FITTING LOG-LINEAR MODELS TO CONTINGENCY TABLES FROM SURVEYS WITH COMPLEX SAMPLING DESIGNS: AN INVESTIGATION OF THE CLOGG-ELIASON APPROACH

Chris SKINNER (*), Louis-André VALLET (**)

(*) University of Southampton
(**) CNRS & CREST – Laboratoire de Sociologie Quantitative

Abstract: Clogg and Eliason (1987) proposed a simple method for taking account of survey weights when fitting log-linear models to contingency tables. The method includes a simple function of the weights as an offset term in the model. This paper investigates the properties of this method. A rationale is provided for the method when the weights are constant within the cells of the table. For more general cases, however, it is shown that the standard errors produced by the method are invalid, contrary to claims in the literature. The method is compared to the pseudo maximum likelihood method both theoretically and through an empirical study of social mobility, relating daughter's class to father's class using data from the 'Formation & Qualification Professionnelle' survey, conducted in France in 1985. The method of Clogg and Eliason (1987) is found to underestimate standard errors systematically. It is also observed empirically to produce standard errors which are virtually the same as ignoring the weights entirely. The paper concludes by recommending against the use of this method, despite its simplicity. The limitations of the method may be overcome by using the pseudo maximum likelihood method.

Keywords: complex sampling; jackknife; log-linear model; pseudo maximum likelihood; stratification; survey weight.

1. Introduction

Sample survey data provide the basis of much statistical modelling in the social sciences. Classical methods of fitting statistical models can, however, be invalid in the presence of complex sampling designs involving, for example, unequal weights, stratification or multi-stage sampling. To address this concern, there has been considerable development of methods which do take account of complex designs (e.g. Rao and Thomas, 1988; Skinner, Holt and Smith, 1989; Korn and Graubard, 1999; Chambers and Skinner, 2003). One approach, pseudo maximum likelihood (PML) estimation (Binder, 1983; Skinner, 1989), has found increasingly wide application and is now implemented in many statistical software packages, such as SPSS Complex Samples™, STATA (version 10+), LISREL (version 8.7+) and MPlus (version 3+). One advantage of this approach is its generality; it is applicable to a very broad class of complex sampling schemes and to a wide range of statistical modelling methods, especially those based upon generalized linear models but also other methods such as latent variable modelling (Asparouhov, 2005).

In this paper we shall consider an alternative approach proposed by Clogg and Eliason (1987), hereafter referred to as CE, for use with one specific class of modelling methods: log-linear modelling of contingency tables. Although their proposal featured as just one of many ideas in their paper, it has received continuing attention, for example in the standard text book of Agresti (2002, p.391) and in the extension to latent class models of Vermunt and Magidson (2007). The primary rationale for the approach is that it provides a simple way of incorporating survey weights into the estimation of the log-linear model to give "appropriate parameter estimates and standard errors" (Agresti, 2002, p.391). It has also been claimed that the approach leads to valid model testing procedures (Vermunt and Magidson, 2007).

An acknowledged shortcoming of the CE approach is that it fails to take account of stratification or multi-stage sampling in the estimation of standard errors. Since it is very common for social surveys to employ multi-stage sampling and since the impact of multi-stage sampling on standard errors is often much greater than the impact of unequal weights, this is a major disadvantage of the CE approach relative to the PML approach. Nevertheless, surveys do arise where there is no clustering and the

survey weights exhibit appreciable variability. Moreover, there do exist software packages, for which log-linear modelling procedures via the PML approach are not available but the CE approach can be implemented easily (SAS®, for example, appears to fall in this category at present).
The purpose of this paper is to investigate the properties of the CE approach and to compare them to those of the PML approach. For an earlier discussion of this comparison, see Patterson et al. (2002) and Vermunt (2002).
The paper is organised as follows. In section 2, we introduce the log-linear model and explain how unequal probability sampling can affect the fitting of this model. This discussion is designed to motivate the CE approach which is set out in section 3. The theoretical properties of the CE approach are assessed in section 4 under a sampling design, chosen to be favourable to the CE approach. The PML approach is set out briefly in section 5 and then compared theoretically to the CE approach in section 6. An empirical comparison is provided in section 7 using data from the 'Formation & Qualification Professionnelle' survey, conducted in France in 1985. Conclusions are drawn in section 8.

2. The Log-linear Model and the Impact of Sampling

As in Clogg and Eliason (1987) (hereafter CE), we may express a log-linear model for a contingency table as a matrix equation:

$$\log(\mu_s) = X\lambda_s \ , \qquad (1)$$

where $\log(\mu_s)$ is an $M \times 1$ vector containing the logarithms of the expected frequencies for the M cells in the table, X is an $M \times p$ *model matrix* (or design matrix) containing specified values, usually either 0 or 1, and λ_s is a $p \times 1$ vector of unknown parameters, where $p \leq M$.

We subscript μ_s and λ_s by *s* to denote sample. This highlights a basic problem with this model for sample survey data: the parameters of the model are dependent upon the sampling scheme if, as is common, μ_s is defined in terms of the expected *sample* frequencies. To explore this dependence, suppose instead that the log-linear model is defined in terms of the expected *population* frequencies. To emphasize the distinction we remove the subscript *s* and write the population-level model as:

$$\log(\mu) = X\lambda \, , \qquad (2)$$

where we suppose the same design matrix X applies. For simplicity, consider a sampling scheme where all units in the k^{th} cell of the table are included in the sample with probability π_k and let $\log(\pi)$ be the $M \times 1$ vector containing the $\log(\pi_k)$. Then we may write:

$$\log(\mu_s) = \log(\pi) + \log(\mu) \, , \qquad (3)$$

since the expected sample frequency in the k^{th} cell is given by π_k times the expected population frequency. Hence from (2) and (3), we may write:

$$\log(\mu_s) = \log(\pi) + X\lambda \, . \qquad (4)$$

Provided the structure of X is appropriate, the expression in (4) may be equated to the original expression in (1) for some special sampling schemes, for example:
(i) equal probability selection: if all the π_k are equal then $\log(\pi)$ will be a multiple of a vector of ones and if the first column of X is specified to be a vector of ones, the vectors λ_s and λ will only differ in their first element. Such a definition of X is standard (e.g. Agresti, 2002, Ch.8) where the first element of λ_s represents the total sample size and the remaining elements determine the proportions falling into the different cells of the table.
(ii) disproportionate stratified sampling according to one of the cross-classifying variables in a multi-way table: provided X is defined to include the main effects for the stratifying variable, the vectors λ_s and λ will only differ with respect to those elements corresponding to these main effect terms.
Thus, for some simple sampling schemes, it may be reasonable to follow the traditional approach of fitting model (1) to the sample frequencies, provided the design matrix is specified to capture the differential sampling effects and some of the parameter estimates are interpreted as absorbing effects

of sampling, e.g. the grand mean term in example (i) and the main effects for the stratifying variable in example (ii). This approach is not suitable, however, for more complex sampling schemes.

3. The Clogg and Eliason Approach

The CE approach may be motivated by equation (4). Suppose the expected sample frequencies in a contingency table are given in the vector μ_s and that the log-linear model in (4) holds. Then, provided the inclusion probabilities π_k are known and provided it is reasonable to make a standard 'sampling model' assumption, such as a Poisson or multinomial distribution, the parameter vector λ may simply be estimated in a conventional way, e.g. using maximum likelihood (ML), by treating $\log(\pi)$ as an offset term in the model. (See Agresti, 2002, p.385 for a discussion of the use of offset terms.)

We shall generally assume in this paper that the inclusion probabilities are known, typically via survey weights. The more critical issue here is whether the sample frequencies obey a conventional 'sampling model'. In this paper we shall take the conventional sampling model to be a Poisson distribution (e.g. Agresti, 2002, sect. 4.3.1), although equivalent arguments could be presented using the multinomial distribution (the main alternative conventional approach). Whether the sample frequencies do follow a Poisson distribution depends upon the nature of the sampling scheme. We suppose that at the *population* level, the population frequencies N_k in cells *k* do indeed follow Poisson distributions, that is they are outcomes of independent Poisson random variables with means μ_k. A sampling scheme which favours the CE approach is Bernoulli sampling within cells, i.e. where each of the N_k units in cell *k* is included independently in the sample with probability π_k. In this case, standard theory for the Poisson distribution implies that the sample frequencies n_k will also be the outcomes of independent Poisson random variables, now with means $\mu_{sk} = \pi_k \mu_k$. In other words, the sample frequencies will follow a conventional sampling model. It follows that, under this Bernoulli sampling scheme, the approach of fitting the model in a conventional way via expression (4) using $\log(\pi)$ as an offset is valid.

The assumption that the inclusion probabilities are uniform within cells is very restrictive, however, and CE address the general case where survey weights vary between individual units. In this case, they replace the cell-level inclusion probabilities π_k appearing in $\log(\pi)$ in (4) by an estimator of the sampling fraction n_k / N_k in cell *k* given by $z_k = n_k / \hat{N}_k$, where $\hat{N}_k$ is the sum of survey weights across sample units in cell *k*. CE then claim that if the model:

$$\log(\mu_s) = \log(z) + X\lambda \qquad (5)$$

is fitted using conventional ML methods, treating $\log(z)$ (the $M \times 1$ vector containing the $\log(z_k)$) as an offset, then inference about the parameter vector λ, and in particular the implied standard errors, will be appropriate.

We have seen that this claim is valid in one special case, i.e. where Bernoulli sampling is employed within cells and where the survey weights (assumed to be inverse inclusion probabilities) are constant within cells. We argue, however, that this claim is not valid in general for two main reasons.

First, as noted in the introduction, complex sampling schemes impact on standard errors not only through unequal weights but also, and often more importantly, through other features of the design such as cluster sampling. It is well known that cluster sampling can seriously inflate standard errors. The use of the offset term in the CE approach takes no account of potential variance inflation from designs such as cluster sampling and thus will generally lead to invalid standard errors.

The second reason why we argue that the CE approach will, in general, be invalid is that it does not adequately take account of the effects of weight variation. Since this is the main purpose of the discussion in CE, it is the theme which we shall focus on. In the next section, we consider a sampling scheme which is designed to be as favourable to the CE approach as possible, while imposing no constraints on the variability of the weights.

4. Theoretical Properties of CE Approach under a Poisson Sampling Design

4.1. The Poisson Sampling Design

We saw in the previous section that the CE approach is valid if the population units are selected independently with probabilities which are constant within cells. In this section, we retain the assumption that population units are selected independently. This favours the CE approach, in particular by ruling out cluster sampling designs which might lead to underestimation of standard errors by the CE approach. We now, however, allow the inclusion probabilities to vary between units within cells. A sampling design which selects units independently with unequal probabilities is sometimes called a Poisson sampling design (e.g. Hájek, 1981). Since we shall treat the survey weights as reciprocals of the inclusion probabilities, we are also allowing the weights to vary between units.

For simplicity, suppose that there is only a finite number of possible values of the sample inclusion probabilities, denoted $\pi_1, \pi_2, ..., \pi_H$. We shall refer to the different parts of the population which are sampled with different probabilities as strata, i.e. units in stratum *h* are sampled with probability π_h ($h = 1, 2, ..., H$). Note, however, that the Poisson sampling scheme does not ensure a fixed sample size within each stratum and hence this design does not correspond to standard stratified sampling.

Let N_{kh} be the population count in cell *k* in stratum *h*, so that $N_k = \sum_{h=1}^{H} N_{kh}$. In order to construct a framework where the CE approach is natural, we shall assume that the N_{kh} are generated independently as Poisson random variables: $N_{kh} \sim Poisson(\mu_{kh})$. This implies that $N_k \sim Poisson(\mu_k)$, where $\mu_k = \sum_{h=1}^{H} \mu_{kh}$, and also that the numbers n_{kh} of sample units which fall into cell *k* and stratum *h* are independently distributed as: $n_{kh} \sim Poisson(\pi_h \mu_{kh})$. It follows that the distribution of $n_k = \sum_{h=1}^{H} n_{kh}$ is also Poisson, as assumed in the CE approach, i.e.

$$n_k \sim Poisson(\mu_{sk}), \text{ where } \mu_{sk} = \sum_{h=1}^{H} \pi_h \mu_{kh}. \qquad (6)$$

4.2. Point Estimation under the CE Approach

The parameter vector λ is estimated under the CE approach using ML estimation based upon (5), treating the $z_k = n_k / \hat{N}_k$ as fixed. As discussed by Vermunt (2002), the log likelihood used in the CE approach may be expressed as:

$$\log L(\lambda) = \sum_k \{ n_k \log[\mu_{sk}(\lambda)] - \mu_{sk}(\lambda) \}, \qquad (7)$$

where, from (5), $\mu_{sk}(\lambda) = \exp(x_k \lambda) z_k$ and x_k denotes the k^{th} row of X. The point estimator in the CE approach is denoted $\hat{\lambda}_{CE}$ and is the value of λ which maximizes (7).

We show in the Appendix that, providing the model in (2) holds, then $\hat{\lambda}_{CE}$ is consistent for λ (under a suitable asymptotic framework). Thus, the CE approach does make use of the weights to correct for bias from unequal probability sampling, at least in large samples. Note, however, that if any of the cells are empty ($n_k = 0$) then z_k is not defined and thus the estimator $\hat{\lambda}_{CE}$ is not defined.

4.3. Standard Error Estimation under the CE Approach

CE propose to obtain standard errors by treating the expression in (7) as a likelihood function with z_k treated as fixed. It is shown in the Appendix that this approach leads to a variance-covariance matrix

of $\hat{\lambda}_{CE}$ of the form $\tilde{J}^{-1}$, where $\tilde{J} = \sum_k \mu_{sk} x_k' x_k$ and this matrix may be estimated by replacing the μ_{sk} by n_k. The CE standard errors are obtained as the square roots of the diagonal elements of this matrix.

We show in the Appendix that in fact, if we properly take account of the fact that z_k is not fixed, the (large sample) variance-covariance matrix of $\hat{\lambda}_{CE}$ can be expressed as:

$$v(\hat{\lambda}_{CE}) = \tilde{J}^{-1} + \tilde{J}^{-1}\{\sum_k c_k^2 \mu_{sk} x_k' x_k\}\tilde{J}^{-1}, \qquad (8)$$

where c_k^2 is the squared coefficient of variation of the survey weights within cell *k*. Thus the CE approach will generally underestimate the standard error of each parameter estimate. It will only provide valid standard errors if the survey weights are constant within cells as discussed in section 2, but this is not a case of great interest since the CE approach was specifically formulated to deal with situations where the weights vary.

When the distribution of the weights is the same in each cell *k* so that $c_k^2 = c^2$ does not depend on *k*, expression (8) simplifies further to $v(\hat{\lambda}_{CE}) = (1+c^2)\tilde{J}^{-1}$. Hence, in large samples, the CE estimator underestimates the variance-covariance matrix by a factor $(1+c^2)$ and, in particular, the variance of each element of $\hat{\lambda}_{CE}$ is underestimated by this factor. In the general case when the c_k^2 depend on *k*, the degree of underestimation may be interpreted as an average of the $1+c_k^2$.

One special case under the assumed sampling design, where the distribution of the weights is the same in each cell *k,* arises when the strata are independent of the cell variables, so that $\mu_{kh} = \mu_k \phi_h$, where ϕ_h denotes the probability of falling in stratum *h* and $\sum \phi_h = 1$. This case may be called 'non-informative stratification'. In this case, c_k^2 is equal to the overall coefficient of variation of the weights across all cells:

$$c_k^2 = c^2 = \sum_{h=1}^{H} \pi_h \phi_h \sum_{h=1}^{H} \pi_h^{-1} \phi_h - 1.$$

The above results imply some possible modifications of the CE approach. For example, if the weights are judged to be roughly independent of the cell variables then the CE standard errors could be modified by multiplying them by a factor $\sqrt{1+\hat{c}^2}$ where $\hat{c}$ is the sample coefficient of variation of the weights. We do not pursue this idea, however.

5. Pseudo Maximum Likelihood (PML) Approach

The point estimator of λ in the pseudo maximum likelihood approach, denoted $\hat{\lambda}_{PML}$, is obtained by fitting the population level model (2) to the weighted population counts $\hat{N}_k$. This may be achieved by using any of the standard algorithms used for ML estimation of log-linear models. The standard errors of any of the elements of $\hat{\lambda}_{PML}$ may be obtained by any of the usual variance estimation methods used in survey sampling. See Rao and Thomas (1988, sect. 5.2) for details and, in particular, the use of linearization variance estimators. We consider just the jackknife method for the case of stratified multi-stage sampling. This involves constructing replicate sets of survey weights $w^{(hc)}$ corresponding to each of the primary sampling units (PSUs) $c = 1, 2, ..., n_h$, within each stratum $h = 1, 2, ..., H$. The set of weights $w^{(hc)}$ is constructed by removing from the sample all units from PSU c in stratum h and inflating the weights of all other units in this stratum by the factor $n_h/(n_h - 1)$ (Rust and Rao, 1996). For each set of replicate weights $w^{(hc)}$, λ is estimated in the same way that the point estimator of interest $\hat{\lambda}$ is computed except that the original survey weights are replaced by the set of

replicate weights $w^{(hc)}$. The resulting estimator is denoted $\hat{\lambda}^{(hc)}$. This is repeated for each set of replicate weights and the jackknife estimator of the variance of $\hat{\lambda}$ is then given by:

$$v_j(\hat{\lambda}) = \sum_{h=1}^{H}\left(\frac{n_h - 1}{n_h}\right)\sum_{c=1}^{n_h}(\hat{\lambda}^{(hc)} - \hat{\lambda})(\hat{\lambda}^{(hc)} - \hat{\lambda})' \qquad (9)$$

There is a number of variants of the jackknife estimator. In particular, it is possible to reduce the computation by calculating $\hat{\lambda}^{(hc)}$ only for $n_h - 1$ values of c within each stratum h. The jackknife estimator is consistent for the variance of $\hat{\lambda}$ under general assumptions about the survey weights and the stratified multi-stage design (Shao and Tu, 1995, Ch.6). Note that this property applies to a broad class of estimators $\hat{\lambda}$, including not just $\hat{\lambda}_{PML}$ but also $\hat{\lambda}_{CE}$, for example.

6. Theoretical Comparison of the CE and PML Approaches

6.1. Comparison of Standard Error Estimators

As we showed in section 4.3, the standard error estimators produced by the CE approach are generally biased and inconsistent as a result either of non-independence between the selection of different units, e.g. via cluster sampling, or because of unequal survey weights within the cells of the table. On the other hand, the PML method is designed so that the standard error estimators are consistent.

6.2. Comparison of Point Estimators

The point estimators, $\hat{\lambda}_{PML}$ and $\hat{\lambda}_{CE}$, for the two approaches are not identical (as noted by Vermunt, 2002) but they are both consistent for the true value of λ if the model in (2) holds. If the model in (2) does not hold then, as the sample size increases, $\hat{\lambda}_{PML}$ and $\hat{\lambda}_{CE}$ will not in general converge to the same quantities. Whether the limiting value of either $\hat{\lambda}_{PML}$ or $\hat{\lambda}_{CE}$ is of interest depends on the scientific objectives. One possible advantage of $\hat{\lambda}_{PML}$ is that it can be shown that its limiting value does not depend on the sampling scheme (cf. comments of Patterson et al., 2002).

We next compare the variances of the elements of $\hat{\lambda}_{PML}$ and $\hat{\lambda}_{CE}$ under the assumption that the model in (2) is correct. To do this we write both estimators as solutions of the estimating equations:

$$\sum_k a_k\{\hat{N}_k - \exp(x_k\lambda)\}x_k = 0 \qquad (10)$$

where for $\hat{\lambda}_{PML}$ we set $a_k = 1$ and for $\hat{\lambda}_{CE}$ we set $a_k = z_k$ (see Appendix).

The variance of any linear combination of the elements of the vector λ solving (10) is minimized by setting $a_k x_k$ proportional to:

$$\frac{1}{\mathrm{var}(\hat{N}_k)}\frac{\partial \exp(x_k\lambda)}{\partial\lambda} = \frac{1}{\sum_{h=1}^{H}\pi_h^{-1}\mu_{kh}}\mu_k x_k .$$

Hence the optimal choice of a_k is

$$a_{kopt} \propto \mu_k / \sum_{h=1}^{H}\pi_h^{-1}\mu_{kh}$$

and an estimate of the optimal a_k is $\hat{a}_{kopt} = S_{1k}/S_{2k}$, where S_{1k} is the sum of weights and S_{2k} is the sum of squared weights in cell *k*. One special case arises when the weights are constant in which case a_{kopt} is constant and both $\hat{\lambda}_{PML}$ and $\hat{\lambda}_{CE}$ are equal and optimal.

The CE point estimator $\hat{\lambda}_{CE}$ is optimal if the weights are constant within cells. In general, however, the CE approach takes no account of weight variation within cells and thus will not be optimally efficient. The PML point estimator $\hat{\lambda}_{PML}$ will be close to efficient when the weights are variable but tend to be unrelated to the cells as, for example, in the case of non-informative stratification mentioned in section 4.3. There seems no reason to expect $\hat{\lambda}_{CE}$ to tend to be always more efficient than $\hat{\lambda}_{PML}$ nor vice versa.

7. Empirical Comparison of the CE and PML Approaches

We now set out to compare the CE and PML approaches empirically. We use data from the *1985 Enquête Formation & Qualification Professionnelle*, a survey with complex sampling design and post hoc reweighting that was conducted by the French Statistical Office and for which a stratum variable and a weight variable are available in the data file. In the following sections, we briefly describe the sampling characteristics of the survey, then the data and contingency table we use and the log-linear model we consider. Finally, we systematically compare the corresponding CE and PML estimators and standard errors.

7.1. Sampling Characteristics of the Survey

The *1985 Formation & Qualification Professionnelle* survey was designed to be representative of the population of ordinary households in the 1982 census and covered all employed and unemployed persons, whatever their age, and all persons not in the labour market aged between 13 and 69 in 1982. The survey was administered to a stratified sample of 46,500 individuals drawn from the 1982 census with sampling fractions that varied between about 1/200 and 1/2500 (Laulhé and Soleilhavoup, 1987; Gollac, Laulhé and Soleilhavoup, 1988a, 1988b).

More precisely, the survey sample was drawn from the (very large) 1982 master sample in order to concentrate interviews within the geographical areas covered by the team of interviewers of the French Statistical Office so as to minimize travelling costs. The sampling was divided into two phases. First, a sample of 38,000 dwellings was drawn from the master sample so that all dwellings in the population had an equal probability of inclusion of 1/200. Then the individuals in these dwellings were stratified according to nationality in two categories (French, foreigners), position as regards the labour market, socio-economic class and age group. Second, the final sample of 46,500 individuals was drawn from the 73 resulting strata using different (sub-)sampling fractions, ranging from 13% to 100%, determined by the objectives of the survey. As a result, the probabilities of inclusion of the different individuals in the census population ranged between 1/200 and 1/2690. These probabilities are referred to as the *initial sampling fractions*. The geographical clustering in the master sample is not identified in the file and will be ignored in our analyses for reasons of practicality and simplicity. The possible clustering of individuals in dwellings will also be treated as negligible since it will happen with very small probability (especially since we shall restrict attention to a subsample of women in a particular age range). In summary, the sample will be treated as being derived by (disproportionate) stratified simple random sampling.

The interviews were completed between mid April and the end of June 1985 with 39,233 completed questionnaires collected. To take account of not only the disproportionate stratification but also other sources of missing data (because of unknown addresses, long term absences and refusals), weights were constructed as ratios of census counts to counts of survey respondents within weighting classes defined by the strata cross-classified with residential area at the census (rural, urban, or Parisian). The resulting final weight variable is available for each case in the data file, for use in producing estimates for the population.

7.2. Data, Contingency Table and Log-linear Model

For our analysis we restrict attention to the sub-sample of 5,159 women, with French nationality at the date of the survey, aged between 35 and 59 at the end of December 1985, currently employed at the date of the survey, and who reported information about their current socio-economic class and their father's socio-economic class when they stopped attending school or university on a regular basis. *Table 1* displays characteristics of this sub-sample across the strata. The 5,159 women belong to 18

different strata with initial sampling fractions varying between 1/310 and 1/2500. The distribution is very uneven as only 2 women appear in the least numerous stratum while 1,581 belong to the most numerous one. For descriptive purposes, Table 1 also presents the mean and standard deviation of the final weight in each stratum. The discrepancy in each stratum between the average final weight and the inverse of the initial sampling fraction reflects the adjustments that result from missing data, and the standard deviation of the final weight reflects the variability of the case-by-case weighting within each stratum.

Table 1 – Characteristics of the sub-sample used for analysis in the different strata

Stratum (status as recorded in master sample in 1982; note that status in 1985 might differ, e.g. on nationality, and age will increase by 3 years)	**Sample size**	**Initial sampling fraction**	**Mean of final weight**	**Standard deviation of final weight**
French women, in the labour market, farmers, aged 32 – 51	234	1/940	959.85	123.67
French women, in the labour market, farmers, aged 52+	83	1/1250	1245.55	47.53
French women, in the labour market, artisans and shopkeepers, 32 – 51	223	1/1040	1144.54	87.61
French women, in the labour market, artisans and shopkeepers, 52 +	28	1/1360	1487.79	120.05
French women, in the labour market, company managers and higher-grade professionals, 32 – 51	747	1/310	344.25	44.21
French women, in the labour market, company managers and higher-grade professionals, 52+	94	1/340	388.76	67.22
French women, in the labour market, lower-grade professionals, 32 – 51	1064	1/600	669.29	111.77
French women, in the labour market, lower-grade professionals, 52 +	101	1/620	720.25	76.06
French women, in the labour market, non manual workers, 32 – 51	1581	1/830	934.72	129.21
French women, in the labour market, non manual workers, 52 +	214	1/830	946.18	111.74
French women, in the labour market, manual workers, 32 – 51	535	1/760	839.90	74.58
French women, in the labour market, manual workers, 52 +	60	1/1080	1193.80	50.60
French women, in the labour market, unemployed who never worked before	13	1/400	491.69	93.49
French women, students	7	1/900	1000.14	110.70
French women, who were previously in the labour market, less than 70	2	1/2270	2464.00	247.49
Other French women, out of the labour market, 32 – 51	146	1/2500	2794.81	620.64
Other French women, out of the labour market, 52 +	17	1/2500	2794.76	389.26
Foreign women, in the labour market, employed or unemployed, 32 – 51	10	1/730	831.20	189.88
Total	5159	-	850.34	451.33

Our analysis is based on the 7 x 7 two-way contingency table that cross-classifies women's socio-economic class with their father's socio-economic class when they stopped attending school or university on a regular basis. The mobility table uses a discrete and unordered socio-economic classification defined as follows: (1) higher-grade salaried professionals; (2) company managers and liberal professions; (3) lower-grade salaried professionals; (4) artisans and shopkeepers; (5) non-manual workers; (6) foremen and manual workers; (7) farmers. *Table 2* presents both unweighted frequencies and weighted frequencies in the mobility table after rescaling the latter to the exact sample size.

We aim at analysing the structure and strength of the association between father's socio-economic class and daughter's socio-economic class in 1985 within French society. For that purpose, we use the log-linear model proposed by Hauser (1978, 1980) that identifies the two-way interaction effects by constraining some of them to be equal across cells of the contingency table. Assuming that i and j respectively index father's class and daughter's class, that the cells (i, j) are assigned to K mutually exclusive and exhaustive subsets and that each of those sets shares a common interaction parameter δ_k, the logged expected frequency in cell (i, j) of the mobility table is expressed as follows:

$$Log\, m_{ij} = \alpha + \beta_i + \gamma_j + \delta_k \qquad \text{if the cell } (i, j) \text{ belongs to subset } k.$$

Thus, aside from total (α), row (β_i), and column (γ_j) effects, each expected frequency is determined by only one interaction parameter (δ_k) which "reflects the density of mobility or immobility

in that cell relative to that in other cells in the table" (Hauser, 1980, p.416). The interaction parameters of the model may therefore "be interpreted as indexes of the social distance between categories of the row and column classifications" (Hauser, 1980, p.416).

Table 2 – Unweighted frequencies and weighted (rescaled) frequencies in the mobility table

Daughter's class / **Father's class**	Frequency	1	2	3	4	5	6	7	Total
1 Higher-grade salaried professionals	Unweighted	164.00	25.00	136.00	12.00	59.00	9.00	0.00	405.00
	Weighted	81.23	13.01	113.18	15.35	66.32	8.08	0.00	297.17
2 Company managers and liberal professions	Unweighted	56.00	27.00	37.00	14.00	28.00	3.00	3.00	168.00
	Weighted	28.78	11.72	38.22	14.46	32.45	2.65	7.01	135.29
3 Lower-grade salaried professionals	Unweighted	95.00	16.00	161.00	15.00	115.00	18.00	4.00	424.00
	Weighted	48.08	11.44	129.70	22.79	131.79	18.20	4.77	366.78
4 Artisans and shopkeepers	Unweighted	97.00	35.00	219.00	78.00	200.00	35.00	8.00	672.00
	Weighted	52.25	21.35	174.45	118.41	223.37	39.57	14.27	643.67
5 Non-manual workers	Unweighted	59.00	7.00	145.00	32.00	182.00	29.00	3.00	457.00
	Weighted	30.18	3.68	120.03	53.42	216.57	28.65	4.17	456.70
6 Foremen and manual workers	Unweighted	128.00	18.00	419.00	124.00	930.00	339.00	37.00	1995.00
	Weighted	64.18	14.88	361.46	184.12	1065.19	355.76	47.06	2092.66
7 Farmers	Unweighted	38.00	8.00	164.00	73.00	342.00	136.00	277.00	1038.00
	Weighted	20.29	5.63	134.71	101.98	394.83	140.49	368.80	1166.73
Total	Unweighted	637.00	136.00	1281.00	348.00	1856.00	569.00	332.00	5159.00
	Weighted	324.99	81.71	1071.75	510.54	2130.52	593.40	446.08	5159.00

Note: Weighted frequencies are rescaled to the sample size by multiplying them by the ratio 5159/4386881.

A previous paper (Vallet, 2005) relied on sociological hypotheses to build such a model of the father-daughter mobility table with $K = 7$ interaction parameters. The allocation of the interaction effects across the cells of the contingency table that characterizes the postulated model is presented in the upper part of *Table 3*.

Table 3 – Initial model and final model for the structure of the association in the mobility table

	1	2	3	4	5	6	7
1 – Higher-grade salaried professionals	II	III	IV	V	VI	VII	VII
2 – Company managers and liberal professions	III	II	IV	IV	VI	VII	VII
3 – Lower-grade salaried professionals	IV	IV	IV	V	V	VI	VII
4 – Artisans and shopkeepers	V	IV	V	IV	V	VI	VI
5 – Non-manual workers	VI	VI	V	V	V	V	VI
6 – Foremen and manual workers	VII	VII	VI	VI	V	IV	V
7 – Farmers	VII	VII	VII	VI	VI	V	I

	1	2	3	4	5	6	7
1 – Higher-grade salaried professionals	II	II	III	IV	V	VI	VII
2 – Company managers and liberal professions	II	II	III	III	V	VI	IV
3 – Lower-grade salaried professionals	III	III	III	IV	IV	V	VI
4 – Artisans and shopkeepers	IV	III	IV	III	V	V	V
5 – Non-manual workers	V	V	IV	IV	IV	V	V
6 – Foremen and manual workers	VI	VI	V	IV	IV	III	IV
7 – Farmers	VII	VI	VI	IV	V	IV	I

Note: Rows and columns in the matrices respectively correspond to father's socio-economic class and daughter's socio-economic class. Among the interaction effects, I is supposed to be the strongest and VII the weakest.

In our underlying hypotheses, we assumed that the association between origin class and destination class is symmetrical across the main diagonal, and we also emphasized that three aspects must be considered to describe the structure and strength of the association: the relative desirability of different socio-economic class positions; the relative advantages afforded to individuals by different socio-economic class origins; and the relative barriers that face individuals in seeking access to different socio-economic class positions. Although this initial model did not satisfactorily fit the data on conventional criteria of statistical significance, the expected frequencies were generally close to the observed frequencies. On the basis of an examination of residuals, a few modifications were introduced to reallocate some cells to a different interaction parameter (Vallet, 2005). The final model, with again $K = 7$ interaction parameters, that resulted from this process and proved to satisfactorily fit the data is presented in the lower part of Table 3.
For the initial and final log-linear models, we now compare estimates and standard errors obtained in four different ways: the standard ML approach for the tables of unweighted frequencies and of weighted rescaled frequencies; the CE approach; and the PML approach.

7.3. Computation

To implement the first three approaches, we used both the CATMOD and GENMOD procedures in SAS® software and obtained exactly the same results. To implement the CE approach, we computed z_{ij} as the ratio of the unweighted frequency to the weighted rescaled frequency in cell (i, j) and then introduced $Log(z_{ij})$ as an offset in the log-linear model (see section 3). As z_{ij} cannot be defined for cell *(1,7)* since it is empty in Table 2, we decided to treat this cell as a structural zero for all four approaches.
No SURVEY procedure for log-linear modelling is available in the SAS® software that could be used for a direct implementation of the PML approach. For that purpose, we therefore use the CATMOD SAS® procedure in the context of the SASMOD module of the IVEware software (Raghunathan, Solenberger and Van Hoewyk, 2002). SASMOD is a SAS macro that provides a framework for performing complex design analysis, with or without missing data, for a collection of SAS® procedures. Before invoking the SAS® procedure, the SASMOD setup file must include the definition of three variables: a stratum variable (here, the variable that identifies to which of the 18 strata (Table 1) each observation belongs); a weight variable (here, the (rescaled) weight variable available in the data file); and a cluster variable (here, as no Primary Sampling Unit (PSU) variable is available, we use a pseudo variable with a different value between 1 and 5,159 for each observation). Then the SASMOD module computes the variance estimates using a variant of the jackknife method in (9) based upon $U - H$ (here 5,141) replicate estimates, where H denotes the number of strata and U the total number of PSUs (personal communication from T. E. Raghunathan, 2006). For either the initial model or the final model estimated on our data, SASMOD computations take about 50 minutes with an Intel® Pentium® IV 2.2 GHz processor.
To estimate the true variance of the CE point estimator, we implemented the jackknife method in (9) by nesting the GENMOD procedures in a loop in SAS®. We also applied this method to each of the other point estimators and found that with the PML estimator we obtained exactly the same results as with SASMOD.

7.4. Comparison of Parameter Estimates and Standard Errors

Table 4 presents parameter estimates and standard errors obtained for the initial and final models under all four approaches. We consider the point estimates first. The estimates obtained by applying the standard ML approach to the weighted rescaled table are identical to those from the PML approach as expected. Thus, there are really just three sets of point estimates to compare. The most marked differences are between the unweighted estimates and the other two (PML and CE) estimates. As discussed in section 6.2, both the latter estimators will be approximately unbiased if the model is true. We cannot be certain that either of the models is true but it seems reasonable to view the differences between the unweighted estimates and the other two estimates as evidence of bias in the former procedure (cf. Clogg and Eliason (1987, p.22)). This bias is especially pronounced in the case of the γ_j parameters and this may be attributed to the strong correlation between the column variable (women's socio-economic class in 1985) and one of the stratifying variables (women's socio-economic class at the census) upon which the sampling is differential. The PML and CE estimates are broadly

similar and should not lead to any difference in substantive interpretation for either model. Leaving aside consideration of the standard errors, there seems no strong reason to prefer one set of estimates to the other. One possible argument in favour of the PML estimator, following Patterson et al. (2002) and mentioned in section 6.2, is that the PML estimator is 'estimating' a well-defined population quantity if the model is false, whereas the CE estimator is then estimating a quantity dependent on the arbitrariness of the sampling scheme.

Table 4 – Comparison of parameter estimates and standard errors (in parentheses)

Parameter	Initial model				Final model			
	Unweighted	Weighted rescaled	Clogg & Eliason	Pseudo maximum likelihood	Unweighted	Weighted rescaled	Clogg & Eliason	Pseudo maximum likelihood
β_1	-1.813	-1.825	-1.828	-1.825	-1.747	-1.754	-1.763	-1.754
(se)	(0.087)	(0.086)	(0.086)	(0.098)	(0.084)	(0.083)	(0.083)	(0.093)
β_2	-2.626	-2.621	-2.612	-2.621	-2.663	-2.610	-2.632	-2.610
(se)	(0.107)	(0.108)	(0.106)	(0.133)	(0.102)	(0.105)	(0.102)	(0.125)
β_3	-1.532	-1.559	-1.549	-1.559	-1.492	-1.517	-1.514	-1.517
(se)	(0.079)	(0.078)	(0.079)	(0.090)	(0.076)	(0.075)	(0.076)	(0.085)
β_4	-0.856	-0.857	-0.855	-0.857	-0.633	-0.614	-0.643	-0.614
(se)	(0.069)	(0.067)	(0.070)	(0.079)	(0.061)	(0.059)	(0.061)	(0.068)
β_5	-1.134	-1.104	-1.111	-1.104	-1.036	-1.013	-1.021	-1.013
(se)	(0.072)	(0.072)	(0.073)	(0.082)	(0.067)	(0.065)	(0.067)	(0.075)
β_6	0.492	0.510	0.505	0.510	0.487	0.507	0.497	0.507
(se)	(0.049)	(0.049)	(0.049)	(0.056)	(0.048)	(0.047)	(0.048)	(0.054)
β_7	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0
γ_1	2.187	1.179	1.261	1.179	2.177	1.196	1.238	1.196
(se)	(0.149)	(0.139)	(0.149)	(0.166)	(0.148)	(0.138)	(0.148)	(0.157)
γ_2	0.585	-0.269	-0.182	-0.269	0.450	-0.373	-0.321	-0.373
(se)	(0.169)	(0.169)	(0.170)	(0.205)	(0.167)	(0.167)	(0.167)	(0.198)
γ_3	2.889	2.360	2.424	2.360	2.855	2.341	2.376	2.341
(se)	(0.140)	(0.120)	(0.140)	(0.150)	(0.139)	(0.119)	(0.139)	(0.146)
γ_4	1.473	1.508	1.555	1.508	1.204	1.253	1.282	1.253
(se)	(0.147)	(0.124)	(0.148)	(0.156)	(0.147)	(0.124)	(0.147)	(0.153)
γ_5	3.089	2.895	2.943	2.895	3.167	2.971	3.003	2.971
(se)	(0.137)	(0.116)	(0.137)	(0.143)	(0.137)	(0.116)	(0.137)	(0.144)
γ_6	1.605	1.297	1.349	1.297	1.638	1.340	1.370	1.340
(se)	(0.146)	(0.126)	(0.146)	(0.150)	(0.146)	(0.126)	(0.146)	(0.150)
γ_7	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0
δ_I	3.561	3.451	3.569	3.451	4.163	4.096	4.138	4.096
(se)	(0.163)	(0.146)	(0.163)	(0.189)	(0.228)	(0.266)	(0.228)	(0.252)
δ_{II}	2.730	2.619	2.660	2.619	3.215	3.104	3.123	3.104
(se)	(0.119)	(0.147)	(0.118)	(0.135)	(0.191)	(0.251)	(0.191)	(0.214)
δ_{III}	2.396	2.297	2.326	2.297	2.276	2.252	2.275	2.252
(se)	(0.150)	(0.189)	(0.149)	(0.186)	(0.187)	(0.245)	(0.187)	(0.208)
δ_{IV}	1.683	1.633	1.700	1.633	1.692	1.658	1.675	1.658
(se)	(0.086)	(0.093)	(0.085)	(0.105)	(0.183)	(0.243)	(0.183)	(0.204)
δ_V	1.161	1.078	1.154	1.078	1.245	1.217	1.240	1.217
(se)	(0.084)	(0.092)	(0.084)	(0.103)	(0.181)	(0.241)	(0.181)	(0.201)
δ_{VI}	0.683	0.641	0.699	0.641	0.731	0.708	0.702	0.708
(se)	(0.072)	(0.080)	(0.072)	(0.087)	(0.177)	(0.239)	(0.177)	(0.196)
δ_{VII}	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0
Deviance	86.11	77.12	75.58	-	47.71	33.69	34.77	-
DF	29	29	29	-	29	29	29	-

As regards standard errors in Table 4, only those for the PML estimator have been estimated in a way which takes appropriate account of the complex sampling. Since the weighted rescaled and the PML point estimators are identical, the differences between the standard errors for these two estimators demonstrate that the former method can often lead to seriously incorrect standard errors, as noted by Clogg and Eliason (1987, p.24). We have also calculated valid standard errors for the unweighted point estimators using the jackknife method and found that these too can differ from the values in Table 4, although the differences are more minor. We do not report or comment on these results further, however, since the unweighted point estimators show clear bias and so their standard errors are of little interest.

Of much more importance to the theme of this paper are the standard errors for the CE approach. The standard errors of the CE point estimator obtained via a valid jackknife approach are compared in *Table 5* with those obtained via the CE approach. We observe that the CE approach uniformly

underestimates the standard errors. The jackknife value is often at least 10% higher and sometimes at least 20% higher. Our empirical investigation therefore illustrates how the CE variance estimator can systematically underestimate the true variability. Moreover, we observe in Table 4 that the standard errors obtained under the CE approach are virtually identical to those of the unweighted approach. Hence the device of including the offset term in the model seems to provide virtually no benefit in capturing the effect of unequal sampling weights on the standard error.

Table 5 – Comparison of estimated standard errors for Clogg-Eliason estimator: Clogg-Eliason approach vs Jackknife method allowing for complex design

Parameter	Initial model		Final model	
	Clogg & Eliason	Jackknife	Clogg & Eliason	Jackknife
β_1	0.086	0.102	0.083	0.096
β_2	0.106	0.130	0.102	0.122
β_3	0.079	0.090	0.076	0.086
β_4	0.070	0.078	0.061	0.068
β_5	0.073	0.081	0.067	0.074
β_6	0.049	0.055	0.048	0.054
β_7	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0
γ_1	0.149	0.158	0.148	0.155
γ_2	0.170	0.204	0.167	0.201
γ_3	0.140	0.144	0.139	0.143
γ_4	0.148	0.152	0.147	0.149
γ_5	0.137	0.140	0.137	0.141
γ_6	0.146	0.147	0.146	0.148
γ_7	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0
δ_I	0.163	0.181	0.228	0.253
δ_{II}	0.118	0.139	0.191	0.218
δ_{III}	0.149	0.192	0.187	0.212
δ_{IV}	0.085	0.101	0.183	0.207
δ_V	0.084	0.099	0.181	0.204
δ_{VI}	0.072	0.084	0.177	0.200
δ_{VII}	Fixed at 0	Fixed at 0	Fixed at 0	Fixed at 0

Finally, we compare the jackknife estimates for the CE estimator in Table 5 with the jackknife estimates for the PML estimator in Table 4. We observe that these are very similar. This is not surprising since the values of the point estimators were similar too. It implies that, at least for this application, there is no evidence of an efficiency advantage of the CE point estimator compared to the PML approach.

8. Conclusions

Clogg and Eliason (1987) proposed, amongst many other ideas, a simple method for handling survey weights in log-linear modelling. This method has continued to be cited. We have investigated the properties of this method using both statistical theory and an empirical study of social mobility using French survey data. Despite its simplicity, we recommend against the use of the method for the following reasons:

- the standard errors produced by the method are invalid in general as a means of capturing the effect of weighting, contrary to claims in the literature. They are only valid in one or two

very special cases. They generally underestimate the true standard errors. In our empirical study the method produced standard errors which were virtually identical to ignoring the survey weights entirely.

- the standard errors produced by the method take no account of the effects of complex sampling other than weights. In the authors' experience, there is a common misperception among survey data users that weights are the only aspect of complex sampling that need to be taken account of in data analyses, whereas for most social surveys, multi-stage sampling has a much greater impact than weighting on standard errors.
- the method does correct for bias in point estimation but we see no clear advantages of this approach compared to the equally simple approach of fitting a model to a weighted table.
- we are not aware of any rigorous theoretical justification of claims in the literature (e.g. Vermunt and Magidson, 2007) that this method leads to valid model testing procedures in the presence of survey weights and, on the basis of the theoretical work in this paper, we do not find this plausible.

We consider that the pseudo maximum likelihood (PML) approach overcomes these limitations of the method of Clogg and Eliason (1987). Although the PML method is not implemented in log-linear modelling procedures in all standard software packages, it is often feasible to employ replication variance estimation methods, such as the jackknife or the bootstrap, where the point estimates are repeatedly computed for different replicates to obtain valid standard errors.

Appendix: Proofs of Results in Section 4

Consider first the consistency of $\hat{\lambda}_{CE}$ which maximizes (7) or, alternatively, may be defined as the solution of the estimating equations:

$$\partial[\log L(\lambda)]/\partial\lambda = \sum_k \{n_k - \exp(x_k\lambda)z_k\}x_k = 0,$$

which may also be expressed as:

$$\sum_k [\hat{N}_k - \exp(x_k\lambda)]z_k x_k = 0 \ . \qquad \text{(A1)}$$

Consider the set-up of section 4.1, where $\hat{N}_k = \sum_{h=1}^{H} n_{kh}\pi_h^{-1}$, and assume an asymptotic framework, where H and $\pi_1, \pi_2, ..., \pi_H$ are fixed and the $\mu_1, \mu_2, ..., \mu_H$ each increase to infinity. In this framework, the $\hat{N}_k / \mu_k$ will each converge in probability to unity. Moreover, if the model in (2) is correct, so that $\mu_k = \exp(x_k\lambda)$, then $\hat{N}_k / \exp(x_k\lambda)$ will converge in probability to unity. It then follows from (A1) that, provided the design matrix is defined in a non-redundant way so that (in large samples) (A1) has a unique solution, $\hat{\lambda}_{CE}$ will be consistent for λ.

Consider now the CE standard errors obtained from the information matrix based on (7), given by:

$$J(\lambda) = -\partial^2[\log L(\lambda)]/\partial\lambda^2 = \sum_k \exp(x_k\lambda)z_k x_k' x_k \ .$$

Hence the CE estimator of the variance covariance matrix of $\hat{\lambda}_{CE}$ is:

$$\hat{V}_{CE}(\hat{\lambda}_{CE}) = J(\hat{\lambda}_{CE})^{-1}. \qquad \text{(A2)}$$

When the model in (2) holds, we may write alternatively that $J(\lambda) = \sum_k \mu_k z_k x_k x_k'$ and in large samples:

$$J(\lambda) \doteq \tilde{J}(\lambda) = \sum_k \mu_{sk} x_k' x_k \ . \qquad \text{(A3)}$$

The actual variance-covariance matrix of $\hat{\lambda}_{CE}$ may be obtained by linearization as follows. The first order Taylor expansion of $\exp(x_k\hat{\lambda}_{CE})$ around $\hat{\lambda}_{CE} = \lambda$ is:

$$\exp(x_k\hat{\lambda}_{CE}) \doteq \exp(x_k\lambda) + \exp(x_k\lambda)x_k(\hat{\lambda}_{CE} - \lambda).$$

Substituting into (A1) gives:

$$\sum_k [\hat{N}_k - \exp(x_k\lambda) - \exp(x_k\lambda)x_k(\hat{\lambda}_{CE} - \lambda)]z_k x_k' \doteq 0$$

or

$$\hat{\lambda}_{CE} \doteq \lambda + \{\sum_k \exp(x_k\lambda)z_k x_k' x_k\}^{-1} \sum_k [\hat{N}_k - \exp(x_k\lambda)]z_k x_k'$$

$$= \lambda + J(\lambda)^{-1} \sum_k [\hat{N}_k - \exp(x_k \lambda)] z_k x_k \,'.$$

Thus, in large samples, we may approximate the variance-covariance matrix of $\hat{\lambda}_{CE}$ by:

$$\text{var}\{J(\lambda)^{-1} \sum_k [\hat{N}_k - \exp(x_k \lambda)] z_k x_k \,'\}, \qquad \text{(A4)}$$

which is equivalent in large samples, using (A3), to:

$$\tilde{J}(\lambda)^{-1} \text{ var}\{\sum_k (\hat{N}_k - \mu_k)(\mu_{sk} / \mu_k) x_k \,'\} \tilde{J}(\lambda)^{-1}$$
$$= \tilde{J}(\lambda)^{-1} \{\sum_k (\mu_{sk} / \mu_k)^2 \text{ var}(\hat{N}_k) x_k \,' x_k\} \tilde{J}(\lambda)^{-1}$$

Now

$$\text{var}(\hat{N}_k) = \sum_{h=1}^{H} \pi_h^{-2} \text{ var}(n_{kh}) \; = \sum_{h=1}^{H} \pi_h^{-1} \mu_{kh} \,.$$

So the (large sample) variance-covariance matrix of $\hat{\lambda}_{CE}$ can be expressed as:

$$= \tilde{J}(\lambda)^{-1} + \tilde{J}(\lambda)^{-1} \{\sum_k c_k^2 \mu_{sk} x_k \,' x_k\} \tilde{J}(\lambda)^{-1}, \qquad \text{(A5)}$$

where $c_k^2 = [\sum_{h=1}^{H} \pi_h \mu_{kh} \sum_{h=1}^{H} \pi_h^{-1} \mu_{kh} - (\sum_{h=1}^{H} \mu_{kh})^2] / \mu_k^2 = [\sum_{h=1}^{H} \pi_h \mu_{kh} \sum_{h=1}^{H} \pi_h^{-1} \mu_{kh}] / \mu_k^2 - 1 \,.$ (A6)

Note that $c_k^2 \geq 0$ from the Cauchy-Schwarz inequality. Hence the CE approach generally underestimates standard errors of any element of $\hat{\lambda}_{CE}$. The CE standard errors will only be appropriate if $c_k = 0$ for each *k* that is if $\pi_h \mu_{kh} \propto \pi_h^{-1} \mu_{kh}$ which requires that the π_h are constant, i.e. $\pi_h = \pi$. To show that c_k is the coefficient of variation of the weights in cell *k*, let

$$S_{1k} = \hat{N}_k = \sum_{h=1}^{H} n_{kh} \pi_h^{-1} \doteq \sum_{h=1}^{H} \mu_{kh} = \mu_k \text{ and } S_{2k} = \sum_{h=1}^{H} n_{kh} \pi_h^{-2} \doteq \sum_{h=1}^{H} \mu_{kh} \pi_h^{-1} \,.$$

The sample variance of the weights within cell *k* is then:

$$v_k = S_{2k} / n_k - (S_{1k} / n_k)^2 \doteq (\sum_{h=1}^{H} \mu_{kh} \pi_h^{-1}) / (\sum_h \pi_h \mu_{kh}) - (\sum_{h=1}^{H} \mu_{kh} / \sum_h \pi_h \mu_{kh})^2$$

and the squared coefficient of variation of the weights in cell *k* is:

$$(\sum_{h=1}^{H} \mu_{kh} \pi_h^{-1})(\sum_{h=1}^{H} \pi_h \mu_{kh}) / (\sum_{h=1}^{H} \mu_{kh})^2 - 1$$

which is identical to c_k^2 in (A6).

Acknowledgement

We are extremely grateful to Natalie Shlomo for programming SAS® to obtain the jackknife variance estimates in Table 5.

Bibliography

[1] Agresti, Alan. 2002. *Categorical Data Analysis*. Hoboken, NJ: John Wiley.
[2] Asparouhov, Tihomir. 2005. "Sampling Weights in Latent Variable Modeling." *Structural Equation Modeling* 12:411-34.
[3] Binder, David A. 1983. "On the Variances of Asymptotically Normal Estimators from Complex Surveys." *International Statistical Review* 51:279-92.
[4] Chambers, Roy L. and Chris J. Skinner. (eds.) 2003. *Analysis of Survey Data*. Chichester: John Wiley.
[5] Clogg, Clifford C. and Scott R. Eliason. 1987. "Some Common Problems in Log-Linear Analysis." *Sociological Methods & Research* 16:8-44.
[6] Gollac, Michel, Pierre Laulhé et Jeanine Soleilhavoup. 1988a. *Mobilité sociale. Enquête Formation-Qualification Professionnelle de 1985*. D 126, Paris, Insee.
[7] Gollac, Michel, Pierre Laulhé et Jeanine Soleilhavoup. 1988b. *Formation. Enquête Formation-Qualification Professionnelle de 1985*. D 129, Paris, Insee.

[8] Hauser, Robert M. 1978. “A Structural Model of the Mobility Table.” *Social Forces* 56:919-53.
[9] Hauser, Robert M. 1980. “Some Exploratory Methods for Modeling Mobility Tables and Other Cross-Classified Data.” *Sociological Methodology* 11:413-58.
[10] Hájek, Jaroslav. 1981. *Sampling from a Finite Population*. New York: Marcel Dekker.
[11] Korn, Edward L. and Barry I. Graubard. 1999. *Analysis of Health Surveys*. New York: John Wiley.
[12] Laulhé, Pierre et Jeanine Soleilhavoup. 1987. *Mobilité professionnelle. Enquête Formation-Qualification Professionnelle de 1985*. D 121, Paris, Insee.
[13] Patterson, Blossom H., C. Mitchell Dayton, and Barry I. Graubard. 2002. “Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data.” *Journal of the American Statistical Association* 97:721-29.
[14] Raghunathan, Trivellore E., Peter W. Solenberger, and John Van Hoewyk. 2002. *IVEware: Imputation and Variance Estimation Software. User Guide*. Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan.
(http://www.isr.umich.edu/src/smp/ive/)
[15] Rao, J. N. K. and D. Roland Thomas. 1988. “The Analysis of Cross-Classified Categorical Data From Complex Sample Surveys.” *Sociological Methodology* 18:213-69.
[16] Rust, Keith F. and J. N. K. Rao. 1996. “Variance Estimation for Complex Surveys using Replication Techniques.” *Statistical Methods in Medical Research* 5:283-310.
[17] Shao, Jun and Dongsheng Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.
[18] Skinner, Chris J. 1989. “Domain Means, Regression and Multivariate Analysis.” in *Analysis of Complex Surveys*, edited by C. J. Skinner, D. Holt, and T. M. F. Smith. Chichester: John Wiley.
[19] Skinner, Chris J., David Holt, and T. M. Fred Smith. (eds.) 1989. *Analysis of Complex Surveys*. Chichester: John Wiley.
[20] Vallet, Louis-André. 2005. “Utiliser le modèle log-linéaire pour mettre au jour la structure du lien entre les deux variables d’un tableau de contingence : un exemple d’application à la mobilité sociale.” *Actes des Journées de Méthodologie Statistique 2005*, Paris, Insee, 21 p.
(http://jms.insee.fr/site/files/documents/2006/410_1-JMS2005_SESSION06_VALLET_ACTES.PDF)
[21] Vermunt, Jeroen K. 2002. “Comment on Latent Class Analysis of Complex Sample Survey Data: Application to Dietary Data.” *Journal of the American Statistical Association* 97:736-7.
[22] Vermunt, Jeroen K. and Jay Magidson. 2007. “Latent Class Analysis with Sampling Weights: A Maximum-Likelihood Approach.” *Sociological Methods & Research* 36:87-111.