

**EU-SILC COMPLEX INDICATORS: THE IMPLEMENTATION OF
VARIANCE ESTIMATION**

Diego Moretti, Claudia Rinaldelli

ISTAT, Servizio Condizioni economiche delle famiglie

Summary

EU-SILC complex indicators: the implementation of variance estimation

Relative poverty measures and inequality indicators will be estimated by EU-SILC survey (Statistics on Income and Living Conditions survey) with their sampling errors and design effect. This paper reports the study and software procedures developed to evaluate the sampling variance of these complex cross-sectional statistics.

Sommario

Procedure software per la stima della varianza campionaria degli indicatori complessi di povertà relativa e disuguaglianza dell'indagine EU-SILC

E' sempre più frequente la produzione di stime complesse da un punto di vista funzionale per le quali non si può applicare direttamente la metodologia standard di stima della varianza campionaria. Una serie di indicatori complessi di povertà relativa e disuguaglianza saranno stimati (con i corrispondenti errori di campionamento ed effetto del disegno) dai dati dell'indagine europea su reddito e condizioni di vita (EU-SILC). Il presente lavoro ha lo scopo di illustrare gli studi effettuati con riferimento a queste misure complesse, le soluzioni metodologiche individuate e le procedure software sviluppate per la valutazione dei corrispondenti errori campionari ed effetto del disegno.

INDEX

1. Introduction
2. Experiences of complex statistics
3. The methodological approach
4. The software procedures
 - 4.1. Important features of the software procedures
- Appendix 1. The EU complex statistics
- Appendix 2. The linearized variables
- Appendix 3. The Balanced Repeated Replications technique

EU-SILC complex indicators: the implementation of variance estimation¹

1. Introduction

Complexity is a feature of data coming from complex sample surveys carried out by the National Statistical Institutes. Users often omit complexity in their analyses because many software packages don't handle complex designs; on the other hand, it could be not simple to use software packages planned to do that. When estimating sampling variance, it's recommended to consider complexity (Skinner, 1989); actually, sampling variance is used to measure the precision of the estimates and in inference techniques. For this reason, at ISTAT two software procedures were developed in SAS (Statistical Analysis System): SGCE² and GENESEES³ provide, among other functions, the evaluation of sampling variance (and the indicators depending on it) of the most common estimates (frequencies, means, totals) handling complex sampling designs (Falorsi, Rinaldelli, 1998; ISTAT, 2002a); these software procedures are based on the standard methodology⁴ for variance estimation (ISTAT, 1989; Särndal et al., 1992; Wolter, 1985). The development of independent and generalized software is a common practice in many National Statistical Institutes.

Since the last three years in ISTAT the production of *complex statistics* has been coming to the attention; we mean statistics that are *complex* because come from *complex* surveys and *above all* they are expressed by *complex* functions. The standard methodology and therefore the current software procedures of variance estimation can not be directly applied to these statistics. This paper reports the Italian experience in evaluating the sampling variance of complex cross-sectional statistics as the relative poverty measures and inequality indicators estimated by EU-SILC survey (Statistics on Income and Living Conditions survey). In particular, section 2 reports the most important experiences of complex statistics studied in

¹ Diego Moretti wrote section 4.1; Claudia Rinaldelli wrote sections 1,2,3,4 and appendixes 1,2,3.

² SGCE, Sistema Generalizzato di Calcolo degli Errori di Campionamento (Generalized System for Sampling Errors Estimation).

³ Software GENESEES (GENERalised Software for Sampling Errors Estimation in Surveys) is on www.istat.it.

⁴ *Standard methodology* means well known and widely applied methodology; literature on sampling theory for finite population provides formulas to calculate sampling variance for the most used sampling designs and estimators.

ISTAT; section 3 summarizes the methodological studies (linearization and resampling) made to calculate the sampling variance of the EU complex indicators while section 4 explains the software solution developed to do that; lastly, section 4.1 describes the most important features of these software procedures. Appendix 1 reports the EU complex statistics involved in this paper while Appendix 2 contains the obtained linearized expressions for some of the EU complex statistics, lastly Appendix 3 summarizes the Balanced Repeated Replication technique.

2. Experiences of complex statistics

One relevant case of *complex* statistics happened following the agreement between ISTAT and ‘Ministero delle Finanze e dell’Economia’⁵ to produce relative poverty estimates⁶ at regional level (ISTAT, 2003); the evaluation of the sampling variance of these estimates was one important methodological aim in the period 2002-2003 (Coccia et al., 2002). To obtain a valid evaluation of sampling variance, linearization and resampling techniques⁷ were used and therefore new computer codes were developed (De Vitiis et al., 2003; Pauselli, Rinaldelli, 2004a-2004b-2004c).

The Statistics on Income and Living Conditions survey⁸ (EU-SILC) proposes a further important case of complex statistics. Actually, this survey has the aim to estimate statistics on income, living conditions, poverty; among these statistics there are the EU relative poverty and inequality cross-sectional indicators that will be estimated with their sampling errors and design effect (deft) (Regulation, 2003). Most of these are complex, in particular: at risk of poverty threshold, at risk of poverty rate, Gini index, gender pay gap, relative median at risk of poverty gap, income quintile share ratio (EUROSTAT, 2004a-2004b) (see

⁵ Ministry of Economy and Finance.

⁶ In Italy official poverty estimates are calculated yearly by ISTAT using the data of the Households Budget sample survey (ISTAT, 2004). It is defined as poor a household whose monthly consumption expenditure is equal or below to a threshold called the *standard poverty line*. The *standard poverty line* is the estimated average monthly pro capite expenditure for consumption of a household with two members; an equivalence scale is used to correct the standard poverty line when households have different sizes (ISTAT, 2002b).

⁷ Jackknife and Balanced Repeated Replication techniques were experimented.

⁸ This survey is carried out under European Regulation; in Italy the first year of survey is 2004.

Appendix 1); once more, the standard methodology and the current software procedures of variance estimation can not be directly applied to these statistics.

3. The methodological approach

Works by Moretti, Pauselli, Rinaldelli (2004a-2004b) report the methodological studies and applications experimented to evaluate the sampling errors and design effect of EU complex statistics. First of all, four of these measures (at risk of poverty threshold, at risk of poverty rate, Gini index, gender pay gap) were linearized by the ‘estimating equations’ approach and the Taylor–Woodruff method (see Appendix 2); furthermore, a resampling approach by Balanced Repeated Replication technique (BRR) was considered (see Appendix 3). After that, the values of sampling errors and design effect were experimentally obtained by linearized variables and Balanced Repeated Replication technique both, in order to check their performance. Data from ECHP (European Community Household Panel) were used in these applications because data from EU-SILC survey were not available (ISTAT, 2002c).

Table 1 shows the values of sampling errors and design effect of EU measures by linearization and Balanced Repeated Replication techniques; we observe that the two different approaches leads to similar values of sampling errors and design effect (Moretti et al., 2004a-2004b).

Considering the results of this study and the previous experience on relative poverty from Households Budget survey, the following way of calculation was chosen: sampling errors and design effect will be estimated by linearization approach for the ‘at risk of poverty threshold’, ‘at risk of poverty rate’, ‘Gini index’, ‘gender pay gap’ and by Balanced Repeated Replication technique for the ‘relative median at risk of poverty gap’ and the ‘income quintile share ratio’.

Table 1. Sampling errors and design effect of EU statistics by linearization and BRR approach (*)

	Relative sampling errors (%)		Deft	
	linearization	BRR	linearization	BRR
At risk of poverty threshold	1.72	1.73	3.08	3.01
At risk of poverty rate	3.55	3.53	2.78	2.76
Gini index	1.71	1.60	2.87	2.80
Gender pay gap	30.51	32.61	1.00	1.07
Relative median at risk of pov. gap	---	6.61	---	2.86
Income quintile share ratio	---	2.88	---	3.00

(*) Table from Moretti, Pauselli, Rinaldelli (2004a-2004b)

4. The software procedures

Two new software procedures were developed in SAS to make available the methodological solution described in paragraph 3 (Moretti et al., 2004a-2004b). Actually the software procedures already used in ISTAT don't automatically manage linearized expressions and don't perform resampling techniques; on the other hand, external software packages don't solve completely the methodological and software problems referring to the complex estimates coming from the sample surveys of ISTAT. The implementation of software for the Balanced Repeated Replications technique has required to manage a complex computational situation⁹ due to the used approach (resampling), to the complex function of such measures, to the complexity and size of the sampling design involved (see section 4.1). For the approach based on linearization, a software procedure was developed to use automatically the linearized variables in software SGCE. At the moment, these software procedures handle the two-stages sampling designs used in the surveys on households.

4.1. Important features of the software procedures

The software procedures were developed on the basis of a user-friendly approach. Actually these procedures display a sequence of screens where users are allowed to type in the statistic of interest, the parameters of the sampling design involved, the name of the input data file

and so on; furthermore, users can easily read the final results. To realize that, these procedures were associated with a basic graphic interface and a post-processing procedure. The graphic interface allows to generate the sequence of screens; for example, figure 1 shows the screen where users can type in the statistic of interest.

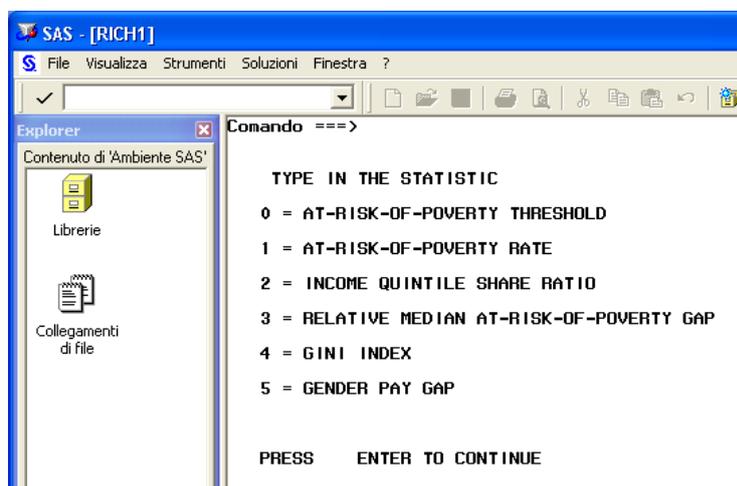


Fig.1. The screen of the EU statistics

The post-processing procedure manipulates the output of calculation to provide a complete and final spreadsheet report for the specified statistic. Figure 2 shows an example of final screen where is reported the table containing the values of the absolute and relative percentage sampling errors and deft for the estimate of the *at risk-of-poverty rate* at national level and for breakdown by age.

⁹ When applying BRR technique to simple random sampling design (to estimate design effect), the dimension of Hadamard's matrix was considerable.

	A	B	C	D	E
1	AT-RISK-OF-POVERTY RATE				
2					
3	ESTIMATE %	ABSOLUTE SAMPLING ERROR	RELATIVE SAMPLING ERROR %	DEFT	
4	20.48	0.73	3.55	2.78	TOTAL
5	21.51	1.84	8.58	1.62	0 <= AGE <= 15
6	25.10	1.45	5.78	1.72	16 <= AGE <= 24
7	21.69	0.80	3.68	1.92	25 <= AGE <= 49
8	19.56	0.85	4.34	1.48	50 <= AGE <= 64
9	17.41	0.90	5.16	1.74	AGE >= 65
10					
11					
12					

Fig. 2. Example of final spreadsheet report

The purpose of the software procedure based on linearization approach is to use automatically the linearized variables in software SGCE. This procedure is composed of different modules; each module performs a different task. When users select one statistic, the procedure will apply the correct sequence of modules to calculate respectively the value of the statistic, the appropriate linearized variable and will call the software SGCE. Furthermore, this procedure allows to compute the values of the statistic of interest, sampling errors and deft for specified breakdowns.

The second software procedure is based on BRR approach; in order to calculate the value of deft, firstly BRR technique is applied to the current sampling design (two-stages sampling design) and then to the simple random sampling (SRS) design. This procedure is composed of two modules, the first for the two-stages sampling design and the second for the SRS design. The post-processing procedure combines the results coming from the complex sampling design with the SRS design to compute the deft. Furthermore, this procedure allows to calculate the values of the statistic of interest, sampling errors and deft for specified breakdowns.

We mention three important features implemented in the BRR software procedure:

- 1) the extension of the basic BRR to the current complex sampling design;
- 2) the memory space management of the Hadamard's matrix on which BRR is based;
- 3) a modular development.

With reference to point 1), we underline that BRR was implemented in the extended version for the complex two-stages sampling design as described in Appendix 3.

Referring to point 2), we experimented some problems about the memory space of our desktop personal computers. The arrangement of the Hadamard's matrix in SRS design needed more of 8 GB of memory, that is an huge amount for a single personal computer; actually, the rank of the Hadamard's matrix depends on the total number of sampling units and on the sampling design (see Appendix 3). So, some solutions were searched; then, the partially-balanced repeated replications technique was implemented in SRS design. This version consists in grouping the sampling units with the aim to reduce the rank of the Hadamard's matrix (Wolter, 1985).

With point 3), we mean that a specified statistic is calculated in one single sub-procedure; so, other complex statistics can be added to this software, changing the code of that sub-procedure only.

Appendix 1. The EU complex statistics

Introducing k as the index of the unit, y_k as the value of the variable Y on the unit k , w_k as the weight of the unit k , \hat{Y}_β as the estimated β th quantile of variable Y , we report briefly the EU complex statistics involved in this paper (EUROSTAT, 2004a-2004b; Moretti et al., 2004a-2004b):

1) at Risk-of-Poverty Threshold (RPT), defined as the 60% of the median national income¹⁰:

$$RPT=60\% \hat{Y}_{0,5} \quad (1)$$

2) at Risk-of-Poverty Rate (RPR), defined as the percentage of persons over the total population with an income below at Risk-of-Poverty Threshold:

$$RPR = \frac{\sum_{k \in S} I_k}{\sum_{k \in S} w_k} * 100 \quad (2)$$

where S is the collected sample, I_k is a binary variable defined as:

$$I_k = \begin{cases} 1 & \text{if } y_k < RPT \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

¹⁰ Income means equivalised disposable income.

3) inequality of income distribution, Gini index defined as (EUROSTAT, 2004b):

$$G=100* \left[\frac{2* \sum_{k=first\ unit}^{last\ unit} \left(y_k * w_k * \frac{\sum_{first\ unit}^{unit\ k} w_k}{\sum_{k=first\ unit}^{last\ unit} y_k * w_k} \right) - \sum_{k=first\ unit}^{last\ unit} y_k * w_k^2}{\left(\sum_{k=first\ unit}^{last\ unit} w_k \right) * \sum_{k=first\ unit}^{last\ unit} y_k * w_k} - 1 \right] \quad (4)$$

to apply formula (4) the values of y_k have to be in ascending order;

4) Gender Pay Gap (GPG), defined as the difference between men's and women's average gross hourly earnings as a percentage of men's average gross hourly earnings (the population consists of all paid employees aged 16-64 'at work x+ hours per week'):

$$GPG = \frac{\frac{\sum_{k \in M} y_k w_k}{\sum_{k \in M} w_k} - \frac{\sum_{k \in F} y_k w_k}{\sum_{k \in F} w_k}}{\frac{\sum_{k \in M} y_k w_k}{\sum_{k \in M} w_k}} * 100 \quad (5)$$

where M are the male sampled paid employees aged 16-64 'at work x+ hours per week', F are the female sampled paid employees aged 16-64 'at work x+ hours per week';

5) Relative Median at Risk-of-Poverty Gap (RPG), defined as the difference between the median income of poor units and the 'at Risk-of-Poverty Threshold', expressed as a percentage of the 'at Risk-of-Poverty Threshold':

$$RPG = \frac{RPT - \hat{Y}_{0.5}^{poor}}{RPT} * 100 \quad (6)$$

where RPT is the 'at Risk-of-Poverty Threshold' expressed by (1) and $\hat{Y}_{0.5}^{poor}$ is the median income calculated for poor units that means units with $I_k=1$ (see formula (3));

6) Income Quintile Share Ratio (QSR), defined as the ratio of total income received by 20% of the country's population with the highest income (top quintile) to that received by 20% of the country's population with the lowest income (lowest quintile):

$$QSR = \frac{\left(\sum_{k \in T} y_k w_k \right) / \left(\sum_{k \in T} w_k \right)}{\left(\sum_{k \in L} y_k w_k \right) / \left(\sum_{k \in L} w_k \right)} \quad (7)$$

where T is the set of sampled units with $y_k > \hat{Y}_{0.80}$ and L is the set of sampled units with $y_k \leq \hat{Y}_{0.20}$.

Appendix 2. The linearized variables

Introducing k as the index of the unit, y_k as the value of the variable Y on the unit k , w_k as the weight of the unit k , \hat{Y}_β as the estimated β th quantile of variable Y , N as the number of units in the population, the linearized variables (obtained by the ‘estimating equations’ approach and the Taylor–Woodruff method) used in this work are (Moretti et al., 2004a-2004b; Berger, Skinner, 2003; Binder, Kovacevic, 1994; Deville, 1999; Kovacevic, Binder, 1997; Preston, 1996; Woodruff, 1952):

1) for the ‘at Risk-of-Poverty Threshold’:

$$u_k = \frac{1}{N \hat{f}(\hat{Y}_{0.5})} \left[0.5 - I(y_k < \hat{Y}_{0.5}) \right] \quad (8)$$

where $\hat{f}(\hat{Y}_{0.5})$ is the estimated density function of variable Y in $\hat{Y}_{0.5}$ and $I(y_k < \hat{Y}_{0.5})$ is a binary variable with value 1 if $y_k < \hat{Y}_{0.5}$ and 0 otherwise.

2) for the ‘at Risk-of-Poverty Rate’:

$$u_k = \frac{1}{N} \left[I(y_k < 0.6 \hat{Y}_{0.5}) - \hat{p}_{0.6,0.5} - 0.6 \hat{R}_{0.6,0.5} (I(y_k < \hat{Y}_{0.5}) - 0.5) \right] \quad (9)$$

and

$$\hat{R}_{0.6,0.5} = \frac{\hat{f}(0.6 \hat{Y}_{0.5})}{\hat{f}(\hat{Y}_{0.5})} \quad (10)$$

where $\hat{p}_{0.6,0.5}$ is the estimated proportion of units with y_k below $0.6 \hat{Y}_{0.5}$, $I(y_k < 0.6 \hat{Y}_{0.5})$ is a binary variable with value 1 if $y_k < 0.6 \hat{Y}_{0.5}$ and 0 otherwise, $I(y_k < \hat{Y}_{0.5})$ is a binary

variable with value 1 if $y_k < \hat{Y}_{0.5}$ and 0 otherwise, $\hat{f}(0.6\hat{Y}_{0.5})$ and $\hat{f}(\hat{Y}_{0.5})$ are the estimated density functions of Y in $0.6\hat{Y}_{0.5}$ and in $\hat{Y}_{0.5}$.

3) for the ‘Gini index’:

$$u_k = \frac{2}{N\hat{\mu}} \left[\hat{A}(y_k)y_k + \hat{B}(y_k) - \frac{\hat{\mu}}{2}(\hat{G}+1) \right] \quad (11)$$

and

$\hat{A}(y) = \hat{F}(y) - (\hat{G}+1)/2$, $\hat{B}(y) = \sum_s w_k y_k I(y_k \geq y)/N$, $\hat{G} = \frac{1}{N\hat{\mu}} \sum_s w_k (2\hat{F}_k - 1)y_k$ is the estimated index of Gini, $\hat{F}_k = \sum w_k I(y_j \leq y_k)/N$ and $\hat{\mu}$ is the estimated mean of Y.

4) for the ‘Gender Pay Gap’:

$$u_k = b_1 y_{F_k} + b_2 y_{M_k} + b_3 d_{M_k} + b_4 d_{F_k} \quad (12)$$

where

$$b_1 = \frac{\sum_{k \in M} w_k}{\sum_{k \in M} y_k w_k + \sum_{k \in F} w_k}, \quad b_2 = -\frac{\sum_{k \in F} y_k w_k + \sum_{k \in M} w_k}{\left(\sum_{k \in M} y_k w_k \right)^2 + \sum_{k \in F} w_k}, \quad b_3 = \frac{\sum_{k \in F} y_k w_k}{\sum_{k \in M} y_k w_k + \sum_{k \in F} w_k},$$

$$b_4 = -\frac{\sum_{k \in F} y_k w_k + \sum_{k \in M} w_k}{\sum_{k \in M} y_k w_k + \left(\sum_{k \in F} w_k \right)^2},$$

M are the male sampled paid employees aged 16-64 ‘at work x+ hours per week’; F are the female sampled paid employees aged 16-64 ‘at work x+ hours per week’; y_{M_k} = the value of the variable of interest if unit k is a male sampled paid employees aged 16-64 ‘at work x+ hours per week’, $y_{M_k} = 0$ otherwise; y_{F_k} = the value of the variable of interest if unit k is a female sampled paid employees aged 16-64 ‘at work x+ hours per week’, $y_{F_k} = 0$ otherwise; $d_{M_k} = 1$ if unit k is a male sampled paid employees aged 16-64 ‘at work x+ hours per week’, $d_{M_k} = 0$ otherwise; $d_{F_k} = 1$ if unit k is a female sampled paid employees aged 16-64 ‘at work x+ hours per week’, $d_{F_k} = 0$ otherwise.

Appendix 3. The Balanced Repeated Replications technique

We describe the basic technique for one stage stratified sampling design; the extension to the multistage design will be showed (ISTAT, 1989; McCarthy, 1969a-1969b; Rao, Wu, 1988; Rao, Shao, 1996; Särndal et al., 1992). Let a finite population divided in H strata and a sample of two units drawn in each stratum. Selecting at random a sample unit per stratum, we get a sub-sample called *replication* whose size is the half original sample size. Let a_h be a coefficient with value +1, if first unit in stratum h belongs to the *replication*, -1 if the second one belongs to the *replication*; then, the set of all replications can be described by a matrix $2^H \times H$. Each cell of this matrix has the value +1 or -1. In this matrix, the following properties hold:

$$\sum_{r=1}^{2^H} a_{rh} = 0 \quad (h=1, \dots, H) \quad (13)$$

$$\sum_{r=1}^{2^H} a_{rh} a_{rk} = 0 \quad (h=1, \dots, H) \quad (14)$$

The (13) implies that each sample unit has to be present in the same number of replications, the (14) that matrix columns are each other orthogonal. Let $\hat{\theta}$ be a linear estimation of a population parameter computed on the whole sample and let $\hat{\theta}_r$ the estimation computed on the replication r, it can be showed that:

$$\hat{\theta} = \frac{1}{2^H} \sum_{r=1}^{2^H} \hat{\theta}_r \quad (15)$$

$$Var(\hat{\theta}) = \frac{1}{2^H} \sum_{r=1}^{2^H} (\hat{\theta} - \hat{\theta}_r)^2 \quad (16)$$

In the case of non linear estimator the (15) doesn't hold and the (16) is an estimate of variance.

If the method is applied to the whole set of replications the computation cost is high also for a small H (if H=30, replications amount to more than one billion). To decrease the number of replications, it is necessary to look for a $R \ll 2^H$; selecting at random R replications implies a variance estimation greater than the estimation computed on the whole set of replications. The Balanced Repeated Replications technique has been given by McCarthy (McCarthy, 1969a-1969b) as a solution to this problem. According to BRR technique a subset of R replication is balanced if:

$$\sum_{r=1}^R a_{rh} a_{rk} = 0 \quad (h, k=1, \dots, H; k \neq h) \quad (17)$$

Moreover if:

$$\sum_{r=1}^R a_{rh} = 0 \quad (h=1, \dots, H) \quad (18)$$

the replications are fully balanced. Variance estimation on R balanced replications give the same estimation of the 2^H replications. To get a subset of balanced replications Hadamard's matrices must be used. Hadamard's matrices are special squared matrices to describe k replications. For each subset of $k' < k$ columns (excluding first column) conditions (17) and (18) hold; for all the columns only condition (17) holds. The Hadamard's matrix of order 2 is:

$$\begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}$$

Let M a Hadamard's matrix of order k; the iterative procedure to generate a 2k order matrix is:

$$\begin{pmatrix} \mathbf{M} & \mathbf{M} \\ \mathbf{M} & -\mathbf{M} \end{pmatrix}$$

It is possible to compute two variance estimation formulas:

$$Var_1(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta} - \hat{\theta}_r)^2 \quad (19)$$

and

$$Var_2(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta} - \hat{\theta}_r^c)^2 \quad (20)$$

where $\hat{\theta}_r^c$ is the estimation of the parameter computed on the sample complement to replication r (that is the replication got multiplying for -1 the matrix coefficients). The average of (19) and (20) gives a more precise estimation of the sampling variance. The Italian sample surveys on households are based on two-stage sampling designs with stratification of the primary sampling units (PSU), that are the municipalities; in order to apply BRR technique some modifications are required. The modifications can be summarized in four steps:

- 1) sampling units are considered at first stage level;

- 2) if a stratum contains one sampling unit only, it will be collapsed to neighbor stratum; this operation implies an overestimation of the variance;
- 3) if a stratum contains more than two sampling units, they will be regrouped in two pseudo-sampling units at random;
- 4) in self-representative strata, every sampling household is considered a PSU and they are divided in two pseudo-PSU as described in step 3.

At the end of these four steps the basic BRR can be applied. Lastly, the estimation was obtained as simple average of (19) and (20).

REFERENCES

Berger, Y.G. and Skinner, C.J. (2003). Variance estimation for a low income proportion. *Applied Statistics*, 52, part 4, pp. 457-468.

Binder, D.A. and Kovacevic, M.S. (1994). Estimating some measures of income inequality from survey data: an application of the estimating equation approach. *Survey Methodology*, 21, pp. 137-145.

Coccia, G., Pannuzi, N., Rinaldelli, C. and Vignani, D. (2002). Verso una misura della povertà regionale: problemi e strategie. Sesta Conferenza Nazionale di Statistica, Roma 6-8 Novembre 2002, on www.istat.it.

Deville, J.C. (1999). Variance Estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, vol. 25 n.2, pp 193-203.

De Vitiis, C., Di Consiglio, L., Falorsi, S., Pauselli, C. and Rinaldelli, C. (2003). La valutazione dell'errore di campionamento delle stime di povertà relativa. Document related at Conference "Povertà Regionale ed Esclusione Sociale", Roma 17 Dicembre 2003.

EUROSTAT (2004a). Technical document on intermediate and final quality reports, *Statistics on Income and Living Conditions (EU-SILC)*. Doc. EU-SILC 132/04.

EUROSTAT (2004b). Common cross-sectional EU indicators based on EU-SILC; the gender pay gap, *Statistics on Income and Living Conditions (EU-SILC)*. Doc. EU-SILC 131/04.

Falorsi, S. and Rinaldelli, C. (1998). Un software generalizzato per il calcolo delle stime e degli errori di campionamento. *Statistica Applicata*, 10, 2, 217-234.

ISTAT (1989). Tecniche di stima della varianza campionaria. *Manuale di tecniche di indagine, Collana Note e Relazioni*, anno 1989, n.1, vol.5.

ISTAT (2002a). GENESEES v1.0 Software per il calcolo delle stime e degli errori campionari. Manuale Utente e Aspetti Metodologici.

ISTAT (2002b). La stima ufficiale della povertà in Italia 1997-2000. Argomenti n.24.

ISTAT (2002c). Panel Europeo sulle famiglie. Collana ISTAT Metodi e Norme n.15.

ISTAT (2003). La povertà e l'esclusione sociale nelle regioni italiane. Statistiche in Breve, 17 dicembre 2003, on www.istat.it.

ISTAT (2004). I consumi delle famiglie, anno 2002. Annuario n.9.

Kovacevic, M.S. and Binder, D.A. (1997). Variance estimation for measures of income inequality and polarization – The estimating equations approach. *Journal Official Statistics*, Vol. 13 N. 1, pp. 41-58.

McCarthy, P.J. (1969a). Pseudoreplication: further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics, Series 2 n.31*, National Center for Health Statistics, Public Health Service, Washington, D.C.

McCarthy, P.J. (1969b). Pseudoreplication: Half-Samples. *Review of the International Statistical Institute*, 37, 239-264.

Moretti, D., Pauselli, C. and Rinaldelli, C. (2004a). La valutazione dell'errore campionario degli indicatori complessi di povertà e disuguaglianza dell'indagine su reddito e condizioni di vita (EU-SILC). Internal document, ISTAT, June 2004.

Moretti D., Pauselli, C. and Rinaldelli, C. (2004b). La stima del “fattore disegno di campionamento” degli indicatori complessi di povertà e disuguaglianza dell'indagine su reddito e condizioni di vita (EU-SILC). Internal document, ISTAT, June 2004.

Pauselli, C. and Rinaldelli, C. (2004a). Stime di povertà relativa: la valutazione dell'errore campionario secondo le Replicazioni Bilanciate Ripetute. *Statistica Applicata*, Vol.16, n.1, 1-13.

Pauselli, C. and Rinaldelli, C. (2004b). La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute. *Rivista di Statistica Ufficiale* (accepted, to be printed).

Pauselli, C. and Rinaldelli, C. (2004c). La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute. *Collana ISTAT Contributi*, n. 9, on www.istat.it.

Preston, I. (1996). Sampling distributions of relative poverty statistics. *Applied Statistics*, 45, pp. 91-99.

Rao, J.N.K. and Wu, C.F.J. (1988). Resampling inference with complex survey data. *American Statistical Association*, 83, 401, 231-241.

Rao, N.K. and Shao, J. (1996). On balanced half-sample variance estimation in stratified random sampling. *Journal of American Statistical Association*, vol. 91, n.433.

REGULATION (EC) No 1177/2003 of the EUROPEAN PARLIAMENT and of the COUNCIL of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC), 3.7.2003 L 165/1 *Official Journal of the European Union*.

Särndal, C-E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Skinner, C. J., Holt, D. and Smith, T.M.F. (editors) (1989). *Analysis of complex surveys*. Wiley and sons.

Woodruff, R. S. (1952). Confidence intervals for medians and other position measures. *J. Am. Statist. Ass.*, 47, 635-646.

Wolter, K. M. (1985). *Introduction to variance estimation*. New York: Springer-Verlag.