# VARIANCE ESTIMATION UNDER MARGINAL MODELS FOR LONGITUDINAL DATA ANALYSIS USING COMPLEX SURVEY DATA

Ioana Ôchiopu-Kratina
Statistics Canada

Key Words: Estimating Equations, Marginal models, Design inference, Jackknife

**1. Introduction**. In longitudinal surveys subjects are observed on at least two different occasions, which makes such surveys suitable for studying change over time at the individual, or unit level. In addition to the production of crossectional estimates, data from longitudinal surveys may be used, for instance, to estimate gross flows (important in the study of labour market dynamics), or in event history modelling, which may be used to uncover determinants of survival for individuals afflicted with a serious health condition. More generally, longitudinal data may be used for modelling a response variable as a function of covariates and time, with applicability in many areas. Rao (1998) and the references therein give a more complete description of possible use of data from longitudinal surveys and the statistical techniques that are available to explore them. The task of sponsoring large scale longitudinal surveys over long periods of time is often undertaken by large organizations like Statistics Canada. Their primary goal in conducting a survey is to obtain design - based estimates of totals, means or proportions for a target population, which is finite. The selection of the sample generally follows a complex plan with goals like reducing the design variability of the sample estimates. The conditions for model based inference are often not met by the data collected according to the survey design, even if the finite population is large. Design - based inference, introduced by Binder (1983), offers a solution, as it allows for the use of modelling techniques in the context of survey randomization. We follow this approach, which is also that of Rao (1998), and consider marginal models for longitudinal data as in Liang and Zeger (1986) in the context of design - based inference. In longitudinal data, observations on the same subject are dependent, and this dependence is different from the clustering effect due to the sampling selection. Liang and Zeger (1986) introduced Generalized Estimating Equations (GEE), which require only specification of the marginal model mean and variance for each individual. Correlation across time for the same individual is assumed to exists, but it is not specifically modelled. Section 2 of this article presents marginal models as in Liang and Zeger (1986). Example 1 illustrates the classical use of estimating equations (EE) in calculating an estimator of a regression coefficient for the linear model. This estimator becomes the census parameter in the context of design - based inference, which is outlined in Section 3. The design that we consider is stratified, multistage and with replacement at the first stage. Example 2 shows the calculation of the design based estimator from the 'weighted' EE in Example 1. Section 4 is devoted to design - consistency. This topic seems to have been neglected in the literature - it is not even mentioned in Liang and Zeger (1986). Yet consistency is an essential ingredient in the proof of asymptotic normality of estimators that are implicitly defined by EE. Example 3 illustrates conditions for consistency on the EE in Example 2. Example 4 is a genuine example of GEE on which we illustrate the conditions for consistency. We give conditions needed for asymptotic normality in the presence of nuisance parameters in Example 5. Section 5 deals with asymptotic normality and the consistency of the jackknife estimator of the asymptotic variance. Complete proofs of all statements made in this article will be presented elsewhere.

**2. Model set-up.** We describe briefly the set-up in Liang and Zeger (1986). Consider M individuals observed on $d_i$ occasions ( i = 1, ... M). The responses $y_{i\,t}$ and the p - covariates $x_{i\,t}$ are recorded, t = 1, ... $d_i$ , i = 1,... M. We assume that $d_i$ = d , i = 1, ... M. Typically, d is small for marginal models. Otherwise, time series techniques may be more appropriate. Here only $E_m\,[y_{i\,t}]$ and $Var_m\,(y_{i\,t})$ are specified, where *m* stands for model- based, for all (œ) t, i \$ 1. Liang and Zeger (1986) consider probability densities of the following type: $D\,(y_{i\,t}) = \exp\,[\{y_{i\,t}\,2_{i\,t} - a\,(2_{i\,t}) + b\,(y_{i\,t})\}N]$ with $2_{i\,t}$ = $h(0_{i\,t})$, $0_{i\,t} = x_{i\,t}^{T}\$$, where a, b and h are known (differentiable) functions, $\{2_{i\,t}\}$, N are parameters , $x_{i\,t}^{T}$ is an 1 × p matrix of covariates and \$ is an p × l vector of main parameters, œ t, i\$ 1. Here T stands for transposition of matrices. Note that for random variables with such densities we have:

$$E_m\,[y_{i\,t}] = \mu_{i\,t} = a'\,(2_{i\,t}),\ œ\ t,\ i\$\ 1: \tag{1}$$

Let $\mu(0) = a'(0)$ , œ 0 in a space of parameters 1. The function g is a link function if $gB\mu(2_{i\,t})$ = $x_{i\,t}^{T}\,\$$, œ t, i\$ 1. If g $= \mu^{-1}$ , then g is called the canonical, or natural link, the function *h* above can be taken to be the identity and the parametric form of the model is the natural one. With binary response, the logit link function g(μ) = log{μ/(1-μ)}is the natural link associated with the logistic regression model. EE's are formed that mimic score functions associated with exponential distributions (e.g. normal, binomial, logistic, Poisson). The idea is to make few assumptions on the distribution of the observed data, produce estimators for \$which are roots of likelihood equations (RLE), when they exist, and study their properties. When GEE's are used, it is assumed that correlation of observations $y_{i\,t}$ across time for the same individual is the same for all individuals, and is represented by a matrix R("), with " a "nuisance parameter". More precisely, let $U_i\,(\$, ", n)$ = $D_i^{T}\,V_i^{-1}\,S_i$ , $V_i = 1/n\,[A_i^{1/2}\,R\,(")A_i^{1/2}]$ , $D_i = A_i\,)_i\,X_i$ , $S_i = Y_i - a'(2_i)$, $A_i$ = diag $a''(2_{i\,t})$ in $R^{d}$ and $)_i$ = diag [d $2_{i\,t}$ / d $0_{i\,t}$] , which could be taken to be the identity matrix $I_d$ , œ i \$ 1. Notice that the covariates are contained in $D_i$ and that $A_i$ as well as $S_i$ (through a') contain the main parameter \$, i \$ 1. The GEE, or equation (7) of Liang and Zeger (1986), is:

$$\sum_{i=1}^{M} U_i\,(\$,\ \hat{\alpha}(\$),\hat{N}(\$)) = 0 \tag{2}$$

Equation (2) above is called a pseudo-likelihood equation in Shao (1999), p. 315. Note that it consists of p scalar equations. We denote by $u_{iR}\,(s,\$\,)$ the R th component of $U_i$ above, R = 1,..p , i = 1,... M. In equation (2) • and $\hat{N}$ are estimates of nuisance parameters that are obtained from the sample and generally contain \$. When the solution to (2) exists and is unique , i.e. when \$ is defined implicitly by (2), it is denoted by $\hat{\$}_G$ in Liang and Zeger (1986). Note that this approach is different from the one presented in Section 5 of Rao (1998). It is important to note that (2) contains only \$ as unknown parameter and that, due to the estimation of the nuisance parameters, the left hand side of (2) is, in general, a nonlinear function of the sample observations. In the special situation when the observations across time are assumed independent for each individual (the working independence assumption), equation (2) becomes the Independence Estimating Equation (IEE). In this case R(") = $I_d$ and there is no need to estimate nuisance parameters in (2). This is the situation discussed, in a design randomization context, by D. Binder (1983). In the context of IEE and survey randomization (see Section 3), $\hat{\$}_G$ becomes the "census" parameter defined in Binder (1983). The example below illustrates the calculation of $\hat{\$}_G$ from an IEE. Notice that the presence of the time dimension is accounted for by the increase in the number of data points (from M to 2×M in this case).

**Example 1** Assume that the individual observations are independent, identically distributed (i.i.d.) and that they follow a normal distribution. Take N = 1 and d = 2 occasions. We have R(") =

$I_2$ (case IEE). Assume that $x_{it}$, $\beta$ are scalars, $i,t \geq 1$ and that *h* is the identity.

$$\Delta(y_{it}) = \exp -\frac{(y_{it}-\theta_{it})^2}{2} = \exp\{y_{it}\,\theta_{it} - a(\theta_{it}) + b(y_{it})\} \Rightarrow a(\theta_{it}) = \frac{\theta_{it}^2}{2},\ b(y_{it}) = -\frac{y_{it}^2}{2}$$

$$E[y_{it}] = \theta_{it} = \frac{d\,a}{d\,\theta_{it}};\ \frac{d^2 a}{d\theta_{it}^2} = 1,\ \theta_{it} = x_{it}\beta,\ i,t \geq 1.$$

Note that each $x_{it}$ has as many components as $\beta$ (p = 1 components here) and, for $i, t \geq 1$:

$$\frac{d \log \Delta(y_{it})}{d\beta} = y_{it}\, x_{it} - x_{it}^2\beta$$

Now $a'(\theta_{it}) = \theta_{it} = x_{it}\beta$, $i, t \geq 1$ and (2) is:

$$(3) \qquad \sum_{i=1}^{M}\sum_{j=1}^{2} x_{ij}\, y_{ij} - \sum_{i=1}^{M}\sum_{j=1}^{2} x_{ij}^2\,\beta = 0 \Rightarrow \hat{\beta}_G = \frac{\sum_{i=1}^{M}\sum_{j=1}^{2} x_{ij}\, y_{ij}}{\sum_{i=1}^{M}\sum_{j=1}^{2} x_{ij}^2}$$

**3. The design and the design-based inference**. In the article, inference is done in the design - based randomization as proposed by D. Binder (1983). As mentioned in his paper, conclusions can be drawn only in designs in which conditions have been given for the Central Limit Theorem (CLT) to hold. The design that we consider here is stratified, multistage in which the p.s.u.'s (clusters) are selected with replacement from a population of M individuals (or 'ultimate' selection units). Conditions for the CLT to hold in such designs have been given by Krewski and Rao (1981) and by Yung (1996). Here the cluster totals (or normalized cluster totals) are i.i.d.'s in the design randomization within each stratum and independent random variables (r.v.'s) across strata. Thus, the r.v.'s involved in the limiting theorems are the clusters rather than the individuals. The populations change with the increase in the number of units involved in the inference. The sampling distributions of these variables change with the changing populations and so does the finite population parameter. It is therefore appropriate to consider CLT's for arrays. To simplify notation, we index the populations by the total number of associated r.v.'s involved in the limiting process, i.e. the total number of clusters N from which *n* p.s.u.'s are selected. Thus, the census parameter defined by (3) for the IEE case will be denoted $\beta_N = \hat{\beta}_G$, rather than $\beta_M$, which would be more appropriate. The parameter to estimate in the design randomization context changes as $n \to \infty$ (which implies that N, M $\to \infty$). The classical proofs that derive the distribution of estimators of $\beta_0$ rely on asymptotic distribution of sums which depend on a fixed value of $\beta$. To apply these results to the finite population context one must require that the functions that make up the EE be equicontinuous in $\beta$ (see condition E below). In this article, the parameter $\beta_0$ plays the role of the fixed point in the asymptotic, e.g.: $\hat{\beta}_N \xrightarrow{p_N} \beta_0$, where $\xrightarrow{p_N}$ means convergence in the design probability, which is consistent with Binder (1983). In some instances, one might wish to link $\beta_0$ to the superpopulation parameter, e.g. if one wishes to give an interpretation to the finite population parameter. We do not attempt to do this here.

For simplicity, we consider that the selected sample *s* consists of respondents only. The generalization to the situation where nonresponse occurs completely at random is straightforward (see J.N.K. Rao, 1998). Consider a population that consists of M individuals and which is partitioned into L strata. Each stratum consists of $M_h$ individuals from which $N_h$ clusters are formed, h =1,...L. From each stratum *h*, $n_h$ clusters are selected with replacement and a further selection of $m_{hi}$ individuals takes place within each cluster i, i = 1,... $n_h$ , h = 1,... L. We denote by *n* the total number of clusters selected. To each individual *k* we attach a basic weight appropriate to the sample selection mechanism. As in Yung (1996), we 'normalize' it by dividing the basic weight by M, the total number of individuals in the finite population. We denote the resulting weight by $w_{hik}$ and, when no confusion may arise, by $w_k$ , k = 1, ... M, i = 1, ... $n_h$ , h = 1, ... L

**<u>Definition 1.</u>** In the case of the GEE (2), the census parameter $\beta_N$ is defined as the solution (when it exists and is unambiguously defined) of equation (4) below:

$$\sum_{k=1}^{M} U_k(\beta, \alpha_N(\beta), \varphi_N(\beta)) = 0 \qquad (4)$$

□

We will define next a sample - based estimator $\hat{\beta}_N$, which will serve to make design based inference on the census parameter $\beta_N$. In conjunction with the GEE (2), we define, for $\beta \in 1$ :

$$\hat{f}_N(\beta) = f_N(s, \beta) = \sum_{k \in s} w_k U_k(\beta, \hat{\alpha}_N(\beta), \hat{\varphi}_N(\beta)) \qquad (5)$$

In (5) $\hat{\alpha}_N(\beta)$ and $\hat{\varphi}_N(\beta)$ are sample based estimators of the census parameters $\alpha_N$, respectively $\varphi_N$. Notice that in case of with - replacement sampling, s is an ordered sample, i.e. the same unit may appear several times in the sample *s* (Särndal et al 1992, p.72)

**<u>Definition 2.</u>** An estimator $\hat{\beta}_N$ of the census parameter $\beta_N$ is defined as a solution to $f_N(s, \beta) = 0$, with $f_N(s, \beta)$ as in (5) above. □

**<u>Example 2.</u>** Consider the simpler situation of an IEE presented in Example 1. The census parameter in Example 1 is $\beta_N = \hat{\beta}_G$ in (3). A design based estimator $\hat{\beta}_N$ is a solution to $\hat{f}_N(\beta) = f_N(s,\beta) = 0$, where:

$$f_N(s, \beta) = \sum_{i \in s} w_i \sum_{j=1}^{2} x_{ij}(y_{ij} - x_{ij}\beta)$$

This estimator can be found explicitly as the EE above has the unique solution:

$$\hat{\beta}_N = \frac{\sum_{i \in s} \sum_{j=1}^{2} w_i x_{ij} y_{ij}}{\sum_{i \in s} \sum_{j=1}^{2} w_i x_{ij}^2} \qquad (6)$$

Note that in (6) the normalized weights can be replaced by the original design weights. □

**4. Consistency of $\hat{\beta}_N$.** We first give conditions for the existence of an RLE estimator $\hat{\beta}_N$ as well as on its convergence to a constant, which is a major step in proving its design consistency.

**Assumption 1** (also included in Binder (1983)):

(i) $f_N(s, \beta) \xrightarrow{p_N} f(\beta), \ \forall \beta \in \Theta$, where *f* is a non random function defined on the space of parameters $\Theta$ which may be unbounded. Recall $p_N$ is the design probability.

(ii) $f(\beta_0) = 0$, and all partial derivatives of *f* exist and are continuous around $\beta_0$.

(iii) $D_\beta[f]|_{\beta_0}$ is invertible (it suffices to have $\det |D_\beta[f]|_{\beta_0}| \neq 0$), where $D_\beta[f]$ is the p×p matrix of partial derivatives of *f*. □

**Remark** Assume that $\beta_0$ is the true superpopulation parameter used in $S_i$, i = 1,...N. Then $E_m[Y_i - a'(\theta_i)] = 0$, $i \geq 1$, by the first model assumption in equation (1). □

**Assumption 2** For $K_0 = K(\beta_0)$ a compact containing $\beta_0$, $K_0 \subset \Theta$ and any $\varepsilon > 0$, $\exists$ a constant $h_0$ and an integer $n_0$ such that, for the partial derivatives of $\hat{f}_N(\beta) = (\hat{f}_j^N(\beta))_{j=1,\ldots p}$, $\beta = (\beta_k)_{k=1,\ldots p}$

(iv)
$$\sup_{n \geq n_0} p_N\left\{s : \sup_{\beta \in K_0} \left| \frac{\partial \hat{f}_j^N(\beta)}{\partial \beta_k} \right| \geq h_0 \right\} \leq \varepsilon$$

for all i, j = 1, ... p . □

We note that (iv) is equation (4.69) of Shao (1999).

**Example 3** : We consider again Example 2 above.

$$f_N(s, \beta) = \sum_{i \in s} w_i \sum_{j=1,2} x_{ij}(y_{ij} - x_{ij}\beta)$$
$$\rightarrow \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1,2} x_{ij}(y_{ij} - x_{ij}\beta)$$

(if design consistency).

If the Strong Law of Large Numbers holds in the superpopulation, we have

$$\xrightarrow{\text{SLLN's}} \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1,2} x_{ij}\left[E_m[y_{ij}] - x_{ij}\beta\right] \rightarrow f(\beta)$$

$$\text{if} \quad \lim \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1,2} x_{ij}^2 = X_2 < \infty$$

by the model assumption (1), where :

$$f(\$) \text{ ' } X_2(\$_0 \& \$)$$

Therefore (i) of Assumption 1 holds. Now clearly $f(\$_0)$ ' 0 and so (ii) also holds. For (iii) to hold, we notice that the derivative of $f$ (\$) is - $X_2$ , which is different from zero if at least one of the covariates is.

To verify Assumption 2, we first take the derivative of $\hat{f}_N(\$)$ with respect to \$, N \$ 1. We note that the survey weights do not depend on \$ and neither does $D_{\$}\hat{f}_N(\$)$ in this example. Furthermore, if we have design consistency of totals, we conclude that:

$$D_{\$}[\hat{f}_N(\$)] \bullet D_{\$}[f_N(\$)] \text{ ' } \frac{\&1}{M} \sum_{i,j} x_{ij}^2 .$$

Note that the right hand side of the equation above is bounded if the covariates are equibounded, or if the right hand side converges, i.e. $X_2$ < 4 (

The proof of the following result was given in the scalar case only ( p = 1 ). To treat the more general situation of a vector-valued parameter \$, one could adapt the proof of Theorem 4.17 in J. Shao (1999) to the case of design-based inference.

**Theorem 1.** ( Existence and convergence of $\hat{\$}_N$ , $N$ \$ 1 ). Under Assumptions 1 and 2, there exist estimators $\hat{\$}_N(s)$ such that, œ 0 , * > 0 , › $n_0(0,*)$ ' $n_0$ such that:

$$\sup_{n \,\$ n_0} p_N 7 s: *\hat{\$}_N(s) \& \$_0 * \# *, \; f_N(s, \hat{\$}_N) \text{ ' } 0 ? \$ 1 \& 0$$

The proof relies on functional properties of $f$ , which is invertible in a neighbourhood of $\$_0$( by (ii) (iii) and the inverse function theorem), and the uniform convergence of the sequence $\hat{f}_N(\$)$, $N$ \$ 1 , which follows from (i) and Assumption 2. The monotonicity of the function $f$ (\$) is also used, which makes this proof unsuitable for p > 1. (

In Example 3 the verification of the assumptions was done in two stages, the first based on assumptions of design consistency and the second on Assumptions 1&2 holding for $f_N$ , $N$ \$ 1 . Under this new set of conditions, we also obtain $\$_N$ 6 $\$_0$ and consequently design consistency:

**Corollary 1.** If Assumptions 1&2 hold for $f_N$ , $N$ \$ 1 and the estimators of totals are design consistent, then $\$_N$ 6 $\$_0$, $\hat{\$}_N \overset{p_N}{6} \$_0$ and so $\hat{\$}_N \& \$_N \overset{p_N}{6} 0$ as n 6 4 . Furthermore, the convergence in (i) is uniform in \$. (

Theorem 1 is also valid in the GEE situation. Conditions for Assumptions 1&2 to hold are more complex because, as mentioned above, the EE are no longer sums of independent r. v.'s in the design, due to the presence of the estimated correlation structure across time. As in Rao (1998), we replace $V_i$ (\$ ) in $U_i$ (\$, ", n) = $D_i^T V_i^{-1} S_i$ by a sample based estimator when R(\$) is completely unspecified, i \$1 (Example 5 of Liang and Zeger (1986)). We illustrate the procedure on an example.

**Example 4.** The marginal model is that of Example 1. Recall that $\frac{d^2 a}{d\theta_{it}^2} = 1$ in this case. An estimator of $V_i(\beta) = V(\beta)$ is a 2×2 matrix $\hat{C}_N(\beta)$ with entries $\hat{c}_N(j,\ell)(\beta) = \sum_{i \in s} w_i s_{ij}(\beta)\, s_{i\ell}(\beta)$, $j, \ell = 1,2$, where $s_{ij}(\beta) = y_{ij} - \mu_{ij}(\beta)$, j = 1, 2 and i = 1, ... M. The entries of the matrix $\hat{G}_N(\beta) = \hat{C}_N^{-1}(\beta)$, which is assumed to exist, will be denoted by $\hat{g}_N^{jk}(\beta) = \hat{g}_{jk}(\beta)$, j, k = 1,2, $\beta \in B$. They could be calculated from the entries of $\hat{C}_N(\beta)$. Substituting in (5), we obtain $\hat{f}_N(\beta) = \sum_{j,k=1,2} \hat{g}_{jk}(\beta)\, \hat{f}_N^{jk}(\beta)$, where $\hat{f}_N^{jk}(\beta) = \sum_{i \in s} w_i x_{ij} s_{ik}(\beta)$ j, k = 1,2. Therefore verifying Assumptions 1&2 reduces to verifying these assumptions on a finite number of terms (d×d = 4 in this case), which we can treat separately.

If we assumed $\hat{g}_{jk}(\beta) \xrightarrow{p_N} g_{jk}(\beta)$, $\forall \beta \in B$, $j,k = 1,2$, then conditions that ensure consistency in the GEE case reduce to conditions on sums of independent variables, as was the case with IEE, and an additional condition :

(v) $$\hat{C}_N(\beta) \xrightarrow{p_N} C(\beta), \forall \beta \in B.$$

In (v), $C(\beta)$ is an invertible covariance matrix, $\forall \beta \in B$. For instance, if Assumption 1 holds for all $\hat{f}_N^{jk}$, j, k = 1,2 with the same $\beta_0$, it also holds for $\hat{f}_N$, $N \geq 1$. When considering Assumption 2, we see that :

$$-\hat{J}_N(\beta) = \frac{d\hat{f}_N(\beta)}{d\beta} = \sum_{j,k=1,2} \frac{d\hat{g}_{jk}(\beta)}{d\beta}\, \hat{f}_N^{jk}(\beta) + \sum_{j,k=1,2} \hat{g}_{jk}(\beta) \frac{d\, \hat{f}_N^{jk}(\beta)}{d\beta}.$$

For $\beta$ close to $\beta_0$, we require the first term to converge to 0 and the second to converge to a constant $-J_0 \neq 0$ as n increases. These conditions are stronger than what is needed for (iv) but are required for asymptotic normality anyhow. To obtain them, we assume that the conditions of Theorem 1 hold for $\hat{f}_N^{jk}(\beta)$. The convergence in (i) is then uniform in $\beta$, j, k = 1,2. Still dealing with the first term, we ask that (iv) hold for $\hat{g}_N^{jk}(\beta)$ : for each j, k = 1,2 , $K_0 = K(\beta_0)$ a compact containing $\beta_0$, $K_0 \subset B$ and $\varepsilon > 0$, $\exists$ a constant $g_0$ and an integer $n_0$ such that:

(vi) $$\sup_{n \geq n_0} p_N \{ s : \sup_{\beta \in K_0} \left| \frac{d\hat{g}_N^{j,k}}{d\beta}(\beta) \right| \geq g_0 \} \leq \varepsilon$$

To obtain uniform convergence to $-J_0$ we require uniform convergence in (v). (

Condition (v) appears in Shao (1999), p. 315. It is, however, not sufficient to ensure the asymptotic normality of $n^{1/2}[\hat{\beta}_N - \beta_N]$. For asymptotic normality to hold, we need a $n^{1/2}$- design consistency to match (i) - (ii) in Theorem 2 of Liang and Zeger (1986). Note that $J_0$ is 'the bread of the sandwich' in the asymptotic variance of $n^{1/2}[\hat{\beta}_N - \beta_N]$, which we discuss next. Condition (vi) seems to appear here for the first time.

**Example 5** Consider the conditions spelt out in Example 4 and:

(vii) $n^{1/2}\,[\hat{g}_{jk}\,(\$_N) - g_{jk}\,(\$_N)]$ is $O_{p_N}\,(1)$, with $g_{jk}\,(\$_N)$ finite population parameters, j, k = 1, 2.

Then $n^{1/2}[\,\hat{\$}_N \,\&\$_N\,]$ has the same asymptotic distribution as $\&J_0 n^{1/2} \sum_{j,k\,'\,1,2} g_{jk}(\$_N)\; \hat{f}_N^{jk}(\$_N)$.

To see this, we first apply the mean value theorem to $\hat{f}_N(\$)\;'\sum_{j,k\,'\,1,2} \hat{g}_{jk}(\$)\;\hat{f}_N^{jk}(\$)$ at points $\hat{\$}_N$, $\$_N$ and use the properties of the derivative $\hat{J}_N(\$)$ in Example 4 at an intermediate point $\bar{\$}_N$,

.

for N $\$$ 1. Since for large $n$ and on a large set we have $\hat{f}_N(\hat{\$}_N)\;'\;0$, we must now study the asymptotic distribution of $n^{1/2}\hat{f}_N(\$_N)$ as n 6 4 . We write $\hat{g}_{jk}\,(\$_N)\;'\;\hat{g}_{jk}(\$_N)\&g_{jk}(\$_N)\%g_{jk}(\$_N)$ substitute in $\hat{f}_N(\$_N)$ and apply (vii) with the results of Example 4. (

**5. The jackknife estimator of the asymptotic variance**. In the context of GEE, Liang and Zeger (1986) stated some of the conditions needed for the CLT to hold for $\hat{\$}_G\&\;\$_0$. A sketch of the proof and the asymptotic form of the variance are given in the Appendix to their paper. A design-based, rigorous proof of the normality of $n^{1/2}(\hat{\$}_N\&\$_N)$ is more complex partly because of the existence of two levels of the main and the nuisance parameters.

The weakest form of the CLT that is needed involves normalized sums of random variables for fixed values of the parameter $\$$:

(CLT$\$$) $$V_N^{\&1/2}(\$)\,[\,\hat{f}_N(\$)\&E\hat{f}_N(\$)]\;\overset{N}{6}\;N(0,I_p)\ \text{as n 6 4}$$

where $V_N(\$)$ is the (design) variance of $\hat{f}_N(\$)$ and $E\hat{f}_N(\$)$ is its (design) expectation. In the case of IEE with our design, Lyapunov type conditions are sufficient for (CLT$\$$) to hold, for any fixed $\$$ (e.g. (6.36) in Shao & Tu (1995), condition C1 of Yung (1996)). This is so because we are dealing with a sum of independent r. v.' s. As the authors above do, we will also assume C2 of Yung (1995), i.e. , with the notation in (CLT$\$$):

(C2$\$$) $n V_N(\$)$ 6 E($\$$), œ $\$$, as n 6 4, where E($\$$) is invertible.

The stronger (CLT$\$_N$) form is needed here , with $\$$ replaced by $\$_N$ in CLT$\$$, even in the context of IEE (see Part I, Section 4 of Rubin - Bleuer (1998). We assume equicontinuity of the components of $U_i$ , i = 1, ... N, in (2) or (4) , i.e. for IEE or GEE with known variance structure, we require:

(E) The family of functions $\{u_{i,R}(\$)\}_{i,R}$ is equicontinuous in $\$$ at $\$_0$ .

Such assumptions appear elsewhere in the technical literature in connection with asymptotic results for GEE or IEE (e.g. Theorem 5.14 of Shao (1999) and Rubin - Bleuer (1998) when p =1). In many interesting instances, condition (E) is implied by continuity of functions of the covariates and boundedness of the covariates (e.g. equation (3) in Example 1). We consider the one step jackknife estimator with the design described in section 3. From the sample of n clusters, let us delete cluster *i*, which we assume belongs to stratum *h*. All individual weights in each of the remaining $n_h$ - 1 clusters in stratum $h$ are multiplied by the factor $n_h/(n_h - 1)$ to compensate for the deletion of one cluster. The weights in other strata are left unchanged. Of course, all individual weights in cluster $i$ are set to 0. Let $w_k^*$ be the new weights, $k \in s$. The estimator that corresponds to $\hat{f}_N(\theta)$ in (5) will be denoted, for simplicity, by $\hat{f}_{-i}(\theta)$. More precisely, we have, for each cluster $i \in s$ :

$$(7) \qquad \hat{f}_{-i}(\theta) = \sum_{k \in s^*_i} w_k^* U_k(\theta, \bullet(\theta), \hat{N}(\theta))$$

We introduce, for cluster $i \in s$, the p×p "information" matrices $\hat{J}_{-i}(\theta)$ along with the p - component vectors $\hat{\theta}_{-i}$, with $\hat{\theta}_N$ as in Definition 2:

$$(8) \qquad \hat{J}_{-i}(\theta) = -\frac{\partial \hat{f}_{-i}(\theta)}{\partial \theta}, \quad \hat{\theta}_{-i} = \hat{\theta}_N + \hat{f}_{-i}(\hat{\theta}_N) \times \hat{J}_{-i}^{-1}(\hat{\theta}_N), \ i \in s,$$

Note that in (8) one must state conditions for the definitions to be valid (i.e. the inverse matrices, should exist at least asymptotically). We can now define the jackknife estimator $v_J(\theta)$. For simplicity of notation, we write $\hat{\theta}$ for $\hat{\theta}_N$ when no confusion may arise.

$$(9) \qquad v_J(\hat{\theta}) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{\theta}_{-i} - \hat{\theta})(\hat{\theta}_{-i} - \hat{\theta})^T$$

The main result presented in this section is the consistency of the jackknife estimator of $V_N(\hat{\theta}_N)$, the variance of $\hat{\theta}_N$, i.e.:

$$(10) \qquad n[v_J(\hat{\theta}_N) - V_N(\hat{\theta}_N)] \xrightarrow{P_N} 0 \text{ as } n \to \infty$$

We state condition C4 of Yung (1996), which also appears in Shao and Tu (1995):

$$(C4) \qquad n \max_{hik} m_{hi} w_{hik} = O_n(1)$$

This condition can be interpreted in terms of the weights associated with the sample design.
It requires that the design weights be comparable in size. We can now state the main result of this article, for the IEE and the case p = 1 only:

**Theorem 2** Assume that the assumptions and thus conclusions of Corollary 1 hold. We assume (C2$\theta$) and that C1 in Yung (1996) holds with $y_{hik}$ replaced by $u_k(\theta)$, k = 1, ... M, for all $\theta$ in the space of parameters. We assume (E) for the family $\{u_k(\theta)\}_{k \geq 1}$ and that condition (C4) on the design also holds. Furthermore, if $J_0$ is a nonzero constant and $\max_i |\hat{J}_{-i}(\hat{\theta}_N) - J_0| \xrightarrow{P_N} 0$ as $n \to \infty$ (see (8) ), the jackknife estimator is consistent as in (10). ∎

Theorem 2 can be extended to the GEE case when the covariance structure is known and then further to the case of pseudo likelihood (equations (2), (5)) in the manner of Example 5. The extension is essentially possible due to the form of the EE in which factors containing the nuisance parameters can be separated from those containing the main parameter, as seen in Example 4.

**6. Conclusions.** Design inference is a useful, interesting and challenging subject. Inference is generally more difficult in finite populations than in infinite populations. In the finite population situation, we have to deal with 2 levels for each of the main and 'nuisance' parameters. Many of the techniques that are used in classical inference can be adapted to the context of survey randomization. However, 'regularity conditions' that involve the interchange of derivatives and expectations taken with respect to the superpopulation model must be replaced by functional conditions. We tried to reduce the model assumptions to a minimum. As in Rao (1998), we retained the first moment model assumption in (1). Even though convergence of census parameters (including population averages in Example 3) can be treated as limits of functions, it is more natural to view them as realizations of sums of r.v.'s, as indicated in Example 3. Therefore, it appears more natural to view design - inference within the more general set-up presented in Rubin-Bleuer (1998), which allows for joint model and design-based inference.

**Acknowledgement**. The author would like to thank Professor J.N.K. Rao for suggesting the research topic and for his continuous assistance and encouragement.

## REFERENCES

Binder, D. A.(1983). On the variances of asymptotically normal estimators from complex surveys, *International Statistical Review*, **51**, 279-292.

Krewski, D. and Rao, J.N.K.(1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication methods, *The Annals of Statistics* **9**, 1010-1019.

Liang, K-Y and Zeger, S. L.(1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Rao, J.N.K. (1998). Marginal models for repeated observations: inference with survey data. *Proceedings of the ASA*, 1998.

Rubin-Bleuer, S.(1998). Sabbatical Research Report, Statistics Canada.

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

Shao, J., Tu, D.(1995) . *The Jackknife and Bootstrap*, Springer Verlag, New York.

Shao, J.(1999). *Mathematical Statistics*, Springer Verlag, New York.

Yung, W.(1996). Contributions to poststartification in stratified multi-stage samples. Ph. D. thesis, Carleton University, Ottawa, Canada.