

A Comparison among Two Stage Sampling Plans Based on Different Time Stratifications for the Redesign of the Italian Labour Force Survey

*Un Confronto tra Disegni Campionari a Due Stadi Basati su Differenti
Stratificazioni nel Tempo per il Ridisegno dell'Indagine sulle Forze di
Lavoro*

Loredana Di Consiglio
Stefano Falorsi
ISTAT
Via A. Depretis 74b
00186 Roma
diconsig@istat.it
stfalors@istat.it

Claudia De Vitiis
ISTAT
Via A. Ravà 150
00142 Roma
devitiis@istat.it

Riassunto: La strategia campionaria dell'indagine sulle Forze di Lavoro è stata ridisegnata per soddisfare i requisiti richiesti dall'Unione Europea relativi alla precisione delle stime e alla continuità della rilevazione nel tempo. In questo lavoro vengono posti a confronto tre disegni alternativi che prevedono differenti stratificazioni nel tempo e la loro efficienza viene valutata mediante approccio simulativo.

Keywords: Time stratification, Labour Force Survey, Sample design

1. Introduction

The sampling strategy of the Italian Labour Force Survey (LFS) has been redesigned in order to meet the UE requirements on precision of the estimates and continuity of data collections. The current survey is carried on in one fixed week each quarter, whereas the new LFS will be based on a uniform and continuous collection. This work will focus on the comparison among three different time stratifications, that allows to meet the last requirement, and will compare their performance in terms of variances. In section 2 the sampling design is described with particular attention to time stratification, in section 3 the carried out empirical work is illustrated and the main results are reported.

2. The sampling design

The sampling plans considered in this work are sampling designs *with stratification in space and in time*. The theoretical basis for this sampling structure is described in Vos (1964), where different sampling schemes are proposed for the selection from a population in the different times in which the reference period is divided. The methods are such that the resulting sample is representative for the estimates of the parameters both in space and in time. In the particular case of the LFS, the reference period for the estimates is the quarter, which is subdivided in thirteen weeks; hence, it is necessary to stratify the sample not only in space but also in time and to select a sample of weeks according to a fixed design.

2.1 Stratification in space

With regards to the *space stratification*, the considered sampling designs retain the main features of the current survey: the sample is a two stage design with stratification of the primary sampling units (PSU); the municipalities are the PSUs and the households are the secondary sampling units. Within each province the PSUs are classified as Self Representative (SR) units and Non Self Representative (NSR) units according to a threshold based on demographic size: a SR municipality is itself a stratum, whereas NSR units are stratified in strata of approximately equal size in terms of resident population. The municipalities are chosen with probabilities proportional to size (Hansen e al. 1953), while the households are selected with equal probabilities. For further details on methodological features see ISTAT (1993).

2.2 Stratification in time

After defining the sampling scheme in space, it has been necessary to define the stratification scheme in time. Three different stratifications have been studied: weekly, monthly and quarterly stratification. With the above time-stratifications each stratum of municipalities is represented every week, every month or every quarter respectively. Space and time stratification methods are strictly tied, because the choice of some parameters, which define the space stratification, result from the choice of the time stratification scheme. In fact, for the weekly stratification, it is necessary to involve each SR municipality every week, to select 13 municipalities from each NSR stratum, to assign each of them to a week of the quarter and to divide the sample size of each stratum in 13 sub-samples. For the monthly stratification, each SR municipality is involved three times in the quarter (one per month), three municipality are selected from each NSR stratum and are randomly assigned a week in different months. Finally, for the quarterly stratification, one municipality is selected from each NSR stratum and a week has to be selected both for SR and NSR municipalities. The random assignment of weeks to municipalities has to guarantee that the sample in the 13 weeks is distributed as uniformly as possible. The weekly stratification has been considered only as a benchmark, since it has been decided not to utilize it because of operational reasons.

3. The comparison of the three designs

3.1 The framework

The three designs have been compared on the basis of their variances. An exact expression of variances is derived in Vos (1964) for the case of simple random sampling of elementary units. In order to be valid for the LFS case, these expressions should be extended to the *pps* two stage sampling; moreover, for their evaluation, some population parameters, such as variance among times, should be known. These parameters are not available since weekly collected data do not exist, so an empirical approach has been pursued: repeated samples were selected from four artificial populations to obtain a numerical evaluation of variance. The four populations have been generated by using 1991 census aggregated data of employed and unemployed people in the Italian municipalities and a series of sixteen quarterly estimates of the same aggregates (employed and unemployed) from April 1996 to October 1999. On the

basis of this information and assuming four different evolution rules in the thirteen weeks of a quarter, the populations have been created as follows:

- 1) estimates of relative differences between consecutive quarters have been calculated either at regional and national level;
- 2) the two consecutive quarters (t, t+1) with median difference have been identified;
- 3) the relative differences at stratum level have been evaluated for the pair of quarters t and t+1;
- 4) the values for week 1 for each of the 8095 municipalities at 1991 census have been given by the census quantities;
- 5) the values for the 13th week have been obtained by multiplying the initial values for the relative difference of the stratum to which the municipality belongs;
- 6) the eleven intermediate weekly values have been obtained by assuming one of the following evolution rules: a) linear trend; b) geometric trend; c) broken linear trend; the evolution rates are obtained from the relative differences of step 3).

A constant population, each weekly value equal to value of week 1, has been taken as reference population.

3.2 Simulation

For each artificial population and for each of the three stratifications under study, R=500 replications of the sample selection have been carried out; the sample selection has consisted of:

- 1) selection of the municipalities from each *space-stratum*;
- 2) random assignment of sample PSUs to weeks;
- 3) generation, for each sample municipality i , of a value from a random variable distributed as a Binomial with parameters $p = p_i = {}_iY_{(w)}/{}_iN$ and $n = {}_in$, where ${}_iN$ is the census population, ${}_in$ the quarter sample size and ${}_iY_{(w)}$ the number of employed (or unemployed) for the week w ($w=1, \dots, 13$), according to one of the four artificial populations; the generation from a Binomial distribution allows us to reproduce approximately the real case, in which a simple random sample of clusters (households) is selected in each municipality; case that can be approximated by a simple random sample of elementary units;
- 4) computation, at regional and national level, of the quarterly estimates of employed and unemployed by using the expansion estimator.

The efficiency of the stratification methods has been evaluated by means of the following index:

- Percent Relative Root Mean Square Error (*RRMSE*)

$$RRMSE(\hat{Y}) = \left(\frac{\sqrt{MSE(\hat{Y})}}{Y} \right) 100, \text{ where } MSE(\hat{Y}) = \frac{1}{R} \sum_{r=1}^R [\hat{Y}(r) - Y]^2.$$

where $\hat{Y}(r)$ indicates the r -th estimate and Y the 'true' value of the parameter.

The Percent Relative Bias (*RB*) has been evaluated to verify whether the number of replications is large enough to ensure the bias to be null (the estimators used are unbiased) so that the simulated MSE values are stable enough:

$$RB(\hat{Y}) = \left(\frac{1}{R} \sum_{r=1}^R \frac{[\hat{Y}(r) - Y]}{Y} \right) \cdot 100$$

In Table 1 the values of *RRMSE* and *RB* of the three time stratifications are reported for the estimates of the unemployed of the Linear Trend Population at regional level. What emerges is that the monthly stratification is the more satisfactory both for its statistical properties and its practicability. The lower efficiency of the weekly stratification could be partially explained by the fact that in practice this design does not allow to stratify the municipality within each province.

For space reasons, it is not possible to report the results for all the examined populations and for the estimates of Employed, although the differences are less pronounced for the latter that has a lower sampling error. Similar conclusions have been obtained with the other populations. The schemes satisfy the EU requirements.

Table 1: *RRMSE and Relative Bias of the different stratification schemes for the estimates of the Unemployed of the Linear Trend Population*

	RRMSE			Relative Bias		
	Week	Month	Quarter	Week	Month	Quarter
Piemonte	5,05	5,87	6,16	0,01	-2,42	-2,88
Valle D'Aosta	11,19	9,32	10,27	-3,46	1,99	2,66
Lombardia	5,25	6,07	6,92	0,81	-1,51	-2,81
Trentino A. A.	13,71	15,05	18,24	4,63	-0,47	5,86
Veneto	6,94	6,28	7,09	3,02	-1,51	1,46
Friuli V. G.	8,93	7,44	10,94	0,45	-2,22	-1,01
Liguria	6,25	7,78	7,80	0,35	1,90	-2,89
Emilia R.	9,92	7,75	6,66	2,73	2,48	-0,24
Toscana	5,04	5,47	5,29	2,22	2,16	1,79
Umbria	6,28	7,93	8,86	-3,91	-2,69	-1,42
Marche	9,13	12,56	10,51	3,39	1,67	-2,57
Lazio	5,18	3,62	3,17	2,57	0,41	0,46
Abruzzo	9,05	6,40	8,41	2,99	1,10	2,86
Molise	6,59	6,08	8,08	0,81	-3,07	2,08
Campania	3,24	2,22	2,58	1,31	-0,48	-0,44
Puglia	2,88	3,01	2,99	-0,73	-0,38	-1,05
Basilicata	8,65	5,14	6,75	2,83	-0,59	-0,70
Calabria	10,10	9,41	8,88	-3,97	-3,06	-3,80
Sicilia	3,65	2,33	2,64	2,79	1,29	-0,53
Sardegna	3,99	5,37	4,44	1,58	-0,74	-1,01
Italia	1,65	1,21	1,52	0,22	-0,52	-1,01

References

- Hansen M. H., Hurwitz W. N. e Madow W. G. (1953) *Sample Survey Methods and Theory*, Wiley, New York .
- ISTAT (1993) *Indagine Multiscopo sulle Famiglie. Anni 1987-91 Vol 1. Obiettivi, Disegno e Metodologia dell'Indagine*.
- Vos J. W. E. (1964) Sampling in space and time, *Review of the International Statistical Institute*, 32, 226-241.