Chris J. Skinner and L.-A. Vallet
Fitting log-linear models to contingency tables from surveys with complex sampling designs: an investigation of the Clogg-Eliason approach

Article (Accepted version)
(Refereed)

Original citation:
DOI: 10.1177/0049124110366239

© 2010 SAGE Publications

This version available at: http://eprints.lse.ac.uk/39118/
Available in LSE Research Online: November 2011

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is the author’s final manuscript accepted version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this version and the published version may remain. You are advised to consult the publisher’s version if you wish to cite from it.
Fitting Log-linear Models to Contingency Tables from Surveys with Complex Sampling Designs: an Investigation of the Clogg-Eliason Approach

Chris Skinner
University of Southampton
United Kingdom

Louis-André Vallet
CNRS & CREST, Paris
France

Abstract: Clogg and Eliason (1987) proposed a simple method for taking account of survey weights when fitting log-linear models to contingency tables. 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 survey data from France. The method of Clogg and Eliason is found to underestimate standard errors systematically. 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 R (Survey Analysis), 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:

\[
\log(\mu_s) = X \lambda_s, \quad (1)
\]

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 \( \lambda_s \) is a \( p \times 1 \) vector of unknown parameters, where \( p \leq M \).

We subscript \( \mu_s \) and \( \lambda_s \) 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, \( \mu_s \) 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:

\[
\log(\mu) = X \lambda, \quad (2)
\]

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 \( \pi_k \) and let \( \log(\pi) \) be the \( M \times 1 \) vector containing the \( \log(\pi_k) \). Then we may write:

\[
\log(\mu_s) = \log(\pi) + \log(\mu), \quad (3)
\]

since the expected sample frequency in the \( k \)th cell is given by \( \pi_k \) times the expected population frequency. Hence from (2) and (3), we may write:
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 $\pi_i$ 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 $\lambda_i$ and $\lambda$ 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 $\lambda_i$ 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 $\lambda_i$ and $\lambda$ 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 $\mu_i$ and that the log-linear
model in (4) holds. Then, provided the inclusion probabilities $\pi_k$ are known and provided it is reasonable to make a standard ‘sampling model’ assumption, such as a Poisson or multinominal distribution, the parameter vector $\lambda$ 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 $\mu_k$. 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 $\pi_k$. 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 $\mu_{lk} = \pi_k \mu_k$. 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 \( \pi_k \) 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:

\[
\log(\mu_i) = \log(z) + X\lambda
\]

(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 \( \lambda \), 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 $\pi_1, \pi_2, \ldots, \pi_H$. 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 $\pi_h$ ($h = 1, 2, \ldots, H$). 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: 
\( N_{kh} \sim \text{Poisson}(\mu_{kh}) \). This implies that 
\( N_k = \sum_{h=1}^{H} \mu_{kh} \), and also that the numbers \( n_{kh} \) of sample units which fall into cell \( k \) and stratum \( h \) are independently distributed as:

\( n_{kh} \sim \text{Poisson}(\pi_{kh} \mu_{kh}) \). 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.

\[ n_k \sim \text{Poisson}(\mu_{sk}), \text{ where } \mu_{sk} = \sum_{h=1}^{H} \pi_{kh} \mu_{kh}. \] (6)

**4.2. Point Estimation under the CE Approach**

The parameter vector \( \lambda \) 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:

\[ \log L(\lambda) = \sum_k \{ n_k \log[\mu_k(\lambda)] - \mu_k(\lambda) \}, \] (7)

where, from (5), \( \mu_k(\lambda) = \exp(x_k \lambda) z_k \) 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 \( \lambda \) which maximizes (7).

We show in the Appendix that, providing the model in (2) holds, then \( \hat{\lambda}_{CE} \) is consistent for \( \lambda \) (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 \( \bar{J}^{-1} \), where \( \bar{J} = \sum_k \mu_a x_k \) and this matrix may be estimated by replacing the \( \mu_a \) 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:

\[
v(\hat{\lambda}_{CE}) = \bar{J}^{-1} + \bar{J}^{-1} \{ \sum_k c_k^2 \mu_a x_k \} \bar{J}^{-1},
\]

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 \( v(\hat{\lambda}_{CE}) = (1 + c^2) \bar{J}^{-1} \). 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 \( \mu_{kh} = \mu_h \phi_h \), where \( \phi_h \) 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 \( \sqrt{1 + c_k^2} \) where \( c \) is the sample coefficient of variation of the weights. We do not pursue this idea, however.

5. **Pseudo Maximum Likelihood (PML) Approach**

The PML approach is motivated by the ‘census’ likelihood for model (2) which would apply if the whole population was sampled and the \( N_k \) were known. In this case the ML estimator of \( \lambda \) would be obtained by solving the ‘census’ likelihood equations (c.f. Agresti, 2002, p.335) given by:

\[
\sum_k [N_k - \mu_k(\hat{\lambda})] v_k = 0,
\]

where \( \mu_k(\hat{\lambda}) = \exp(x_k \hat{\lambda}) \). The PML estimator of \( \lambda \), denoted \( \hat{\lambda}_{pml} \), is defined as the solution of (9) when the \( N_k \) are replaced by the weighted counts \( \hat{N}_k \). It may be obtained by using one of the standard ML fitting routines for log linear modelling (Agresti, 2002,
sect. 8.7) with the weighted counts. The variance-covariance matrix of $\hat{\lambda}_{pml}$ may be obtained by either linearization or replication methods.

The linearization method may be implemented by first taking the standard estimator of the variance-covariance matrix $v_0 = \{X'\text{diag}[\mu_k(\hat{\lambda}_{pml})]X\}^{-1}$ (Agresti, 2002, p.339), obtained from the ML fit to the weighted counts, as above. The linearization estimator may then (c.f. Rao and Thomas, 1988, sect. 5.2) be expressed as

$$v_L(\hat{\lambda}_{pml}) = v_0X'VXv_0,$$

where $V$ is an estimator of the variance-covariance matrix of the vector of $\hat{N}_k$, which takes account of the complex sampling. The estimator $v_L$ does not appear to be implemented in standard statistical software at present. (Although there is now much software implementing PML for generalized linear models (GLMs) and log-linear models can be represented as GLMs, this representation treats cell counts as observations and does not allow for complex sampling at the unit level.) Instead, $v_L$ can be computed by matrix multiplication after first calculating $V$ from survey software, such as Survey Analysis in R (Lumley, 2004). The $\hat{N}_k$ may first be scaled by a constant before calculating $\hat{\lambda}_{pml}$ and $v_0$ to avoid numerical problems (if the population is large) and to enable $v_0$ to be more interpretable. If the constant is set as $f = \sum_k n_k / \sum_k \hat{N}_k$, the weighted counts sum to the sample size and $v_0$ is interpretable as an approximate variance-covariance matrix ignoring the effects of complex sampling. Scaling by $f$ should not affect $\hat{\lambda}_{pml}$ nor $v_L(\hat{\lambda}_{pml})$ but should multiply $v_0$ by $f^{-1}$ and $V$ by $f^2$. 
Such matrix manipulations may be avoided by employing a replication method, such as the jackknife or bootstrap, where different sets of weights are constructed for each of a series of ‘replicates’. The point estimate \( \hat{\lambda} \) (which could be either \( \hat{\lambda}_{PML} \) or \( \hat{\lambda}_{CE} \)) is computed for each replicate and the estimated variance is obtained as a simple measure of the variability (across replicates) of these estimates. In the case of the jackknife and stratified multi-stage sampling, the replicates correspond to the primary sampling units (PSUs) \( j = 1, 2, \ldots, n_j \) within strata \( h = 1, 2, \ldots, H \). The set of weights \( w^{(h)}_j \) for replicate \( h_j \) is constructed by assigning zero weight to all sample units from PSU \( j \) in stratum \( h \), inflating the weights of other units in this stratum by the factor \( n_h/(n_h - 1) \) and leaving the remaining weights the same (Rust and Rao, 1996). For replicate \( h_j \), \( \hat{\lambda} \) is estimated just as \( \hat{\lambda} \) is computed except that the original survey weights are replaced by the replicate weights \( w^{(h)}_j \). The resulting estimator is denoted \( \hat{\lambda}^{(h)}_j \). This is repeated for each replicate and the jackknife estimator of the variance of \( \hat{\lambda} \) is then given by:

\[
v_j(\hat{\lambda}) = \frac{1}{n_h - 1} \left( \sum_{i=1}^{n_h} (\hat{\lambda}^{(h)}_i - \hat{\lambda}) \right)^2.
\]

This 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).

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, \( \hat{\lambda}_{\text{PML}} \) and \( \hat{\lambda}_{\text{CE}} \), for the two approaches are not identical (as noted
by Vermunt, 2002) but they are both consistent for the true value of \( \lambda \) if the model in (2)
holds. If the model in (2) does not hold then, as the sample size increases, \( \hat{\lambda}_{\text{PML}} \) and \( \hat{\lambda}_{\text{CE}} \)
will not in general converge to the same quantities. Whether the limiting value of either
\( \hat{\lambda}_{\text{PML}} \) or \( \hat{\lambda}_{\text{CE}} \) is of interest depends on the scientific objectives. One possible advantage of
\( \hat{\lambda}_{\text{PML}} \) 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 \( \hat{\lambda}_{\text{PML}} \) and \( \hat{\lambda}_{\text{CE}} \) under the assumption
that the model in (2) is correct. To do this we write both estimators as solutions of the
estimating equations:

\[
\sum_k a_k \{ \hat{N}_k - \mu_k(\lambda) \} x_k = 0, \tag{11}
\]

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

The variance of any linear combination of the elements of the vector \( \lambda \) solving (11)
is minimized by setting \( a_k x_k \) proportional to:

\[
\frac{1}{\var(N_k)} \frac{\partial \exp(x_k \lambda)}{\partial \lambda} = \frac{1}{\sum_{h=1}^{H} \pi_h \mu_{lh}} \mu_k x_k.
\]

Hence the optimal choice of \( a_k \) is

\[ \boxed{ } \]
\[ a_{\text{opt}} \propto \mu_k / \sum_{h=1}^{M} \pi_h^{-1} \mu_{kh} \]

and an estimate of the optimal \( a_k \) is \( \hat{a}_{\text{opt}} = S_{1k} / S_{2k} \), 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 \( a_{\text{opt}} \) is constant and both \( \hat{\lambda}_{\text{PML}} \) and \( \hat{\lambda}_{\text{CE}} \) are equal and optimal.

The CE point estimator \( \hat{\lambda}_{\text{CE}} \) 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 \( \hat{\lambda}_{\text{PML}} \) 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 \( \hat{\lambda}_{\text{CE}} \) to tend to be always more efficient than \( \hat{\lambda}_{\text{PML}} \) 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.

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 (see scaling by \( f \) in section 5).

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 \( \delta_k \),
the logged expected frequency in cell \( (i, j) \) of the mobility table is expressed as follows:

\[
\log \mu_{ij} = \alpha + \beta_i + \gamma_j + \delta_k \quad \text{if the cell (}i,j\text{) belongs to subset } k.
\]

Thus, aside from total (\( \alpha \)), row (\( \beta_i \)), and column (\( \gamma_j \)) effects, each expected
frequency is determined by only one interaction parameter (\( \delta_k \)) 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).

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. 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 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 \( \gamma_j \) 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.

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. It should be noted
that the standard errors in Tables 4 and 5 are only sample estimates. However, it seems
quite implausible that the systematic patterns observed are a result of sampling variation
when the sample size is over 5,000 and the patterns are so similar for the different
parameters.

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. This has been shown theoretically in the case of
unequal probability Poisson sampling and empirically in the case of disproportionate stratified sampling. 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 does not appear to be implemented currently in log-linear modelling procedures in standard software packages, it is often feasible to employ replication variance estimation methods, such as the jackknife or 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:

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

which may also be expressed as:

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

Consider the set-up of section 4.1, where $\hat{N}_k = \sum_{i=1}^H n_{ik} x_{ik}^{-1}$, 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 $\exp(x_k \hat{\lambda}) = \mu_k$, then $\hat{N}_k / \exp(x_k \hat{\lambda})$ 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 $\lambda$.

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

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

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

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

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

$$J(\hat{\lambda}) \approx \hat{J}(\hat{\lambda}) = \sum_k \hat{\mu}_k x_k ' x_k .$$  \hspace{1cm} (A3)
The actual variance-covariance matrix of $\hat{\lambda}_{CE}$ may be obtained by linearization as follows. The first order Taylor expansion of $\exp(x_i\hat{\lambda}_{CE})$ around $\hat{\lambda}_{CE} = \lambda$ is:

$$\exp(x_i\hat{\lambda}_{CE}) \approx \exp(x_i\lambda) + \exp(x_i\lambda)x_i(\hat{\lambda}_{CE} - \lambda).$$

Substituting into (A1) gives:

$$\sum_k [\hat{\lambda}_k - \exp(x_k\lambda) - \exp(x_k\lambda)x_k(\hat{\lambda}_{CE} - \lambda)]z_kx_k' \approx 0$$

or

$$\hat{\lambda}_{CE} \approx \lambda + \sum_k [\hat{\lambda}_k - \exp(x_k\lambda)]z_kx_k' = \lambda + J(\lambda)^{-1}\sum_k [\hat{\lambda}_k - \exp(x_k\lambda)]z_kx_k'.$$

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

$$\text{var}\{J(\lambda)^{-1}\sum_k [\hat{\lambda}_k - \exp(x_k\lambda)]z_kx_k'\}. \quad \text{(A4)}$$

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

$$\tilde{J}(\lambda)^{-1} \text{var}\{\sum_k (\hat{\lambda}_k - \mu_k)(\mu_{ik}/\mu_k)x_k'\} \tilde{J}(\lambda)^{-1}$$

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

Now

$$\text{var}(\hat{\lambda}_k) = \sum_{h=1}^H \pi_{ik}^{-2} \text{var}(n_{ik}) = \sum_{h=1}^H \pi_{ik}^{-1} \mu_{ik}.$$

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

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

where $c_k^2 = [\sum_{h=1}^H \pi_{ik} \mu_{ih}/\sum_{h=1}^H \pi_{ik} \mu_{ih} - (\sum_{h=1}^H \mu_{ih})^2]/\mu_k^2 = [\sum_{h=1}^H \pi_{ik} \mu_{ih}/\sum_{h=1}^H \mu_{ih}]/\mu_k^2 - 1. \quad \text{(A6)}$
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 \( \pi_h \mu_{kh} \approx \pi_h^{-1} \mu_{kh} \) which requires that the \( \pi_h \) are constant, i.e. \( \pi_h = \pi \). To show that \( c_k \) is the coefficient of variation of the weights in cell \( k \), let
\[
S_k = \hat{N}_k = \sum_{h=1}^{H} n_{kh} \pi_h^{-1} \approx \sum_{h=1}^{H} \mu_{kh} = \mu_k \quad \text{and} \quad S_{2k} = \sum_{h=1}^{H} n_{kh} \pi_h^{-2} \approx \sum_{h=1}^{H} \mu_{kh} \pi_h^{-1}.
\]
The sample variance of the weights within cell \( k \) is then:
\[
\nu_k = S_{2k} / n_k - (S_k / n_k)^2 \approx \left( \sum_{h=1}^{H} \mu_{kh} \pi_h^{-1} \right) \left( \sum_{h=1}^{H} \pi_h \mu_{kh} \right) - \left( \sum_{h=1}^{H} \mu_{kh} / \sum_{h=1}^{H} \pi_h \mu_{kh} \right)^2
\]
and the squared coefficient of variation of the weights in cell \( k \) is:
\[
\left( \sum_{h=1}^{H} \mu_{kh} \pi_h^{-1} \right) \left( \sum_{h=1}^{H} \pi_h \mu_{kh} \right) / \left( \sum_{h=1}^{H} \mu_{kh} \right)^2 - 1
\]
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.
References


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

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Sample size</th>
<th>Initial sampling fraction</th>
<th>Mean of final weight</th>
<th>Standard deviation of final weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>French women, in the labour market, farmers, aged 32 – 51</td>
<td>234</td>
<td>1/940</td>
<td>959.85</td>
<td>123.67</td>
</tr>
<tr>
<td>French women, in the labour market, farmers, aged 52+</td>
<td>83</td>
<td>1/1250</td>
<td>1245.55</td>
<td>47.53</td>
</tr>
<tr>
<td>French women, in the labour market, artisans and shopkeepers, 32 – 51</td>
<td>223</td>
<td>1/1040</td>
<td>1144.54</td>
<td>87.61</td>
</tr>
<tr>
<td>French women, in the labour market, company managers and higher-grade professionals, 32 – 51</td>
<td>747</td>
<td>1/1360</td>
<td>1487.79</td>
<td>120.05</td>
</tr>
<tr>
<td>French women, in the labour market, company managers and higher-grade professionals, 52+</td>
<td>94</td>
<td>1/340</td>
<td>388.76</td>
<td>67.22</td>
</tr>
<tr>
<td>French women, in the labour market, lower-grade professionals, 52+</td>
<td>1064</td>
<td>1/600</td>
<td>669.29</td>
<td>111.77</td>
</tr>
<tr>
<td>Foreign women, in the labour market, employed or unemployed, 32 – 51</td>
<td>10</td>
<td>1/400</td>
<td>1000.14</td>
<td>110.70</td>
</tr>
<tr>
<td>French women, students</td>
<td>2</td>
<td>1/2270</td>
<td>2464.00</td>
<td>247.49</td>
</tr>
<tr>
<td>Other French women, out of the labour market, 32 – 51</td>
<td>146</td>
<td>1/2500</td>
<td>2794.81</td>
<td>620.64</td>
</tr>
<tr>
<td>Other French women, out of the labour market, 52+</td>
<td>17</td>
<td>1/2500</td>
<td>2794.76</td>
<td>389.26</td>
</tr>
<tr>
<td>Foreign women, in the labour market, employed or unemployed, 32 – 51</td>
<td>10</td>
<td>1/730</td>
<td>831.20</td>
<td>189.88</td>
</tr>
<tr>
<td>Total</td>
<td>5159</td>
<td></td>
<td>850.34</td>
<td>451.33</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Father’s class</th>
<th>Frequency</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Higher-grade salaried</td>
<td>Unweighted</td>
<td>164.00</td>
<td>25.00</td>
<td>136.00</td>
<td>12.00</td>
<td>59.00</td>
<td>9.00</td>
<td>0.00</td>
<td>405.00</td>
</tr>
<tr>
<td>professionals</td>
<td>Weighted</td>
<td>81.23</td>
<td>13.01</td>
<td>113.18</td>
<td>15.35</td>
<td>66.32</td>
<td>8.08</td>
<td>0.00</td>
<td>297.17</td>
</tr>
<tr>
<td>2 Company managers and</td>
<td>Unweighted</td>
<td>56.00</td>
<td>27.00</td>
<td>37.00</td>
<td>14.00</td>
<td>28.00</td>
<td>3.00</td>
<td>3.00</td>
<td>168.00</td>
</tr>
<tr>
<td>liberal professions</td>
<td>Weighted</td>
<td>28.78</td>
<td>11.72</td>
<td>38.22</td>
<td>14.46</td>
<td>32.45</td>
<td>2.65</td>
<td>7.01</td>
<td>135.29</td>
</tr>
<tr>
<td>3 Lower-grade salaried</td>
<td>Unweighted</td>
<td>95.00</td>
<td>16.00</td>
<td>161.00</td>
<td>15.00</td>
<td>115.00</td>
<td>18.00</td>
<td>4.00</td>
<td>424.00</td>
</tr>
<tr>
<td>professionals</td>
<td>Weighted</td>
<td>48.08</td>
<td>11.44</td>
<td>129.70</td>
<td>22.79</td>
<td>131.79</td>
<td>18.20</td>
<td>4.77</td>
<td>366.78</td>
</tr>
<tr>
<td>4 Artisans and shopkeepers</td>
<td>Unweighted</td>
<td>97.00</td>
<td>35.00</td>
<td>219.00</td>
<td>78.00</td>
<td>200.00</td>
<td>35.00</td>
<td>8.00</td>
<td>672.00</td>
</tr>
<tr>
<td></td>
<td>Weighted</td>
<td>52.25</td>
<td>21.35</td>
<td>174.45</td>
<td>118.41</td>
<td>223.37</td>
<td>39.57</td>
<td>14.27</td>
<td>643.67</td>
</tr>
<tr>
<td>5 Non-manual workers</td>
<td>Unweighted</td>
<td>59.00</td>
<td>7.00</td>
<td>145.00</td>
<td>32.00</td>
<td>182.00</td>
<td>29.00</td>
<td>3.00</td>
<td>457.00</td>
</tr>
<tr>
<td></td>
<td>Weighted</td>
<td>30.18</td>
<td>3.68</td>
<td>120.03</td>
<td>53.42</td>
<td>216.57</td>
<td>28.65</td>
<td>4.17</td>
<td>456.70</td>
</tr>
<tr>
<td>6 Foremen and manual workers</td>
<td>Unweighted</td>
<td>128.00</td>
<td>18.00</td>
<td>419.00</td>
<td>124.00</td>
<td>930.00</td>
<td>339.00</td>
<td>37.00</td>
<td>1995.00</td>
</tr>
<tr>
<td></td>
<td>Weighted</td>
<td>64.18</td>
<td>14.88</td>
<td>361.46</td>
<td>184.12</td>
<td>1065.19</td>
<td>355.76</td>
<td>47.06</td>
<td>2092.66</td>
</tr>
<tr>
<td>7 Farmers</td>
<td>Unweighted</td>
<td>38.00</td>
<td>8.00</td>
<td>164.00</td>
<td>73.00</td>
<td>342.00</td>
<td>136.00</td>
<td>27.00</td>
<td>1038.00</td>
</tr>
<tr>
<td></td>
<td>Weighted</td>
<td>20.29</td>
<td>5.63</td>
<td>134.71</td>
<td>101.98</td>
<td>394.83</td>
<td>140.49</td>
<td>36.80</td>
<td>1166.73</td>
</tr>
<tr>
<td>Total</td>
<td>Unweighted</td>
<td>637.00</td>
<td>136.00</td>
<td>1281.00</td>
<td>348.00</td>
<td>1856.00</td>
<td>569.00</td>
<td>332.00</td>
<td>5159.00</td>
</tr>
<tr>
<td></td>
<td>Weighted</td>
<td>324.99</td>
<td>81.71</td>
<td>1071.75</td>
<td>510.54</td>
<td>2130.52</td>
<td>593.40</td>
<td>446.08</td>
<td>5159.00</td>
</tr>
</tbody>
</table>

Note: Weighted frequencies are rescaled to the sample size by multiplying them by the ratio 5159/4386881.
### Table 3 – Initial model and final model for the structure of the association in the mobility table

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – Higher-grade salaried professionals</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
<td>VII</td>
</tr>
<tr>
<td>2 – Company managers and liberal professions</td>
<td>III</td>
<td>II</td>
<td>IV</td>
<td>IV</td>
<td>VI</td>
<td>VII</td>
<td>VII</td>
</tr>
<tr>
<td>3 – Lower-grade salaried professionals</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
<td>VII</td>
</tr>
<tr>
<td>4 – Artisans and shopkeepers</td>
<td>V</td>
<td>IV</td>
<td>V</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
</tr>
<tr>
<td>5 – Non-manual workers</td>
<td>VI</td>
<td>VI</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>VI</td>
</tr>
<tr>
<td>6 – Foremen and manual workers</td>
<td>VII</td>
<td>VII</td>
<td>VI</td>
<td>VI</td>
<td>V</td>
<td>IV</td>
<td>V</td>
</tr>
<tr>
<td>7 – Farmers</td>
<td>VII</td>
<td>VII</td>
<td>VII</td>
<td>VI</td>
<td>VI</td>
<td>V</td>
<td>I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 – Higher-grade salaried professionals</td>
<td>II</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
<td>VII</td>
</tr>
<tr>
<td>2 – Company managers and liberal professions</td>
<td>II</td>
<td>II</td>
<td>III</td>
<td>III</td>
<td>V</td>
<td>VI</td>
<td>IV</td>
</tr>
<tr>
<td>3 – Lower-grade salaried professionals</td>
<td>III</td>
<td>III</td>
<td>III</td>
<td>IV</td>
<td>IV</td>
<td>V</td>
<td>VI</td>
</tr>
<tr>
<td>4 – Artisans and shopkeepers</td>
<td>IV</td>
<td>III</td>
<td>IV</td>
<td>III</td>
<td>V</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>5 – Non-manual workers</td>
<td>V</td>
<td>V</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>6 – Foremen and manual workers</td>
<td>VI</td>
<td>VI</td>
<td>V</td>
<td>IV</td>
<td>IV</td>
<td>III</td>
<td>IV</td>
</tr>
<tr>
<td>7 – Farmers</td>
<td>VII</td>
<td>VI</td>
<td>VI</td>
<td>IV</td>
<td>V</td>
<td>IV</td>
<td>I</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial model</th>
<th></th>
<th></th>
<th></th>
<th>Final model</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unweighted</td>
<td>Weighted</td>
<td>Clogg &amp; Eliaison</td>
<td>Pseudo maximum likelihood</td>
<td>Unweighted</td>
<td>Weighted</td>
<td>Clogg &amp; Eliaison</td>
<td>Pseudo maximum likelihood</td>
</tr>
<tr>
<td>$\beta_1$ (se)</td>
<td>-1.813 (0.087)</td>
<td>-1.825 (0.086)</td>
<td>-1.828 (0.086)</td>
<td>-1.825 (0.090)</td>
<td>-1.747 (0.084)</td>
<td>-1.754 (0.083)</td>
<td>-1.763 (0.083)</td>
<td>-1.754 (0.093)</td>
</tr>
<tr>
<td>$\beta_2$ (se)</td>
<td>-2.626 (0.107)</td>
<td>-2.621 (0.108)</td>
<td>-2.612 (0.106)</td>
<td>-2.621 (0.133)</td>
<td>-2.663 (0.102)</td>
<td>-2.610 (0.105)</td>
<td>-2.632 (0.102)</td>
<td>-2.610 (0.125)</td>
</tr>
<tr>
<td>$\beta_3$ (se)</td>
<td>-1.532 (0.079)</td>
<td>-1.559 (0.078)</td>
<td>-1.549 (0.079)</td>
<td>-1.559 (0.090)</td>
<td>-1.492 (0.076)</td>
<td>-1.517 (0.075)</td>
<td>-1.514 (0.076)</td>
<td>-1.517 (0.085)</td>
</tr>
<tr>
<td>$\beta_4$ (se)</td>
<td>-0.856 (0.069)</td>
<td>-0.857 (0.067)</td>
<td>-0.855 (0.070)</td>
<td>-0.857 (0.079)</td>
<td>-0.633 (0.061)</td>
<td>-0.614 (0.059)</td>
<td>-0.643 (0.061)</td>
<td>-0.614 (0.068)</td>
</tr>
<tr>
<td>$\beta_5$ (se)</td>
<td>-1.134 (0.072)</td>
<td>-1.104 (0.072)</td>
<td>-1.111 (0.073)</td>
<td>-1.104 (0.082)</td>
<td>-1.036 (0.067)</td>
<td>-1.013 (0.065)</td>
<td>-1.021 (0.067)</td>
<td>-1.013 (0.075)</td>
</tr>
<tr>
<td>$\beta_6$ (se)</td>
<td>0.492 (0.049)</td>
<td>0.510 (0.049)</td>
<td>0.505 (0.049)</td>
<td>0.510 (0.056)</td>
<td>0.487 (0.048)</td>
<td>0.497 (0.047)</td>
<td>0.497 (0.048)</td>
<td>0.507 (0.054)</td>
</tr>
<tr>
<td>$\gamma_1$ (se)</td>
<td>2.187 (0.149)</td>
<td>1.179 (0.139)</td>
<td>1.261 (0.149)</td>
<td>1.179 (0.166)</td>
<td>2.177 (0.148)</td>
<td>1.196 (0.138)</td>
<td>1.238 (0.148)</td>
<td>1.196 (0.157)</td>
</tr>
<tr>
<td>$\gamma_2$ (se)</td>
<td>0.585 (0.169)</td>
<td>-0.269 (0.169)</td>
<td>-0.182 (0.170)</td>
<td>-0.269 (0.205)</td>
<td>0.450 (0.167)</td>
<td>-0.373 (0.167)</td>
<td>-0.321 (0.167)</td>
<td>-0.373 (0.198)</td>
</tr>
<tr>
<td>$\gamma_3$ (se)</td>
<td>2.889 (0.140)</td>
<td>2.360 (0.120)</td>
<td>2.424 (0.140)</td>
<td>2.360 (0.150)</td>
<td>2.855 (0.139)</td>
<td>2.341 (0.119)</td>
<td>2.376 (0.139)</td>
<td>2.341 (0.146)</td>
</tr>
<tr>
<td>$\gamma_4$ (se)</td>
<td>1.473 (0.147)</td>
<td>1.508 (0.124)</td>
<td>1.555 (0.148)</td>
<td>1.508 (0.156)</td>
<td>1.204 (0.147)</td>
<td>1.253 (0.124)</td>
<td>1.282 (0.147)</td>
<td>1.253 (0.153)</td>
</tr>
<tr>
<td>$\gamma_5$ (se)</td>
<td>3.089 (0.137)</td>
<td>2.895 (0.116)</td>
<td>2.943 (0.137)</td>
<td>2.895 (0.143)</td>
<td>3.167 (0.137)</td>
<td>2.971 (0.116)</td>
<td>3.003 (0.137)</td>
<td>2.971 (0.144)</td>
</tr>
<tr>
<td>$\gamma_6$ (se)</td>
<td>1.605 (0.146)</td>
<td>1.297 (0.126)</td>
<td>1.349 (0.146)</td>
<td>1.297 (0.150)</td>
<td>1.638 (0.146)</td>
<td>1.340 (0.126)</td>
<td>1.370 (0.146)</td>
<td>1.340 (0.150)</td>
</tr>
<tr>
<td>$\delta_1$ (se)</td>
<td>3.561 (0.163)</td>
<td>3.451 (0.146)</td>
<td>3.569 (0.163)</td>
<td>3.451 (0.189)</td>
<td>4.163 (0.228)</td>
<td>4.096 (0.226)</td>
<td>4.138 (0.228)</td>
<td>4.096 (0.252)</td>
</tr>
<tr>
<td>$\delta_2$ (se)</td>
<td>2.730 (0.119)</td>
<td>2.619 (0.147)</td>
<td>2.660 (0.147)</td>
<td>2.619 (0.185)</td>
<td>3.215 (0.191)</td>
<td>3.104 (0.251)</td>
<td>3.123 (0.191)</td>
<td>3.123 (0.214)</td>
</tr>
<tr>
<td>$\delta_3$ (se)</td>
<td>2.396 (0.150)</td>
<td>2.297 (0.189)</td>
<td>2.326 (0.189)</td>
<td>2.297 (0.186)</td>
<td>2.276 (0.187)</td>
<td>2.252 (0.245)</td>
<td>2.275 (0.187)</td>
<td>2.275 (0.208)</td>
</tr>
<tr>
<td>$\delta_4$ (se)</td>
<td>1.683 (0.086)</td>
<td>1.633 (0.093)</td>
<td>1.700 (0.085)</td>
<td>1.633 (0.105)</td>
<td>1.692 (0.183)</td>
<td>1.658 (0.243)</td>
<td>1.675 (0.183)</td>
<td>1.675 (0.204)</td>
</tr>
<tr>
<td>$\delta_5$ (se)</td>
<td>1.161 (0.084)</td>
<td>1.078 (0.092)</td>
<td>1.154 (0.084)</td>
<td>1.078 (0.103)</td>
<td>1.245 (0.181)</td>
<td>1.217 (0.241)</td>
<td>1.240 (0.181)</td>
<td>1.240 (0.201)</td>
</tr>
<tr>
<td>$\delta_6$ (se)</td>
<td>0.683 (0.072)</td>
<td>0.641 (0.072)</td>
<td>0.699 (0.072)</td>
<td>0.641 (0.087)</td>
<td>0.731 (0.177)</td>
<td>0.708 (0.239)</td>
<td>0.702 (0.177)</td>
<td>0.702 (0.196)</td>
</tr>
<tr>
<td>$\delta_{VI}$</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
</tr>
<tr>
<td>Deviance</td>
<td>86.11 (0.084)</td>
<td>77.12 (0.083)</td>
<td>75.58 (0.083)</td>
<td>-</td>
<td>47.71 (0.084)</td>
<td>33.69 (0.083)</td>
<td>34.77 (0.083)</td>
<td>-</td>
</tr>
<tr>
<td>DF</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>-</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 5 – Comparison of estimated standard errors for Clogg-Eliason estimator: Clogg-Eliason approach vs Jackknife method allowing for complex design

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial model</th>
<th>Final model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clogg &amp; Eliason</td>
<td>Jackknife</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.086</td>
<td>0.102</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.106</td>
<td>0.130</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.079</td>
<td>0.090</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.070</td>
<td>0.078</td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.073</td>
<td>0.081</td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.049</td>
<td>0.055</td>
</tr>
<tr>
<td>$\beta_7$</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.149</td>
<td>0.158</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.170</td>
<td>0.204</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.140</td>
<td>0.144</td>
</tr>
<tr>
<td>$\gamma_4$</td>
<td>0.148</td>
<td>0.152</td>
</tr>
<tr>
<td>$\gamma_5$</td>
<td>0.137</td>
<td>0.140</td>
</tr>
<tr>
<td>$\gamma_6$</td>
<td>0.146</td>
<td>0.147</td>
</tr>
<tr>
<td>$\gamma_7$</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
</tr>
<tr>
<td>$\delta_I$</td>
<td>0.163</td>
<td>0.181</td>
</tr>
<tr>
<td>$\delta_{II}$</td>
<td>0.118</td>
<td>0.139</td>
</tr>
<tr>
<td>$\delta_{III}$</td>
<td>0.149</td>
<td>0.192</td>
</tr>
<tr>
<td>$\delta_{IV}$</td>
<td>0.085</td>
<td>0.101</td>
</tr>
<tr>
<td>$\delta_V$</td>
<td>0.084</td>
<td>0.099</td>
</tr>
<tr>
<td>$\delta_{VI}$</td>
<td>0.072</td>
<td>0.084</td>
</tr>
<tr>
<td>$\delta_{VII}$</td>
<td>Fixed at 0</td>
<td>Fixed at 0</td>
</tr>
</tbody>
</table>
**Chris Skinner** is Leverhulme Professor of Social Statistics at the University of Southampton, U.K. His main research interests concern statistical aspects of survey methodology, including methods for the analysis of complex survey data. He has co-edited *Analysis of Survey Data* (Wiley, 2003 with R. Chambers) and *Analysis of Complex Surveys* (Wiley, 1989, with D.Holt and T.Smith).

**Louis-André Vallet** is research professor in the French National Centre for Scientific Research (CNRS) and works within the Quantitative Sociology Unit in the Centre for Research in Economics and Statistics (CREST), the research centre of the French Statistical Office. His main research interests and publications are in the areas of social stratification and mobility, sociology of education, and statistical modelling of categorical variables.