the m small areas

d a constant value which is for all units in the population

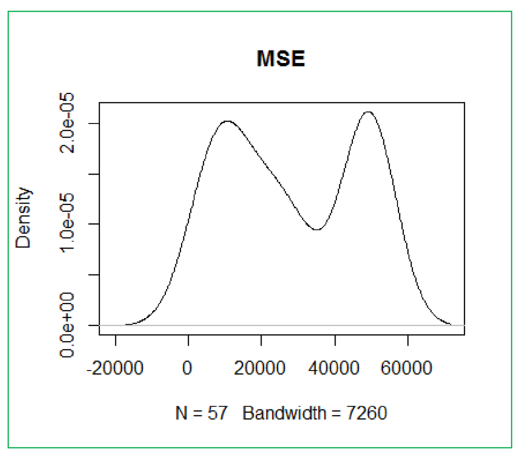
e error term associated with the j-th unit in the i-th area

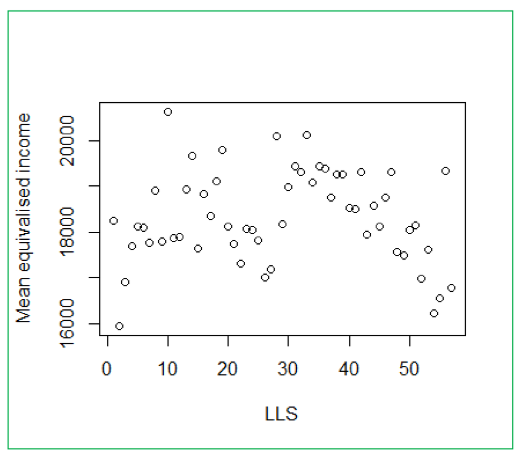
Eblup of the mean of the area

$$\theta_i^{MX} = \left(\sigma_u, \sigma_e\right) = N_i^{-1}\left[\sum_{j \in S_i} y_j + \sum_{j \in S_i} \hat{y}_j\right] \qquad \hat{y}_{ij} = x_{ij}^T \hat{\beta} + d_{ij} \hat{u}_i$$

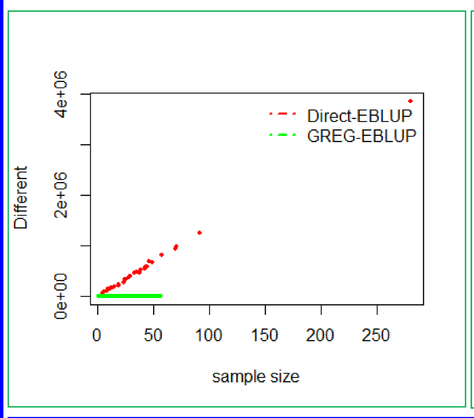


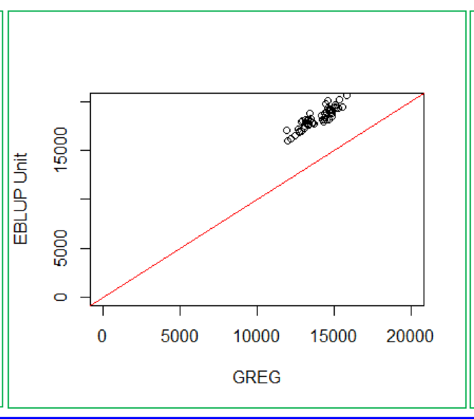
0 %	25 %	50 %	75 %	100 %
15932.50	17734.69	18152.04	19102.82	20644.41

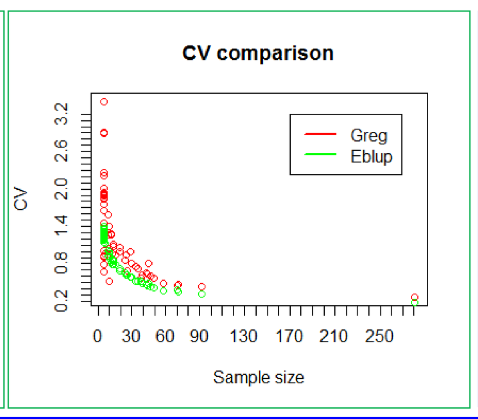




Comparison and Conclusion







1. While GREG and EBLUP are unbiased; **2.** The coefficient of variation (CV) of the EBLUP is lower than the GREG; **3.** With some domains have small sample size (less than 10), The GREG estimator has lower CV than EBLUP; **4.** Variance may be small even for small domains in EBLUP; **5.** All of estimators using GREG and EBLUP method satisfy Canadian Statistical Institute Standard (CV<16.5%), while only 12/57 of domain estimators using HT has CV<16.5%; **6.** With our dataset and domains, we can conclude that EBLUP estimator is better because it has small CV

References:

1. J. N. K. Rao and Isabel Molina, Small Area Estimation (2nd Edition), Wiley Interscience, 2003
2. Michael Hidiroglou, Small-Area Estimation: Theory and Practice , American Statistical Association, 2007