CONSISTENCY OF GEE ESTIMATORS AND APPLICATION TO DESIGN-BASED INFERENCE

Ioana Şchiopu-Kratina, Statistics Canada
R.H. Coats Building, 16th floor, Ottawa, Ontario K1A 0T6, e-mail: ioana.schiopu-kratina@statcan.ca

Key Words: Estimating Equations, Marginal models, Design- based inference

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.

Large scale longitudinal surveys are often carried out by 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. In the special situation when the observations across time are assumed independent for each individual (the working independence assumption), GEE becomes the Independence Estimating Equation (IEE). Consistency is an essential ingredient in the proof of asymptotic normality. In design we can only define weak consistency, i.e. in terms of convergence in probability. The sets on which the estimators are roots of estimating equations (REE) have asymptotic probability 1 (see Theorem 1). We do not make any assumptions regarding the uniqueness of the REE. We first state a result on the existence and consistency of the main estimator (Theorem 1), then show how it applies in the GEE situation (Corollary 3).

The technical problems that we had to overcome were due to the estimation of the variance structure across time and to obtaining asymptotic results in finite populations with survey randomization. The first problem was solved by Liang and Zeger (1986) in a model - based context. They do not supply proofs for their asymptotic results. In order to do design inference(as in Binder 1983) , we tried to give simple analytical proof (not included here) which do not depend on model assumptions in a superpopulation. However, some of our conditions are more natural if a superpopulation is assumed to exist and some model assumptions were present (e.g. (i) and (ii) of Assumption 1 - see Example 3 and model assumption (1)). The results of Binder (1983) had to be extended from the IEE case to GEE.

This article is organized as follows: Section 2 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. Example 3 illustrates conditions for consistency on the EE in Example 2. Some conclusions are presented in Section 5.

2. Model set-up. The is essentially the set-up in Liang and Zeger (1986). Consider M individuals observed on

d_i occasions (i = 1, ... M). The univariate responses y_{it} and the p - covariates x_{it} 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. Only $E_m[y_{it}]$ and $Var_m(y_{it})$ are specified, where *m* stands for model-based, for all (∀) t, i ≥ 1. Liang and Zeger (1986) consider probability densities $\rho(y_{it}) = \exp[\{y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})\}\phi]$ with $\theta_{it} = h(\eta_{it})$, $\eta_{it} = x_{it}^T\beta$, where a, b and h are known (differentiable) functions, $\{\theta_{it}\}, \phi$ are parameters , x_{it}^T is an 1 × p matrix of covariates and β is an p × 1 vector of main parameters, ∀ t, i≥ 1. Here T stands for transposition of matrices. Note that for random variables with such densities we have:

(1) $$E_m[y_{it}] = \mu_{it} = a'(\theta_{it}), \forall t, i \geq 1:$$

Let $\mu(\eta) = a'(\eta)$, ∀ η in a space of parameters Θ. The function g is a link function if $g \circ \mu(\theta_{it}) = x_{it}^T\beta$, ∀ 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. The logit link function $g(\mu) = \log\{\mu/(1-\mu)\}$ is the natural link associated with the logistic regression model. EE's are formed that mimic log likelihood equations associated with exponential distributions (e.g. normal, binomial, logistic, Poisson). These are quasi - likelihood equations if the original distributions belong to the normal family upon further restrictions, i.e. knowledge on the dispersion parameter ϕ (Shao 1999, p. 242). The idea is to produce estimators for β which are REE by making few assumptions on the distribution of the observed data, and then study the properties of these estimators.

When GEE's are used, it is assumed that correlation of observations y_{it} across time for the same individual is the same for all individuals. X. More precisely, let $U_i(\beta, \alpha, \phi) = D_i^T V_i^{-1} S_i$, $V_i = 1/\phi\,[A_i^{1/2} R(\alpha) A_i^{1/2}]$, $D_i = A_i \Delta_i X_i$, $S_i = Y_i - a'(\theta_i)$, $A_i = \text{diag}\, a''(\theta_{it})$ in R^d and $\Delta_i = \text{diag}\,[d\theta_{it}/d\eta_{it}]$, which could be taken to be the identity matrix I_d, ∀ 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 β, i ≥ 1. The GEE, or equation (7) of Liang and Zeger (1986), is:

(2) $$\sum_{i=1}^{M} U_i(\beta, \hat{\alpha}(\beta), \phi(s, \beta)) = 0$$

Equation (2) above is called a pseudo-likelihood equation in Shao (1999), p. 315. Note that it consists of p scalar equations in which α fully characterizes R. In equation (2) $\hat{\alpha}$ and $\phi(s, \beta)$ are estimates of nuisance parameters that are obtained from the sample and generally contain β. When the solution to (2) exists and is unique , i.e. when β is defined implicitly by (2), it is denoted by $\hat{\beta}_G$ 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 β 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.

When the observations across time are assumed independent for each individual (the working independence assumption), equation (2) becomes IEE. In this case $R(\alpha) = I_d$ and there is no need to estimate nuisance parameters in (2). This is the situation discussed, in a design randomization context, by Binder (1983). In the context of IEE and survey randomization (see Section 3), $\hat{\beta}_G$ becomes the "census" parameter defined in Binder (1983). The example below illustrates the calculation of $\hat{\beta}_G$ 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 ϕ and d =2 occasions. We have $R(\alpha) = I_2$ (case IEE). Assume that x_{it}, β are scalars, i,t ≥ 1 and that *h* is the identity.

$$\rho(y_{it}) = \exp - \frac{(y_{it}-\theta_{it})^2}{2} = \exp\{y_{it}\theta_{it} - a(\theta_{it}) +$$

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

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

Note that each x_{it} has as many components as β (p = 1 components here) and, for i, t ≥ 1:

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

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

(3) $$\sum_{i=1}^{M}\sum_{t=1}^{2} x_{it}y_{it} - \sum_{i=1}^{M}\sum_{t=1}^{2} x_{it}^2\beta = 0 \text{ so } .$$

$$\hat{\beta}_G = \frac{\sum_{i=1}^{M}\sum_{t=1}^{2} x_{it}y_{it}}{\sum_{i=1}^{M}\sum_{t=1}^{2} x_{it}^2} \qquad \blacksquare$$

3. The design and the design-based inference. In the article, inference is done in the design - based randomization as proposed by 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. 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 β_M, which would be more appropriate. The parameter to estimate in the design randomization context changes as $n \to \infty$ (which implies that $N, M \to \infty$). In this article, the parameter β_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 β_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.

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 $w_{h\,i\,k}$ and, when no confusion may arise, by w_k, k = 1, ... M, i = 1, ... n_h , h = 1, ... L

Definition 1. In the case of the GEE (2), the census parameter β_N is defined as the solution (when it exists and is unambiguously defined) of equation (4) below:

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

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

$$(5) \quad \hat{\psi}_N(\beta) = \psi_N(s,\beta) = \sum_{k \in s} U_k(\beta, \hat{\alpha}_N(\beta), \phi_N(s,\beta))$$

In (5) $\hat{\alpha}_N(\beta)$ and $\phi_N(s,\beta)$ are sample based estimators of the census parameters α_N, respectively ϕ_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)

Definition 2. The REE estimator $\hat{\beta}_N$ of the census parameter β_N is defined as a solution to:

$\psi_N(s,\beta) = 0$, with $\psi_N(s,\beta)$ as in (5) above. $\blacksquare$

Example 2. 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 $\hat{\psi}_N(\beta) = \psi_N(s,\beta) = 0$, where:

$$\psi_N(s,\beta) = \sum_{k \in s} w_k \sum_{t=1}^{2} x_{kt}(y_{kt} - x_{kt}\beta)$$

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

$$(6) \qquad \hat{\beta}_N = \frac{\sum_{k \in s}\sum_{t=1}^{2} w_k x_{kt} y_{kt}}{\sum_{k \in s}\sum_{t=1}^{2} w_k x_{kt}^2}$$

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

4. Consistency of $\hat{\beta}_N$. We first give conditions for the existence of an RLE estimator β_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) $\psi_N(s,\beta) \xrightarrow{p_N} \psi(\beta), \forall \beta \in \Theta$, where $\psi(\beta)$ is a non random function defined on the space of parameters Θ which may be unbounded. Recall p_N is the design probability.

(ii) $\psi(\beta_0) = 0$, and all partial derivatives of $\psi(\beta)$ exist and are continuous around β_0.

(iii) $D_\beta[\psi(\beta)]|_{\beta_0} = -J_0$ is invertible (it suffices to have $\det | D_\beta[\psi(\beta)]|_{\beta_0} \neq 0$), where $D_\beta[\psi(\beta)]$ is the pxp matrix of partial derivatives of $\psi(\beta)$. ■

Remark 1 Assume that β_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 β_0, $K_0 \subseteq \Theta$ and any $\eta > 0, \exists$ a constant h_0 and an integer n_0 such that, for the partial derivatives of $\psi_N(s,\beta) = (\psi_j^N(s,\beta))_{j=1,\dots p}$, $\beta = (\beta_k)_{k=1,\dots p}$

$$\text{(iv)} \quad \sup_{n \geq n_0} p_N\left\{s: \sup_{\beta \in K_0} \left| \frac{\partial \psi_j^N(s,\beta)}{\partial \beta_k} \right| \geq h_0 \right\} \leq \eta$$

for all j, k = 1, ... p .■

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

Example 3 : We consider again Example 2 above.

$$\psi_N(s,\beta) = \sum_{k \in s} w_k \sum_{t=1,2} x_{kt}(y_{kt} - x_{kt}\beta)$$

$$\cong \frac{1}{M} \sum_{k=1}^{M} \sum_{t=1,2} x_{kt}(y_{kt} - x_{kt}\beta)$$

(if design consistency). If the Strong Law of Large Numbers holds in the superpopulation, we have

$$\underset{\text{SLLN's}}{\cong} \frac{1}{M} \sum_{k=1}^{M} \sum_{t=1,2} x_{kt}\left[E_m[y_{kt}] - x_{kt}\beta\right] \to \psi(\beta)$$

$$\text{if } \lim \frac{1}{M} \sum_{k=1}^{M} \sum_{t=1,2} x_{kt}^2 = X_2 < \infty$$

by the model assumption (1), where :

$$\psi(\beta) = X_2(\beta_0 - \beta)$$

Therefore (i) of Assumption 1 holds. Now clearly $\psi(\beta_0) = 0$ and so (ii) also holds. For (iii) to hold, we notice that the derivative of $\psi(\beta)$ 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 $\hat{\psi}_N(\beta)$ with respect to β, $N \geq 1$. We note that the survey weights do not depend on β and neither does $D_\beta \hat{\psi}_N(\beta)$ in this example. Furthermore, if we have design consistency of totals, we conclude that:

$$(7) \qquad D_\beta[\hat{\psi}_N(\beta)] \approx D_\beta[\psi_N(\beta)] = \frac{-1}{M} \sum_{k,t} x_{kt}^2 .$$

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. $X_2 < \infty$ ■

The proof of the following result was given in the scalar case only (p = 1).

Theorem 1. (Existence and convergence of $\hat{\beta}_N, N \geq 1$). Assume that $\hat{\psi}_N, N \geq 1$ are continuous and that the convergence in (i) is uniform in β. Assume further that (ii) and (iii) of Assumption 1 hold.
There exist then estimators $\hat{\beta}_N$ such that, $\forall \eta, \delta > 0$, $\exists\ n_0(\eta, \delta) = n_0$:

$$\sup_{\imath \geq n_0} p_N\{s: |\hat{\beta}_N - \beta_0| \leq \delta,\ \psi_N(s, \hat{\beta}_N) = 0\} \geq 1 - \eta$$

The same conclusion holds if Assumptions 1&2 hold. ■

Corollary 1 Assume that:

$$(8) \qquad \hat{\psi}_N(\beta_0) \xrightarrow{p_N} 0$$

$$(9) \qquad D_\beta\ \hat{\psi}_N|_\beta \xrightarrow{p_N} D(\beta), \text{ uniformly}$$

in $\beta \in \Theta \cap K$, where $D(\beta)$ is a nonrandom function continuous on K, a compact set containing β_0.

$$(10) \qquad J_0 = -D(\beta_0) \text{ is positive definite and invertible.}$$

We conclude then that there exists a function $\psi(\beta)$, $\beta \in K$, where K is a compact set in Θ, such that uniform convergence in probability holds for $\hat{\psi}_N(\beta)$ and $\psi(\beta)$, as well as for $D_\beta\ \hat{\psi}_N|_\beta$ and $D_\beta\ \psi|_\beta = D(\beta)$. Furthermore, the conclusion of Theorem 1 also holds. ■

Remark 2: Note that uniform convergence in (9) is implied by pointwise convergence and:

(11) Condition (iv) of Assumption 2 holds for $D_\beta(\hat{\psi}_N)$ rather than $\hat{\psi}_N$. ■

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 $\psi_N(s,\beta)$, $N \geq 1$. Under this new set of conditions, we also obtain $\beta_N \to \beta_0$ and consequently design consistency:

Corollary 2. If Assumptions 1&2 hold and

$\psi_N(s,\beta)$, $D_\beta \psi_N(s,\beta)$ $N \geq 1$ are design consistent,

then $\beta_N \to \beta_0$, $\hat{\beta}_N \xrightarrow{p_N} \beta_0$ so $\hat{\beta}_N - \beta_N \xrightarrow{p_N} 0$

as $n \to \infty$. Furthermore, the convergence in (i) of

Assumption 1 is uniform in β. ■

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.

In order to do statistical inference for GEE, we must find a sample based estimator of $V_k = V_k(\alpha, \beta)$ (see Rao (1998)), and replace it in $U_k(\alpha, \beta, \varphi) = D_k^T V_k^{-1} S_k$, $k = 1,\ldots M$. This corresponds to the case when $R(\alpha)$ is completely unspecified in Example 5 of Liang and Zeger (1986). In this instance there is no need to estimate the overdispersion parameter ϕ. To estimate $V_k(\beta) = A_k^{1/2} C_N(\alpha,\beta) A_k^{1/2}$, $k \geq 1$ for fixed values of the parameters, we estimate the common correlation structure across time, denoted here $C_N(\alpha,\beta)$, by $\sum_{k\in s} w_k A_k^{-1/2}(\beta) S_k(\beta) S_k^T(\beta) A_k^{-1/2}(\beta)$. The entries of this matrix are:

$$\hat{c}_N^{ij}(\beta) = \sum_{k\in s} w_k [a''(\eta_{ki}(\beta)) a''(\eta_{kj}(\beta))]^{-1/2} s_{ki}(\beta) s_{kj}(\beta),$$

where $s_{ki}(\beta) = y_{ki} - \mu_{ki}(\beta)$ $k = 1,\ldots M$, $i, j = 1,\ldots d$, $\beta \in \Theta$. Let $\hat{g}_N^{ij}(\beta)$, $i, j = 1, \ldots d$, $\beta \in \Theta$, be the entries of $\hat{C}_N^{-1}(\beta)$, which is assumed to exist. Then $\hat{V}_k^{-1}(\beta)$ has entries $\hat{g}_N^{ij}(\beta)[a''(\eta_{ki}(\beta)) a''(\eta_{kj}(\beta))]^{-1/2}$, $i, j = 1,\ldots d$. We substitute in GEE (5):

$$U_k(\beta, \hat{V}_k(\beta)) = D_k^T \hat{V}_k^{-1}(\beta) S_k, \quad S_k = Y_k - a'(\theta_k), \forall\, k \geq 1$$

and obtain:

$$\psi_N(s,\beta) = \sum_{i,j=1,\ldots d} \hat{g}_N^{ij}(\beta)\ \psi_N^{ij}(s,\beta),$$

with:

$$\psi_N^{ij}(s,\beta) = \sum_{k\in s} w_k \left[\frac{a''(\eta_{ki}(\beta))}{a''(\eta_{kj}(\beta))}\right]^{1/2} x_{ki} s_{kj}(\beta), \quad i, j = 1, \ldots d.$$

Therefore, the GEE in (5) can be written as a finite sum of terms with each of these terms equal to a product of two estimators. Furthermore, each $\psi_N^{ij}(s,\beta)$, $i, j = 1,\ldots d$, is a sum of random variables for which the conditions for consistency in Theorem 1 can easily be applied. We assume that a symmetric, invertible matrix $C(\beta)$ exists and is continuous at β_0 and that:

(v) $$\hat{C}_N(\beta) \xrightarrow{p_N} C(\beta),$$

This implies:

$$\hat{g}_N^{ij}(\beta) \xrightarrow{p_N} g^{ij}(\beta), \forall\, \beta \in \Theta,\ i,j = 1,\ldots d$$

However, we often need uniform convergence on a compact space containing β_0 and so we would have to impose conditions stronger than (v). Consider now the information matrices $\hat{J}_N(\beta) = -\dfrac{d\hat{\Psi}_N(\beta)}{d\beta}$, $N \geq 1$.

Corollary 3. Assume that (9) holds for $\hat{\psi}_N^{ij}(\beta)$, $\hat{g}_N^{ij}(\beta)$, $i, j = 1\ldots d$. Then, by the proof of Corollary 1, $\psi_N^{ij}(s,\beta) \xrightarrow{p_N} \psi^{ij}(\beta)$ and $\hat{g}_N^{ij}(\beta) \xrightarrow{p_N} g^{ij}(\beta)$ for some $\psi^{ij}(\beta)$, $g^{ij}(\beta)$ $i, j = 1,\ldots d$. Let $\psi(\beta) = \sum_{i,j=1,\ldots d} g^{ij}(\beta)\ \psi^{ij}(\beta)$. We assume that $\psi(\beta_0) = 0$ and that $-J_0 = \dfrac{d\psi(\beta_0)}{d\beta} \neq 0$.

Then the conclusions of Theorem 1 hold for $\hat{\psi}_N(\beta)$ and $\psi(\beta)$. Furthermore, $\hat{J}_N(\beta)$ are equicontinuous at β_0 and $\hat{J}_N(\beta) \xrightarrow{p_N} J_0$. Under the additional conditions

$\psi^{ij}(\beta_0) = 0$, $i, j = 1, \ldots d$, we have

$$J_0 = -\sum_{ij} g^{ij}(\beta_0) \frac{d\psi^{ij}(\beta_0)}{d\beta} \neq 0. \ ■$$

5. 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. We suggest therefore to view design- based inference within the more general set-up presented in Rubin-Bleuer (1998), which allows for join model and design-based inference.

REFERENCES

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

Binder , D.A. and Patak, Z. (1994). Use of estimating functions for estimation from complex surveys, *JASA*, **89**, 1035- 1043.

Hájek, J. (1960). Limiting distributions in simple random sampling from finite populations, *Publ. Math. Inst. Hungarian Acad. Sci.*, **5**, 361 - 374.

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

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

Lipsitz, S.R., Laird, N.M. and Harrington, D.P. (1990). Using the jackknife to estimate the variance of regression estimators from repeated measurement studies, *Communications in Statistics* 1990, 821-844.

Rao, C. R. (1974). *Linear Statistical Inference and its Applications*. Wiley, New York.

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

Rudin,W.(1964). *Principles of Mathematical Analysis*, second edition, Mc Graw-Hill, New York.

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.(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.