

Unemployment Estimates for Local Labour Market Areas on the basis of Labour Force Survey

Stime della Disoccupazione a livello dei Sistemi Locali del Lavoro basate sull'Indagine Forze di Lavoro

Loredana Di Consiglio, Stefano Falorsi, Fabrizio Solari
ISTAT MPS/A, Via A. Depretis 74/b, 00184, Roma,
diconsig@istat.it - stfalors@istat.it - solari@istat.it

Riassunto: In questo lavoro vengono presentati i risultati di uno studio empirico per la scelta dello stimatore *indiretto* da adottare nella produzione delle stime sull'occupazione a livello dei sistemi locali del lavoro (SLL), che sono aree sub-provinciali costituite da aggregazioni di comuni. Tali aree sono costruite sulla base di informazioni di tipo socio-economico desunte dall'ultimo censimento generale della popolazione. Gli stimatori *diretti* delle suddette quantità, ottenuti tramite l'indagine Forze Lavoro (FL), risultano generalmente poco efficienti in quanto, in molti SLL, il numero di comuni campione osservati è esiguo ed in alcuni di essi addirittura nullo. Lo studio empirico si basa su simulazioni del campione FL condotte a partire dalla popolazione censuaria.

Keywords: Small area estimation, Synthetic estimator, Composite estimator

1. Introduction

The present work is part of a study undertaken for the production of estimates of employment and unemployment at Local Labour Market Areas' (LLMAs) level. LLMAs are geographical areas designed every ten years on the basis of census population data. The sample information available for the evaluation of the estimates is provided by the Labour Force Survey (LFS). Since the survey design is planned at provincial level, the sample data at LLMA level may be limited or absent and the direct estimates, obtained using *calibration estimator*, currently applied for the dissemination of the survey estimates, may be too unstable or not defined. The aim of this work is to evaluate, on the basis of an empirical study, the performances, expressed in terms of bias and mean square error, of different composite estimators.

2. Parameter of interest and estimator

Let d ($d=1, \dots, D$) be a LLMA of a given geographical region. Since the LLMAs may cut across provinces, the total of interest referred to the small area d is defined by

$$Y = \sum_{a=1}^A \sum_{p=1}^{L_d} \sum_{h=1}^{H_{pd}} \sum_{i=1}^{N_{hd}} \sum_{j=1}^{M_{hi}} Y_{ahij} = \sum_{a=1}^A d Y_a, \quad (1)$$

where: a ($a=1, \dots, A$) is a post-stratum; L_d denotes the provinces including part of territory of the small area d ; H_{pd} are the strata of province p including the small area d ; N_{hd} is the number of PSUs of stratum h belonging to the small area d ; M_{hi} is the number of households of stratum h belonging to PSU hi .

Besides the calibration estimator ${}_{PV}\hat{Y}_{td}$ currently applied for the LFS estimation (Falorsi and Falorsi, 1994) and the post-stratified ratio estimator

$${}_R\hat{Y}_d = \sum_{a=1}^A \frac{{}_{PV}\hat{Y}_{tda}}{{}_{PV}\hat{P}_{tda}} P_{tda}$$

several composite estimators have been considered. All the examined composite estimators are obtained as linear combinations of ${}_R\hat{Y}_d$ with the synthetic estimator

$${}_S\hat{Y}_{td} = \sum_{a=1}^A \frac{{}_{PV}\hat{Y}_{tda}}{{}_{PV}\hat{P}_{tda}} P_{tda} = \sum_{a=1}^A \frac{{}_{PV}\hat{Y}_{tda}}{P_{tda}} P_{tda}.$$

The first form of composite estimator considered is given by

$${}_{C_1}\hat{Y}_t = \alpha_d {}_R\hat{Y}_{td} + (1 - \alpha_d) {}_S\hat{Y}_{td}, \quad (2)$$

where, for each LLMA d , the weight α_d ($0 \leq \alpha_d \leq 1$) is obtained minimizing the MSE of ${}_{C_1}\hat{Y}_t$ and is computed according to the following formula

$$\alpha_d = \frac{\text{MSE}({}_S\hat{Y}_{td})}{\text{MSE}({}_S\hat{Y}_{td}) + \text{Var}({}_R\hat{Y}_{td})} = \frac{\text{Var}({}_S\hat{Y}_{td}) + \text{Bias}^2({}_S\hat{Y}_{td})}{\text{Var}({}_S\hat{Y}_{td}) + \text{Bias}^2({}_S\hat{Y}_{td}) + \text{Var}({}_R\hat{Y}_{td})}. \quad (3)$$

The variances $\text{Var}({}_S\hat{Y}_{td})$ and $\text{Var}({}_R\hat{Y}_{td})$ are evaluated on census data by means of the linearization method proposed by Woodroof (1971). Also the bias of ${}_S\hat{Y}_{td}$ is evaluated using 1991 census population data.

A second form of composite estimator ${}_{C_2}\hat{Y}_{td}$ (Purcell and Kish, 1979) is

$${}_{C_2}\hat{Y}_{td} = \alpha {}_R\hat{Y}_{td} + (1 - \alpha) {}_S\hat{Y}_{td}, \quad (4)$$

in which a common weight α ($0 \leq \alpha \leq 1$), minimizing the MSE of ${}_{C_2}\hat{Y}_{td}$, is applied to the direct estimator in all the LLMAs, and is given by

$$\alpha = \frac{\sum_{d=1}^D \text{MSE}({}_S\hat{Y}_{td})}{\sum_{d=1}^D \text{MSE}({}_S\hat{Y}_{td}) + \sum_{d=1}^D \text{Var}({}_R\hat{Y}_{td})}. \quad (5)$$

The third composite estimator under study, $_{C_3}\hat{Y}_{td}$, is similar to the one in formula (4) with weights α_g s constant on subsets of the LLMA. Three subsets have been defined on the basis of LLMA's demographic sizes. The percent proportion population over the total regional population in the first group ranges from 0.12% to 1.73%; the group is composed by 19 LLMA; the percent proportion in the second group ranges from 1.9% to 5.60%; the group is composed by 7 LLMA; the third group is constituted by the largest LLMA representing a percent proportion equal to 64%. The LLMA are divided up into these three groups because the MSE is expected to be larger for the smaller LLMA which have smaller estimates.

Since the weights (3) and (5) are computed using 1991 census data and the estimates, obtained on the basis of the composite estimators (2) and (4), refer to the period 1996-1999, in the empirical study a particular analysis has been carried out in order to evaluate the impact on MSE due to the use of out of date information for the computation of the α_d s and α . For this reason, the composite estimators have been evaluated with the α_d s and α computed using 1981 census data. The corresponding estimators are identified as $_{C_4}\hat{Y}_{td}$ and $_{C_5}\hat{Y}_{td}$.

3. Empirical study

In order to illustrate the properties of the small area estimators considered before, a simulation study has been carried out. Two thousands repeated draws, according to a two stage sample design with strata, clusters and sample size alike to those used in LFS survey, were taken from 1991 census population related to one Italian region (Lazio). Estimates of employed, unemployed and persons seeking their first job are based on auxiliary information given by sex and age of the members of the sampled households; the LLMA considered are the 27 LLMA centred in the region. The following performance measures were evaluated to assess the bias and the MSE of the examined estimators on the overall R=2.000 replications and over all the D=27 small areas: Average Absolute Relative Bias (AARB); Average Relative Root Mean Square Error (ARRMSE), expressed respectively by formulas:

$$\text{AARB} \left({}_T\hat{Y} \right) = \frac{1}{D} \sum_{d=1}^D \left| \frac{1}{R} \sum_{r=1}^R \left[\frac{{}_{T,d}\hat{Y}(r) - {}_dY}{{}_dY} \right] \right| 100,$$

$$\text{ARRMSE} \left({}_T\hat{Y} \right) = \frac{1}{D} \sum_{d=1}^D \sqrt{\frac{1}{R} \left(\sum_{r=1}^R \left[\frac{{}_{T,d}\hat{Y}(r) - {}_dY}{{}_dY} \right]^2 \right)} 100,$$

in which ${}_{T,d}\hat{Y}$, ${}_{T,d}\hat{Y}(r)$ indicate, respectively, the generic small area estimator T examined in section 2 and its value obtained in the r -th of the 2.000 samples.

4. Analysis of the results

In table 1 the AARBs and the ARRMSSEs of the proposed estimators are reported. As expected, the composite estimator with optimal weights, α_d , computed on 1991 data, i.e. on the same data used for the analysis, gives better results in terms of ARRMSSE. All the examined composite estimators behave very similarly and the use of out of date information does not greatly affect the performance of the estimators. The composite estimators perform almost similarly, also in terms of AARB, but the high values, especially for the variable Unemployed, indicates that the underlying model should be improved.

Table 1. AARB and ARRMSSE% for Employed/Unemployed/Searching 1st job.

	Employed		Unemployed		Search 1st Job	
	AARB	ARRMSSE	AARB	ARRMSSE	AARB	ARRMSSE
Calibration PV	3.62	76.52	4.21	97.99	3.86	84.44
Post-stratified R	25.49	38.48	25.84	58.05	25.35	46.13
Synthetic	4.66	4.82	18.12	19.38	9.55	10.29
Composite C ₁	4.09	4.36	15.51	17.25	8.75	9.70
Composite C ₂	4.14	4.53	16.42	18.73	8.98	10.06
Composite C ₃	4.21	4.53	16.78	18.57	9.06	9.97
Composite C ₄	4.13	4.41	17.07	19.38	8.89	9.89
Composite C ₅	4.23	4.55	16.63	18.71	9.04	10.06

REFERENCES

- Falorsi, P.D., Falorsi S. (1994) Stime trimestrali a livello provinciale per l'indagine sulle forze di lavoro, *Quaderni di Ricerca Metodologia e Informatica*, 3, ISTAT.
- Gosh, M. and Rao, J.N.K. (1994). Small Area Estimation: An Appraisal. *Statistical Science*, 9, 55-93.
- Purcell, N.J., Kish L. (1979) Estimation for small domain, *Biometrics*, 35, 365-384
- Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag. New York.
- Schaible, W. L. (1978). Choosing Weights for Composite Estimators for Small Area Statistics. In *Proceedings of the American Statistical Association, Survey Research Section*, 741-746.
- Singh, M.P., Gambino, J. and Mantel, H. J. (1994). Issues and Strategies for Small Area Data. *Survey Methodology*, 20, 3-22.
- Woodroof, R. (1971), A simple method for approximating the variance of complicated estimate, *Journal of the American Statistical Association*, 66, 411-414 .