Evaluating reliability of combined responses through latent class models

Marcello D'Orazio[1]

Abstract

The evaluation of the potential impact of the response errors on the final survey estimates requires ad hoc studies. Often these studies consist in additional reinterview surveys: a subsample of the respondent units at the main survey is interviewed again. In such cases, the evaluation can be done by means of the theory introduced by Hansen et al. (1964) and further investigated in Biemer and Forsman (1992). More recent studies (Biemer, 2004) present an approach based on the fitting of latent class models. These models allow for a more detailed analysis of the impact of response errors on the final survey estimates but, on the other hand, they require some additional assumptions to hold. In this paper, the usage of latent class models is extended to tackle the case of couples of survey questions involved in a questionnaire skip. An application of such models to the data of the control survey on the 2001 Population and Housing Census in presented.

Keywords: Response Errors, Simple Response Variance, Latent Class Models, Questionnaire Skip

1. Introduction

A measurement error consists in the differences between the value observed for a variable and the true value of this variable for the investigated unit. The measurement errors arising in the survey data collection are called *response errors*. When dealing with categorical variables these errors are also known as *misclassification errors*. In order to deal with the response errors, Hansen *et al.* (1964) introduced a simple model widely used at the US Bureau of the Census (1985). They showed that the measurement errors represent a source of additional variability and may introduce a bias (*response bias*) when estimating the population characteristics. The additional variability can be decomposed into the *simple response variance* (SRV) and the *correlated response variance* (CRV). The first term reflects the variability of the responses that can be collected for a single question in a series of repetitions of the data collection on the same unit. The second term reflects the correlation of the responses collected for the same question on different units in a given survey occasion. A well known source of CRV is represented by the interviewers. Usually, in a self-administered interview the CRV is assumed to be zero.

A common practice for estimating the bias and the variance due to response errors consists in carrying out a reinterview study. A reinterview survey with the purpose of obtaining error-free responses (*gold standard*) permits to estimate the response bias. On the

[1] Researcher (Istat), e-mail: madorazi@istat.it.

other hand, an independent survey consisting in a perfect replication of the main survey on a subset of the responding units (*test-retest reinterview*) has to be carried out to estimate of the SRV (for details see Biemer and Lyberg, 2003, pp. 291-301). In most of the cases, a reinterview survey consists in a combination of both (cf. Biemer and Forman, 1992).

In this work the focus is on the usage of the reinterview survey data in order to estimate the SRV for categorical variables. The paper is structured as follows: Section 2 introduces the approach used at the US Bureau of the Census in order to evaluate the impact of SRV on the final survey estimates. Some results obtained using this approach with the data of the quality control survey of the 2001 Population and Housing census in Italy are reported in Section 2.1. In Section 3 it is introduced the problem of couples of questions involved in a questionnaire skip. Section 4 provides a summary of the theory underlying the LC models for the analysis of repeated measurements. In Section 4.1 it is presented a relatively new approach, based on the application of Latent Class (LC) models, with the objective of evaluating the reliability of the couples of variables involved in a questionnaire skip. Section 4.2 reports the results obtained applying this new approach to the data of the quality control survey of the 2001 Census.

2. Evaluating reliability: the US Bureau approach

In the approach of the US Census Bureau (Hansen *et al.* 1964; US Census Bureau 1985; Biemer and Forsman, 1992; Biemer, 2004) a key role is played by the *index of inconsistency*:

$$I = \frac{\text{SRV}}{\text{SV} + \text{SRV}} \tag{2.1}$$

It represents the proportion of total variance (SV is the sampling variance) due to response errors (cf. Biemer, 2004), hence $0 \le I \le 1$. It is worth noting that the quantity $R = 1 - I$ is known as *reliability ratio*. Once estimated *I*, the following rule of thumb can be considered (US Census Bureau, 1985, p. 95):

$0 \le \hat{I} \le 0.20$	high reliability (low inconsistency)
$0.20 < \hat{I} < 0.50$	moderate reliability (moderate inconsistency)
$0.5 \le \hat{I} \le 1$	low reliability (high inconsistency)

In order to estimate *I*, let consider a categorical variable Y with J ($J \ge 2$) response categories and let assume $y_k^{(1)}$ to be the value observed for Y on the *k*th unit in the main survey ($t = 1$) and $y_k^{(2)}$ the value observed on the same unit in the reinterview survey ($t = 2$). Usually, the reinterview is carried out on a subsample of the respondents to the main survey, selected according to a given probabilistic sampling design $p(s^{(2)})$, being $w_k^{(2)}$ the survey weight of the *k*th unit (inverse of inclusion probability, maybe corrected for unit nonresponse in reinterview survey). In this framework I is estimated by (US Bureau of the Census, 1985, p. 88):

$$\hat{I} = \frac{g}{1-\sum_{j=1}^{J}\hat{P}_{+j}\hat{P}_{j+}} \tag{2.2}$$

with

$$g = 1-\sum_{j=1}^{J}\hat{P}_{jj} \tag{2.3}$$

Note that $\hat{P}_{ij} = \hat{N}_{ij}/\hat{N}$ ($i, j = 1,\dots,J$) are the relative frequencies of the cells of the contingency table $Y^{(2)} \times Y^{(1)}$. In particular, $\hat{N} = \sum_{s^{(2)}} w_k^{(2)}$ and $\hat{N}_{ij} = \sum_{s^{(2)}} w_k^{(2)} C\left(y_k^{(2)} = i, y_k^{(1)} = j\right)$, being $C(\cdot) = 1$ if the condition within parenthesis is satisfied and 0 otherwise. Note that these frequencies are estimated before the editing and imputation phase and all units with missing values (at the main survey, at the reinterview survey or at both) are discarded from the computation.

The quantity g is known as the *gross difference rate* (GDR) or *disagreement rate*. It can be shown that $g/2$ provides an unbiased estimate of the SRV under the assumption of: (A1) equal probabilities of misclassification of the responses at original interview and at reinterview and, (A2) independence between the responses at the main survey and those at the reinterview (absence of *between-trial correlation*) (for details see Biemer and Forsman, 1992; Biemer 2004). They are assumed to hold when the reinterview survey is a perfect independent replication of the main survey.

When the assumption (A1) does not hold, then $g/2$ does not provide an unbiased estimate of the SRV at $t = 1$ and, consequently, the estimated I is unreliable. When both the assumptions do not hold, the negative bias introduced by the presence of positive between-trial correlation (failure of A2) tends to determine an underestimation of the SVR. On the contrary, when only (A2) holds, the sign of the bias is determined by the relationship between misclassification probabilities at the two survey occasions. If misclassification probabilities at the control survey ($t = 2$) are lower than those at the original survey ($t = 1$), then an underestimation of SRV at $t = 1$ it is expected. Overestimation occurs in the opposite situation (cf. Biemer and Forsman, 1992, pp. 919-920).

Note that in case of binary variables ($J = 2$), the assumption (A1) (equal error distribution) corresponds to assuming $H_0 : P^{(t=1)} = P^{(t=2)}$ (cf. Biemer, 2004, p. 430). In the more general case ($J > 2$) it corresponds to assuming the *marginal homogeneity* (cf. D'Orazio, 2008).

Finally, it is worth noting that $\hat{I} = 1-\kappa$, being κ the Cohen's *kappa* measure of reliability (cf. Biemer, 2004, p. 423).

2.1 Reliability in 2001 Population and Housing Census in Italy

The quality evaluation program of the 2001 Population and Households Census (CEN) consisted in a single control survey, referred as Post Enumeration Survey (PES), aimed at estimating the impact on Census estimates of both coverage and measurement errors.

The PES was based on a stratified two stage sample of approximately 1,100 census Enumeration Areas (EAs), located into 98 sample municipalities (the Primary Sampling Units). In each sample area a new complete enumeration of the households was carried out. At the end of the data collection, approximately 68,000 households and about 180,000 individuals were surveyed. A complex procedure of record linkage allowed the units covered by both the CEN and the PES to be identified. In particular, in order to evaluate the impact of response errors, $n = 172{,}620$ linked individuals were considered, out of the 182,519 people found in the sample EAs at the CEN. The linkage procedure was very successful given that linkage rate was quite high (CEN linkage rate was 95%) and some consistency checks led to the conclusion that the false links (couples of units erroneously linked) had a very low chance of occurring. This is an important result because the presence of false links may negatively affect the estimation of the response variance as shown in Brancato *et al.* (2004).

In order to compare the responses provided by the same individuals at the two survey occasions (CEN and PES) the questionnaire of the PES was enlarged with a subset of questions (about fifteen) selected from those in the census form. The PES data were collected by means of a self-administered paper questionnaire - as for the CEN - in the period November-December 2001, about a month after the CEN data collection (CEN reference date is 21st of October 2001). The overlapping with the CEN field operations was accurately avoided. In practice, the PES data collection was designed as a perfect replication of the census one (test-retest reinterview) in order to fulfill assumption (A1). Given the time lag between the two surveys, the between-trial correlation, due to the situation of respondents that at the PES recall the answers provided at the CEN and repeat them, can be considered negligible (assumption A2 holds). Biemer and Lyberg (2003, pp. 298-299) report 5-10 days to be a sufficient time lag for carrying out an independent reinterview.

Table 1 reports the GDRs and the estimated values for I for some of the most important variables (cf. Istat, 2009, pp. 121-122).

Table 1 - GDR and estimates of *I* for some of the variables.

Census Questions	Number of response categories	GDR (x 100)	$\hat{I}$ (x 100)
Relationship with the household head	16	5.25	7.40
Gender	2	0.81	1.62
Age *(in classes)*	16	2.19	2.34
Marital status	6	1.80	3.04
Education level *(Age > 6)*	16	12.72	15.71
Labour status *(Age > 14)*	10	12.09	16.24
Full-time/part-time occupation (*for who declared to have an occupation*)	2	3.56	20.97
Professional status (*for who declared to have an occupation*)	6	7.79	18.33
Limited/unlimited duration of contract (*for occupied as employees*)	2	5.76	28.95

Reliability is quite high when dealing with the first four variables. As expected, Gender is the most reliable variable. The estimates of I increase starting from the question concerning the education level. According to previously mentioned rule of thumb, reliability is always "high" ($\hat{I} \leq 0.20$) with the exception of the question "full time/part time" occupation and for the one concerning the duration of the contract (for those who declared to be occupied as employees).

3. Evaluating the reliability for questions involved in a questionnaire skip

The census form had several *filter questions*. A filter question is crucial because the answer to it determines the response path, i.e. which section of the questionnaire has to be filled in and which one has to be skipped. For instance, the question related to the professional status had to be filled in only by those individuals who responded to be occupied at the question concerning the labour status. A complex questionnaire can contain several filter questions; hence, assessing the reliability of these key questions is crucial in order to assess the overall reliability. On the other hand, evaluating the reliability of the questions which depends on the responses to the filter questions may give rise to some difficulties. For simplicity, let consider the case of two binary variables, X and Y, such that: (i) an answer to Y (i.e. $y = 1$ or $y = 2$) is due if and only if $x = 1$ and, (ii) Y should be skipped (i.e. $y = \text{"Not Applicable"}$) if $x = 2$. In the ideal situation of absence of errors in the skip from one question to another one (e.g. when the data collection is assisted by PC and the questionnaire skips are managed by the software) the following combinations of responses are admitted:

X	Y
1	1
1	2
2	NA

Unfortunately, when dealing with self-administered paper questionnaire, the errors in the questionnaire skip have to be taken into account. In the previous example, when evaluating the reliability of the Y variable there are two possibilities: (a) consider all the cases without caring of the values of the X variable and, (b) limit the attention to the subset of the cases with $x^{(2)} = x^{(1)} = 1$. In the case (a) the disagreements between responses to the filter question (X) are implicitly considered when evaluating the disagreement for Y variable. On the other hand, in the case (b), by considering only the disagreements on Y conditioning to an agreement on the X variable, the response errors in the X are discarded. In this latter case the GDR is computed using:

$$g_{Y|X1=X2=j} = \frac{\sum_{k=1}^{n} w_k^{(2)} C(y_k^{(1)} \neq y_k^{(2)} \big| x_k^{(1)} = x_k^{(2)} = j)}{\sum_{k=1}^{n} w_k^{(2)} C(x_k^{(1)} = x_k^{(2)} = j)} \tag{3.1}$$

and can be referred as a *"net" disagreement rate* on the *Y* variable. In practical situations, the two options (a) and (b) lead to similar GDR values, unless there are many errors in the questionnaire skips (in absence of such errors they provide the same estimate).

In the PES survey, when computing the GDRs (see Table 1) for the variables depending on a filter question, the formula (3.1) was used. In any case, by comparing the GDRs under option (a) (i.e. all the units are considered) and (b) the "net" case, similar values come out as shown in Table 2, thus denoting the presence of few errors in the skip among the questionnaire sections.

Table 2 - GDR and estimates of *I* for some variables in the Census form involved in questionnaire skips.

Variables	Number of categories	GDR (x100) All units	GDR (x100) "net"	$\hat{I}$ (x 100) All units	$\hat{I}$ (x 100) "net"
Education level *(Age > 6)*	16	12.79	12.72	15.79	15.71
Labour status *(Age > 14)*	10	12.01	12.09	16.01	16.24
Full-time/part-time occupation (*for who declared to have an occupation*)	2	3.69	3.56	21.03	20.97
Professional status (*for who declared to have an occupation*)	6	8.00	7.79	18.65	18.33
Limited/unlimited duration of contract (*for occupied as employees*)	2	6.04	5.76	28.77	28.95

A more detailed investigation on the relationship between the GDRs of two variables, *X* and *Y*, involved in a questionnaire skip can be found in D'Orazio (2008).

4. Latent Class Models in presence of repeated measurements

Biemer (2004) shows how to use LC models to investigate the reliability of responses when dealing with a categorical variable. These models allow the assumption of equal probabilities of misclassification in the two survey occasions to be relaxed but, on the other hand, their application relies on the following assumptions:

(B1) the misclassification probabilities at both survey occasions do not vary among individuals:

$$\phi_{kij}^{(t)} = \Pr\left(y_k^{(t)} = i \middle| y_k = j\right) = \Pr\left(y^{(t)} = i \middle| y = j\right) = \phi_{ij}^{(t)},$$

for $t = 1,2$, $k = 1,\dots,n$, $i, j = 1,\dots,J$, being y_k the true unobserved (*latent*) *Y* classification of the *k*th unit;

(B2) the *local independence* holds:

$$\begin{aligned}\Pr\left(y^{(1)} = i,\ y^{(2)} = j,\ y = h\right) = \\ = \Pr\left(y^{(1)} = i \middle| y = h\right) \times \Pr\left(y^{(2)} = j \middle| y = h\right) \times \Pr\left(y = h\right); \quad i, j, h = 1,\dots,J\end{aligned}$$

In the traditional approach to LC models, if the model is identifiable, i.e. the maximum value of the likelihood exists and is unique (for details see Goodman, 1974; Huang, 2005) the ML estimates of the cell probabilities involved in (B2) are derived using well known iterative algorithms such as Newton-Raphson or EM. Then, these estimates can be combined to get an estimate of SRV and I in correspondence of each survey occasion (cf. Biemer, 2004, p. 427). As far as identifiability of the model is concerned, it is worth noting that the factorization in (B2) involves $(J-1)\times(2\times J+1)$ parameters but, given that the observed table of $Y^{(2)}\times Y^{(1)}$ has only $(J\times J-1)$ degrees of freedom, the model is not identifiable (the number of parameters is greater than the degrees of freedom), unless a new additional auxiliary variable G is introduced, with the constraint that the error probabilities $\phi_{ij}^{(t)}$ are equal across the groups identified by the categories of G (cf. Biemer, 2004, pp. 425).

Given that the application of LC models does not require the assumption (A1) of equal misclassification probabilities at both survey occasions, they can be applied to the general situation of reinterview studies carried out in different conditions with respect to the main survey. Moreover, fitting a LC model provides an estimate of the true unobserved $\Pr(y=h)$, hence it is possible to derive an estimate of the response bias, without having to carry out a reinterview study aimed at ascertaining the gold standard. For instance, an estimate of the response bias at the main survey ($t=1$) is:

$$\widetilde{B}_P^{(t=1)}=\hat{\Pr}\left(y^{(1)}=h\right)-\hat{\Pr}_{\mathrm{LC}}(y=h)=\frac{\hat{N}_h^{(t=1)}}{\hat{N}}-\hat{\Pr}_{\mathrm{LC}}(y=h). \tag{4.1}$$

Unfortunately, the usage of a LC model poses some problems. The crucial assumption of the local independence (B2) may not hold; in this case it has to be relaxed by resorting to one of the approaches proposed in literature (see e.g. Hagenaars, 1988; Vermunt, 1997). Assumption (B1) can also be relaxed by introducing the dependence of the distributions of both latent and observed variables on a set of individual covariates (see e.g. Huang *et al.*, 2004). In both cases the risk is that of increasing too much the complexity of the model, thus affecting its identifiability. Another known problem when fitting LC models is represented by local maxima: the iterative algorithm may stop at local maximum and therefore the resulting estimates for the parameters would not be the ML ones. Usually, the chance of having local maxima increases with increasing number of categories of the latent variable. Finally, the ordering of the estimated latent classes is arbitrary and may become difficult to identify the effective response category associated to each latent class (likely to happen with a high number of latent classes).

It is worth noting that in the traditional LC framework, the ML estimates of the parameters are derived by assuming i.i.d. observations, therefore when the reinterview study is based on a complex sample (with stratification and clustering) the ML estimates can be considered valid under the further assumption that the sampling design plays no role in the inference (*model-based inference*, cf. Särndal *et al.*, 1992, pp. 513-520). If the sampling design can not be ignored, it has to be taken into account jointly with the sampling weights. This can be done by resorting to one of the approaches available in literature such as the *Pseudo-ML* estimation (cf. Patterson *et al.* 2002) or to the two-step approach suggested by Vermunt (2002). Vermunt and Magidson (2007) compared these and other approaches to fit LC models when dealing with complex samples.

4.1 Latent class models to evaluate reliability for coupled variables involved in questionnaire skips

Let consider the simple case of two binary variables, X and Y, such that: (i) an answer to Y (i.e. $y=1$ or $y=2$) is due if and only if $x=1$ and, (ii) Y should be skipped (i.e. $y=\text{"Not Applicable"}$) if $x=2$. In the census questionnaire there were a number of situations situation that could be summarised in such a simple situation. Two examples are reported in Tables 3a and 3b.

Table 3a - Relationship between the labour status and the type of occupation.

X="Labour Status"	*Y="Type of occupation"*
1 = "With occupation"	1 = "Full time"
1 = "With occupation"	2 = "Part time"
2 = "Without occupation"	NA = "Not Applicable"

Table 3b - Relationship between the professional status and the duration of contract (for those with an occupation).

X="Professional Status"	*Y="Duration of contract"*
1 =""Employee"	1 = "Unlimited duration"
1 = "Employee"	2 = "Limited duration"
2 = NOT "employee" (self-employed, family worker,..)	NA = "Not Applicable"

Let Z be the variable obtained by combining the response categories of Y and X. Following (i) and (ii), only three categories are admitted for Z, as can be seen in Table 4 (last column).

Table 4 - Response categories obtained by combining *X* and *Y*.

$X^{(t)}$	$Y^{(t)}$	$Z^{(t)}$	Z
1	1	1	1
1	2	2	2
1	"NA"	4	-
2	1	5	-
2	2	6	-
2	"NA"	3	3

If response errors in the skip pattern are considered, then the variable $Z^{(t)}$, obtained by combining responses to $X^{(t)}$ and to $Y^{(t)}$, can admit up to 6 categories (third column in the Table 4). Therefore, the contingency table $Z^{(2)} \times Z^{(1)}$, built before the editing and imputation phase, may show more than the expected 9 ($=3\times3$) cells. This is the case, for instance, of the variables in Table 3a observed at the CEN ($t=1$) and at the PES ($t=2$).

Table 5 - Occupation and type of occupation (estimated relative frequencies x 100).

Reinterview		Census						
		With occupation = "Yes"			With occupation = "No"			
With occupation	Full/part time	1	2	NA	1	2	NA	Total
1 = "Yes"	1= "Full-time"	33.465	0.925	1.293	0.103	0.034	1.025	36.844
	2= "Part-time"	0.429	2.810	0.104	0.008	0.081	0.287	3.720
	NA	0.249	0.030	0.027	0.002	0.002	0.043	0.352
2 = "No"	1= "Full-time"	0.048	0.003	0.004	0.010	0.002	0.113	0.182
	2= "Part-time"	0.006	0.013	0.001	0.001	0.017	0.037	0.076
	NA	0.794	0.334	0.108	0.121	0.231	57.238	58.827
	Total	34.991	4.116	1.537	0.245	0.367	58.744	100.000

Note: NA corresponds to "Not Applicable" or "Not Answer"

Table 5, due to response errors and to errors in the skip pattern, has 36 non-empty cells instead of the expected 9. In any case, the frequencies estimated for the cells relating erroneous skips are very close to 0.

This undesired situation may turn out useful in fitting LC models. In fact let consider the variable *Z*, and assume that (B1) and (B2) hold for *Z* too. Then, the factorization

$$\begin{aligned}\Pr\left(z^{(1)} = i,\, z^{(2)} = j,\, z = h\right) &= \\ &= \Pr\left(z^{(1)} = i \middle| z = h\right) \times \Pr\left(z^{(2)} = j \middle| z = h\right) \times \Pr\left(z = h\right)\end{aligned} \tag{4.2}$$

($i, j = 1,2,\ldots,6$, and $h = 1,2,3$) involves 32 ($= 5\times 3 + 5\times 3 + 2$) free parameters (probabilities) while the starting contingency table (e.g. Table 5) has 35 ($= 6\times 6 - 1$) degrees of freedom. In practice, the degrees of freedom are enough to make the model identifiable without having to introduce a grouping variable as in the case fitting a LC to *X* or to *Y*. This condition is necessary but not sufficient to ensure identifiability; to this purpose it is required that the model admits unique ML estimates of the unknown parameters too.

A further advantage in fitting a LC model to $Z^{(t=2)} \times Z^{(t=1)}$ consists in using the estimates of the probabilities in (4.2), to derive the estimates of *P*, the true population proportion of a given characteristics, and I for both the variables, *X* and *Y*, involved in the questionnaire skip. In other words, a single LC model is fitted instead of two separate models. It is worth noting that similar results can be obtained by fitting *path model* with latent variables to the repeated observations for the starting variables *X* and *Y*.

4.2 An application of the LC model to couples of variables from the census control survey

The theory presented in the previous Section has been applied to the data from the 2001 census control survey. In particular, repeated measurements for the couples of variables in the Tables 3a and 3b have been considered. Note that, due to the complex sampling design underlying the PES survey, the LC models has been applied starting from the contingency table $Z^{(t=2)} \times Z^{(t=1)}$ estimated by summing up the sample weights (scaled in order to sum up to the total sample size) of the cases falling in each cell of the observed table. This approach corresponds to *pseudo-Maximum Likelihood estimation* (cf. Patterson *et al.*, 2002; Vermunt and Magidson, 2007). Note that the identifiability of the model has been verified empirically by checking that the same ML estimates for the parameters are obtained by running the iterative estimation procedure with different starting values.

Table 6 shows the estimates obtained for *P* and I .

Table 6 - Results obtained applying LC models to the data of examples 3a and 3b.

	P estimates (x 100)					*I* estimates (x 100)		
Variables	CEN $t = 1$	PES $t = 2$	NDR	LCM	Bias at CEN	Standard formula	LCM CEN	LCM PES
X=1 ("With occupation")	41.99	42.34	-0.35	42.95	-0.96	5.81	6.85	4.76
Y=1 ("Full-time")	89.92	91.34	-1.42	88.19	1.73	20.97	13.82	27.18
X=1 ("Employee")	73.98	75.00	-1.02	74.58	-0.60	8.34	8.97	7.63
Y=1 ("Unlimited contract")	87.68	89.95	-2.27	85.01	2.67	28.95	20.28	36.20

As far as *P* estimates are concerned, it is interesting to observe that the *net difference rate* ($\mathrm{NDR} = \hat{P}_j^{(t=1)} - \hat{P}_j^{(t=2)}$,) is always negative; it is larger for the variables denoted as *Y*. Due to response errors, a negative response bias comes out when estimating $\Pr(x=1)$. The estimated bias, on the contrary, is positive when $\Pr(y=1)$ is concerned. The bias for *Y* variables is larger, in absolute terms, than that associated to *X*.

Finally, all the estimates of *I* concerning the *X* variables tend to be close, while marked differences emerge for the *Y* variables. The estimates of *I* obtained using the LC models, suggest a higher reliability of *Y* at CEN. This result is in contrast with the evidence from various studies that found lower misclassification probabilities in control surveys than in the main survey (cf. Biemer and Forsman, 1992, p. 920). As far as *Y* variables are considered, the discrepancies found in the estimates of *I*, jointly with the differences found for the NDR and the estimated response bias, lead to conclude that the assumption of equal misclassification probabilities (B1) may not hold in this case and, as a consequence, the estimates of SRV and *I* obtained with standard methods (formulas (2.2) and (2.3)) are misleading.

5. Conclusions

The study of methodologies involving the usage of LC models is important in order to evaluate the reliability when dealing with complex questionnaires with many skip patterns. The approach based on the LC models has the advantage of considering all the available data, including also not coherent answers according to the skip pattern in the questionnaire. Moreover, the estimates of the response bias, SRV and *I*, for each of the variables involved in a questionnaire skip are derived just by applying a single model to a new variable obtained by combining the answers to couples of questions. In addition, there is no need to resort to an additional grouping variable to make the model identifiable. Finally, LC models, providing an estimate of the true probabilities of an event, permit the estimation of the response bias, usually not allowed in the classic approach based on "test-retest" reinterview.

On the other hand, the usage of LC models presents some well known drawbacks. A crucial assumption is the local independence; it can be relaxed at the cost of increasing the complexity of the model. The same happens as far as homogeneity (assumption B1) is concerned. Moreover, when dealing with complex survey data, different approaches are available. The pseudo-ML provides good estimates of the parameters but it does not permit to evaluate the goodness of fit of the model using the standard tools (cf. Vermunt and Magidson, 2007). All these problems require further investigation in order to widely apply the LC models to evaluate responses reliability in complex surveys that collect data using questionnaires with many filter questions.

References

Biemer P.P. (2004), "The Twelfth Morris Hansen Lecture. Simple response Variance: Then and Now", *Journal of Official Statistics*, 3, pp. 417-439.

Biemer P.P., Forsman G. (1992), "On the quality of reinterview data with application to the current population survey", *Journal of the American Statistical Association*, 87, pp. 915-923.

Biemer P.P, Lyberg L.E. (2003), *Introduction to Survey Quality*, Wiley, New York.

Brancato G., D'Orazio M., Fortini M. (2004), "Response Error Estimation in Presence of Record Linkage Errors: The Case of the Italian Population Census". *Proceeding of the European Conference on Quality and Methodology in Official Statistics* (Q2004), Mainz, Germany, 24-26 May 2004.

D'Orazio M. (2008), "Reliability of responses at the 14th Italian Population and Housing Census in 2001". *Proceedings of Q2008 European Conference on Quality in Official Statistics*, Rome, 8-11 July 2008.

Goodman L.A. (1974), "Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models", *Biometrika*, 61, pp. 215-231.

Hagenaars J.A. (1988), "Latent Structure Models with Direct Effects Between Indicators: Local Dependence Models", *Sociological Methods & Research*, 16, pp. 379-405

Hansen M., Hurwitz W.N., Pritzker L. (1964), "The Estimation and the Interpretation of Gross Differences and the Simple Response Variance". In C.R. Rao (ed.) *Contributions to Statistics*, Statistical Publishing Society, Calcutta, pp. 111-136.

Huang G.H. (2005), "Model Identifiability". In B.S. Everitt and D.C. Howell (eds.) *Encyclopedia of Statistics in Behavioral Science*, Wiley, Chichester, Society, Calcutta, Volume 3, pp. 1249-1251.

Huang G.H., Bandeen-Roche K. (2004), "Building an Identifiable Latent Class Model with Covariate Effects on Underlying and Measured Variables", *Psychometrika*, 66, pp. 109-132.

Istat (2009), "La qualità dei dati", *Conoscere il censimento, 14° Censimento Generale della popolazione e delle abitazioni*. Istat, Roma.

http://www.istat.it/dati/catalogo/20091217_00/la_qualita_dei_dati_conoscere_il_censimento.pdf

Patterson B.H., Dayton C.M., Graubard B.I. (2002), "Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data", *Journal of the American Statistical Association*, 97, pp. 721-729.

Särndal C.E., Swensson B., Wretman J. (1992), *Model Assisted Survey Sampling*: Springer, New York.

US Bureau of the Census (1985), "Evaluating Censuses of Population and Housing", *Statistical Training Document,* ISP-TR-5, Washington DC.

Vermunt J.K. (2002), "Comment on Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data", *Journal of the American Statistical Association*, 97, pp. 736-737."

Vermunt J.K., Magidson J. (2007), "Latent Class Analysis with Sampling Weights: A Maximum Likelihood Approach", *Sociological Methods & Research*, 36, pp. 87-111.