MARGINAL MODELS FOR REPEATED OBSERVATIONS: INFERENCE WITH SURVEY DATA

J.N.K. Rao, Carleton University
School of Mathematics and Statistics

Key Words: Jackknife, quasi-score tests

1. Introduction

In a longitudinal survey sample subjects are observed over two or more time prints. Such surveys are suited to study individual changes over time, unlike cross-sectional surveys. Applications of longitudinal surveys include: (a) Gross flows: estimation of transition counts between a finite number of states for individuals in a population from one point in time to the next. Such flow estimates are important to researchers and policy analysts for understanding labor market dynamics. (b) Event history modelling; for example unemployment spells. (c) Elimination of the effect of latent variables in regression models using individual changes in the response and explanatory variables between two consecutive time points. (d) Modelling the marginal means of responses as functions of covariates. (e) Conditional modelling of the response at a given time point as a function of past responses and present and past covariates. Such models can provide better understanding of the underlying dynamics than the marginal models (d). Binder (1998) gave an excellent account of the issues related to longitudinal surveys.

Longitudinal surveys typically lead to dependent observations on the same subject, in addition to the customary cross-sectional correlations induced by the clustering in the sample design. In this paper we focus on marginal modelling and analysis of such longitudinal survey data. The case of a simple random sample of individuals has been studied extensively in the literature, especially in the analysis of data occuring in biomedical and health sciences. Liang and Zeger (1986) used generalized estimating equations, requiring only correct specification of the marginal mean. They obtained standard errors of regression parameter estimates and associated "Wald" tests, assuming a "working" correlation structure for the repeated measurements on a sample subject. Rotnitzky and Jewell (1990) developed "quasi-score" tests and "Rao-Scott" adjustments to working quasi-score tests, under marginal models. These methods are asymptotically valid regardless of the true within-subject correlation structure, but assume independence of sample subjects which is not satisfied for complex longitudinal survey data based on stratified cluster samples.

In this paper Wald and quasi-score tests for longitudinal survey data are proposed, using the Taylor linearization and jackknife methods. These methods take account of the survey design features (clustering, stratification, unequal sampling weights etc.) as well as the longitudinal feature and thus are asymptotically valid.

2. Independence Estimating Equations

Suppose the survey population U consists of M individuals and a sample, s, of individuals is selected using stratified multistage sampling. Let w_{hik} denote the basic design weight attached to the k-th sample individual in the i-th sample cluster ($i = 1, \dots, n_h$) from the h-th stratum ($h = 1, \dots, L$). In a longitudinal survey, the sample s is observed over a specified number of time points, say T, but in practice some of the sample individuals may not respond. The sample weights of respondents, s_r, on the first occasion are first adjusted for unit nonresponse, and then subjected to post-stratification adjustment to ensure consistency with known benchmark totals, e.g., age-sex counts obtained from external sources. We denote the final weights as w^*_{hik}, often called as longitudinal weights.

Suppose that the i-th respondent is observed for T_i occasions ($1 \leq T_i \leq T$ and $i \in s_r$). We assume that the responses are missing completely at random (MCAR), i.e., the response probabilities for an individual do not depend on the missing responses and the observed responses, following Liang and Zeger (1986). The data for the i-th sample individual ($i \in s_r$) consists of $\{y_{it}, \mathbf{x}_{it}\}$, $t = 1, \ldots, T\}$ where y_{it} is response on occasion t and $\mathbf{x}_{it}$ is a $p \times 1$ vector of associated covariates. In the case of binary response, $y_{it} = 1$ if subject i has the attribute at time t, and 0 otherwise.

The marginal model assumes that the mean response $\mu_{it} = E_m(y_{it})$ is a specified function of $\mathbf{x}_{it}$ and regression parameters $\boldsymbol{\beta}$; in particular $g(\mu_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}$ where $g(\cdot)$ is called the link function. With binary responses, the logit link function $g(\mu) = \log\{\mu/(1-\mu)\}$ is a natural choice, leading to a logistic regression model. In this section, we assume "working" independence so that $\text{cov}(\mathbf{y}_i) = \mathbf{V}_{0i} = \underset{1 \le t \le T_i}{\text{diag}}(V_{0it})$, where $V_{0it} = V_0(\mu_{it}) = \text{var}(y_{it})$ is the working variance. For example, in the binary response case, $V_0(\mu_{it}) = \mu_{it}(1-\mu_{it})$.

The above formulation permits time varying regression coefficients. For example if $T = 2$ and $g(\mu_{it}) = \alpha_t + \beta_t z_{it}$, $t = 1, 2$, then we can define $\mathbf{x}_{i1} = (1\ z_{i1}\ 0\ 0)^T$, $\mathbf{x}_{i2} = (0\ 0\ 1\ z_{i2})^T$ and $\boldsymbol{\beta} = (\alpha_1\ \beta_1\ \alpha_2\ \beta_2)^T$. In this case, it might be of interest to test the constancy of the slope coefficient over time, i.e., $H_0 : \beta_1 = \beta_2$ which is of the form $H_2 : \mathbf{c}^T\boldsymbol{\beta} = 0$ with $\mathbf{c} = (0\ 1\ 0\ -1)^T$.

We assume that the marginal model holds for the whole population of M subjects so that we get the "census" model

$$g(\mu_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}, \quad t = 1, \ldots T_i;\ i = 1, \ldots, M \tag{2.1}$$

where T_i now refers to the number of consecutive occasions the i-th population subject would respond if contacted. We further assume that the population of M subjects is a self-weighting sample from a super population obeying the marginal model. It is not necessary to regard the population as a random sample from the super population. The census generalized estimating equations (GEE) are then given by

$$S_\ell(\boldsymbol{\beta}) = \sum_{i=1}^{M} u_{i\ell}(\boldsymbol{\beta}) = 0, \quad \ell = 1, \ldots, p \tag{2.2}$$

where

$$u_{i\ell}(\boldsymbol{\beta}) = \sum_{t=1}^{T_i} \frac{\partial \mu_{it}}{\partial \beta_\ell} \frac{(y_{it} - \mu_{it})}{V_{0it}}. \tag{2.3}$$

Under general conditions, the solution of (2.2), $\boldsymbol{\beta}_M$, is a consistent estimator of $\boldsymbol{\beta}$. We denote $\boldsymbol{\beta}_M$ as the census regression parameter and make statistical inferences on $\boldsymbol{\beta}_M$, following Binder (1983). Such inferences are also valid for $\boldsymbol{\beta}$ under certain conditions. For simplicity, we do not distinguish between $\boldsymbol{\beta}_M$ and $\boldsymbol{\beta}$ in this paper.

Noting that the left hand side of (2.2) is the population total of $u_{i\ell}(\boldsymbol{\beta})$, a design-consistent estimator of (2.2), sample GEE, are given by

$$\sum_{hik \in s_r} w_{hik}^* \mathbf{u}_{hik}(\boldsymbol{\beta}) = \mathbf{0}, \tag{2.4}$$

where $\mathbf{u}_{hik}(\boldsymbol{\beta}) = [u_{hik1}(\boldsymbol{\beta}), \ldots, u_{hikp}(\boldsymbol{\beta})]^T$ with $u_{hik\ell}(\boldsymbol{\beta})$ obtained from (2.3) by changing "i" to "hik". The solution of (2.4), $\hat{\boldsymbol{\beta}}$, is a design-consistent estimator of $\boldsymbol{\beta}_M$.

3. Inference Under Working Independence

It is a common practice among social scientists and others to use normalized weights $\tilde{w}_{hik} = mw_{hik}^*/\sum_{s_r} w_{hik}^*$, where m is the size of s_r, and then apply standard methods using SAS or other standard programs. Using the normalized weights in the standard "sandwich" covariance estimator of $\hat{\boldsymbol{\beta}}$, we get the following naive covariance estimator:

$$\begin{aligned} &\mathbf{v}_N(\hat{\boldsymbol{\beta}}) \\ &= [\tilde{\mathbf{I}}(\hat{\boldsymbol{\beta}})]^{-1}\Big(\sum_{s_r} \tilde{w}_{hik}\mathbf{u}_{hik}(\hat{\boldsymbol{\beta}})\mathbf{u}_{hik}(\hat{\boldsymbol{\beta}})^T\Big)[\tilde{\mathbf{I}}(\hat{\boldsymbol{\beta}})]^{-1} \end{aligned} \tag{3.1}$$

where $\tilde{\mathbf{I}}(\hat{\boldsymbol{\beta}})$ is the estimated information matrix with

$$\tilde{\mathbf{I}}(\boldsymbol{\beta}) = -\sum_{s_r} \tilde{w}_{hik} E_m[\partial \mathbf{u}_{hik}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^T]. \tag{3.2}$$

This follows by applying the Liang-Zeger sandwich covariance estimator formula to the census parameter:

$$\mathbf{v}(\beta_M) \tag{3.3}$$
$$= [\mathbf{I}(\beta_M)]^{-1}\Big[\sum_{hik\in U}\mathbf{u}_{hik}(\beta_M)\mathbf{u}_{hik}(\beta_M)^T\Big] \times [\mathbf{I}(\beta_M)]^{-1}$$

and then replacing each term in (3.3) by its estimator based on the normalized weights, $\tilde{w}_{hik}$, where $\mathbf{I}(\beta_M)$ is the census information matrix with

$$\mathbf{I}(\beta) = -\sum_{hik\in U} E_m[\partial\mathbf{u}_{hik}(\beta)/\partial\beta^T]. \tag{3.4}$$

In the case of a simple random sample and no post-stratification adjustant, we have $\tilde{w}_i = 1$ for all the sample subjects i and (3.3) reduces to the Liang-Zeger formula.

Suppose we are interested in testing a hypothesis of the form $H_0 : \beta_2 = \beta_{20}$, using the sample data $\{(y_{hikt}, \mathbf{x}_{hikt});\ hik \in s_r,\ t = 1,\ldots,T_{hik}\}$, where β is partitioned as $\beta = (\beta_1^T, \beta_2^T)^T$ with β_2 a $r\times 1$ vector and β_1 and $q \times 1$ vector $(q+r = p)$. For example, β_2 could represent interaction terms and we are interested in testing for the absence of interactions, i.e., $\beta_{20} = \mathbf{0}$. A "naive" Wald test of H_0 treats

$$W_N = (\hat{\beta}_2 - \beta_{20})^T[\mathbf{v}_{N22}(\hat{\beta})]^{-1}(\hat{\beta}_2 - \beta_{20}) \tag{3.5}$$

as a χ^2 variable with r degrees of freedom (d.f.), where $\mathbf{v}_{N22}(\hat{\beta})$ is the submatrix of $\mathbf{v}_N(\hat{\beta})$ corresponding to β_2. This test, however, is asymptotically incorrect under stratified multistage sampling or any other complex sampling design. In fact, $\mathbf{W}_N$ is asymptotically distributed as a weighted sum of independent χ_1^2 variables, where the weights are the eigenvalues of a "design effects" matrix. As a result, the naive test could lead to inflated significance levels relative to the nominal level, say 0.05.

We assume that the sampling design provides consistent, asymptotically normal estimators of totals. Following Binder (1983), under certain regularity conditions, $\hat{\beta}$ is then asymptotically normal with mean β_M and its covariance matrix, $\text{cov}(\hat{\beta})$, can be consistently estimated by

$$v_L(\hat{\beta}) = [\hat{\mathbf{J}}(\hat{\beta})]^{-1}\mathbf{v}(\hat{\mathbf{S}})[\hat{\mathbf{J}}(\hat{\beta})]^{-1}. \tag{3.6}$$

Hence

$$\hat{\mathbf{J}}(\beta) = -\partial\hat{\mathbf{S}}(\beta)/\partial\beta^T = -\sum_{hik\in s_r} w^*_{hik}\partial\mathbf{u}_{hik}/\partial\beta^T \tag{3.7}$$

and $\mathbf{v}(\hat{\mathbf{S}})$ is the estimated covariance matrix of $\hat{\mathbf{S}}(\beta)$ under the specified sampling design evaluated at $\beta = \hat{\beta}$. Note that $\mathbf{v}(\hat{\mathbf{S}})$ is obtained from standard survey variance estimator, noting that $\hat{\mathbf{S}}(\beta)$ is the vector of estimated totals of $u_{hik\ell}(\beta)$, $\ell = 1,\ldots,p$. However, the variance estimator used should account for post-stratification and nonresponse adjustment. For example, if the post-stratification indicator variables are denoted by $\mathbf{z}_{hik}$, $hik \in S$, and nonresponse is absent, then $\mathbf{v}(\hat{\mathbf{S}})$ is the estimated covariance matrix of $\hat{\mathbf{E}}(\beta) = \sum_{hik\in s} w_{hik}\mathbf{e}_{hik}(\beta)$, evaluated at $\beta = \hat{\beta}$, where $e_{hik\ell}(\beta) = u_{hik\ell}(\beta) - \mathbf{z}_{hik}^T\hat{\mathbf{B}}_\ell$ with

$$\hat{\mathbf{B}}_\ell = \Big(\sum_s w_{hik}\mathbf{z}_{hik}\mathbf{z}_{hik}^T\Big)^{-1} \times \Big(\sum_s w_{hik}\mathbf{z}_{hik}u_{hik\ell}(\beta)\Big), \quad \ell = 1,\ldots,p.$$

Letting $\mathbf{e}^*_{hi} = n_h\sum_k w_{hik}\mathbf{e}_{hik}(\hat{\beta})$, we have

$$\mathbf{v}(\hat{\mathbf{S}}) = \sum_h \frac{1}{n_h(n_h-1)}\sum_i(\mathbf{e}^*_{hi} - \mathbf{e}^*_{h\cdot})(\mathbf{e}^*_{hi} - \mathbf{e}^*_{h\cdot})^T \tag{3.8}$$

where $\mathbf{e}^*_{h\cdot} = \sum_i \mathbf{e}^*_{hi}/n_h$. The formula (3.8) assumes that the first stage clusters are either drawn with replacement in each stratum or the first stage sampling fractions are negligible. In the case of nonresponse with weighting classes cutting across post-strata, the formula for $\mathbf{v}(\hat{\mathbf{S}})$ becomes more complicated (see Yung, 1996, Chapter 4).

An alternative version of $v_L(\hat{\boldsymbol{\beta}})$ is obtained by changing $\hat{\mathbf{J}}(\hat{\boldsymbol{\beta}})$ to $\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})$, where $\hat{\mathbf{I}}(\boldsymbol{\beta}) = E_m\hat{\mathbf{J}}(\boldsymbol{\beta})$ in (3.6). We suspect that $\hat{\mathbf{I}}(\hat{\boldsymbol{\beta}})$ is more stable than $\hat{\mathbf{J}}(\hat{\boldsymbol{\beta}})$. Note that $\hat{\mathbf{J}}(\boldsymbol{\beta}) = E_m\hat{\mathbf{J}}(\boldsymbol{\beta})$ for logistic regression with binary response, but this is not necessarily true under a working correlation structure on the repeated measurements.

It may be noted that post-stratification may not lead to increased efficiency because the model residuals $u_{hik\ell}(\hat{\boldsymbol{\beta}})$ may be unrelated to the post-stratifiers $\mathbf{z}_{hik}$, particularly when the model fits the data well.

The jackknife method can be used in a straightforward manner to estimate the variance matrix of $\hat{\boldsymbol{\beta}}$. An advantage of the jackknife is that post-stratification and unit-nonresponse adjustment are automatically taken into account, unlike the linearization method.

Using the estimated $\text{cov}(\hat{\boldsymbol{\beta}})$, Wald tests of hypothesis of the form $H_0 : \boldsymbol{\psi} = \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ are readily obtained, where $\mathbf{C}$ is a $r \times p$ full rank matrix of known constants and $\boldsymbol{\beta}$ is $p \times 1$ vector $(r < p)$. Under H_0,

$$X_W^2 = \hat{\boldsymbol{\psi}}^T(\mathbf{C}\mathbf{v}_L(\hat{\boldsymbol{\beta}})\mathbf{C}^T)^{-1}\hat{\boldsymbol{\psi}} \qquad (3.9)$$

is distributed asymptotically as χ_r^2, a χ^2 variable with r d.f., where $\hat{\boldsymbol{\psi}} = C\hat{\boldsymbol{\beta}}$. Therefore, the p-value associated with H_0 is computed as $P[\chi_r^2 > X_W^2 \text{ (obs)}]$, where X_W^2 (obs) is the observed value of the Wald statistic X_W^2. More general hypotheses of the form $H_0 : \boldsymbol{\psi} = \mathbf{h}(\boldsymbol{\beta}) = \mathbf{0}$ can also be tested using the Wald method, where $\mathbf{h}(\boldsymbol{\beta})$ is a $r \times 1$ vector. Under H_0, we have

$$X_W^2 = \hat{\boldsymbol{\psi}}^T\hat{\boldsymbol{\Sigma}}_{\psi}^{-1}\hat{\boldsymbol{\psi}} \qquad (3.10)$$

is asymptotically χ_r^2, where $\hat{\boldsymbol{\psi}} = \mathbf{h}(\hat{\boldsymbol{\beta}})$ and $\hat{\boldsymbol{\Sigma}}_{\psi} = \mathbf{H}(\hat{\boldsymbol{\beta}})\mathbf{v}_P(\hat{\boldsymbol{\beta}})\mathbf{H}(\hat{\boldsymbol{\beta}})^T$ with $\mathbf{H}(\boldsymbol{\beta}) = \partial\mathbf{h}(\hat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}^T$, a $r \times p$ full rank matrix.

4. Quasi-Score Tests Under Working Independence

For the Wald tests, we have to fit the full model $g(\mu_{it}) = x_{it}^T\boldsymbol{\beta}$ which could lead to unstable estimates if the full model contains a large number of terms. For example, with a factorial structure of explanatory variables containing a large number of interactions we may be interested in testing the significance of interaction effects, denoted as $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_{20} = \mathbf{0}$. On the other hand, for the quasi-score tests we need only to fit the simple null model, $g(\mu_{it}) = \mathbf{x}_{1it}^T\boldsymbol{\beta}_1$, where $\mathbf{x}_{it} = (\mathbf{x}_{1it}^T, \mathbf{x}_{2it}^T)^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$. Moreover, the quasi-score tests are invariant to nonlinear transformations of $\boldsymbol{\beta}$, unlike the Wald tests (Boos, 1992). Rao and Scott (1996) studied quasi-score tests in the context of cross-sectional survey data. It may be noted that score tests were first introduced in a seminal paper by C.R. Rao (1947).

Let $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1^T, \boldsymbol{\beta}_{20}^T)^T$ be the solution of $\hat{\mathbf{S}}_1(\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_{20}^T) = \mathbf{0}$, where $\hat{\mathbf{S}} = (\hat{\mathbf{S}}_1^T, \hat{\mathbf{S}}_2^T)$ is partitioned in the same way as $\boldsymbol{\beta}$. The analogue of the score test, called quasi-score test, is given by

$$X_S^2 = \tilde{\mathbf{S}}_2^T[\mathbf{v}(\tilde{\mathbf{S}}_2)]^{-1}\tilde{\mathbf{S}}_2, \qquad (4.1)$$

where $\tilde{\mathbf{S}}_2 = \hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ and $\mathbf{v}(\tilde{\mathbf{S}}_2)$ is a design-consistent estimator of $\text{cov}(\tilde{\mathbf{S}}_2)$. We now sketch a proof to show that X_S^2 is asymptotically χ_r^2 under H_0.

Expanding $\hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}})$ and $\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}})$ around the true value $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_{20}^T)^T$ gives

$$\mathbf{0} = \hat{\mathbf{S}}_1(\tilde{\boldsymbol{\beta}}) \approx \hat{\mathbf{S}}_1(\boldsymbol{\beta}^*) - \mathbf{J}_{11}^*(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*) \qquad (4.2)$$

and

$$\hat{\mathbf{S}}_2(\tilde{\boldsymbol{\beta}}) \approx \hat{\mathbf{S}}_2(\boldsymbol{\beta}^*) - \mathbf{J}_{21}^*(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^*) \qquad (4.3)$$

where $\mathbf{J}^* = \hat{\mathbf{J}}(\boldsymbol{\beta}^*)$ is the value of $\hat{\mathbf{J}}(\boldsymbol{\beta}) = -\partial\hat{\mathbf{S}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ and $\mathbf{J}^*$ is partitioned as

$$\mathbf{J}^* = \begin{bmatrix} \mathbf{J}_{11}^* & \mathbf{J}_{22}^* \\ \mathbf{J}_{21}^* & \mathbf{J}_{22}^* \end{bmatrix}.$$

Now replacing $\mathbf{J}^*$ by its expected value $\mathbf{I}^*$ and substituting for $\tilde{\beta}_1 - \beta_1^*$ from (4.2) into (4.3), we get

$$\tilde{\mathbf{S}} = \hat{\mathbf{S}}_2(\tilde{\beta}) \approx \hat{\mathbf{S}}_2(\beta^*) - \mathbf{I}_{21}^*\mathbf{I}_{11}^{*-1}\hat{\mathbf{S}}_1(\beta^*) = \sum_{hik \in s_r} w_{hik}^* \tilde{\mathbf{u}}_{2hik}(\beta^*), \quad (4.4)$$

where $\tilde{\mathbf{u}}_{2hik}(\beta^*) = \mathbf{u}_{2hik}(\beta^*) - \mathbf{A}^*\mathbf{u}_{1hik}(\beta^*)$ with $\mathbf{A}^* = \mathbf{I}_{21}^*\mathbf{I}_{11}^{*-1}$ and $\mathbf{u}_{hik} = (\mathbf{u}_{1hik}^T, \mathbf{u}_{2hik}^T)^T$. It follows from (4.4) that $\tilde{\mathbf{S}}_2$ is approximately equal to a vector of estimated totals so that $\tilde{\mathbf{S}}_2$ is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\mathrm{cov}(\tilde{\mathbf{S}}_2)$. Thus, X_S^2 is asymptotically χ_r^2 under H_0. Note that $E(\tilde{\mathbf{S}}_2) \approx \mathbf{0}$ under H_0.

Calculation of the quasi-score test X_S^2 requires an estimator of $\mathrm{cov}(\tilde{\mathbf{S}}_2)$. A jackknife estimator, $\mathbf{v}_J(\tilde{\mathbf{S}}_2)$ is obtained in a straight forward manner. The jackknife final weights, $w_{hik(gj)}^*$, when the (gj)-th sample cluster is deleted are obtained in the same manner as w_{hik}^*, using the jackknife basic weights $w_{hik(gj)} = w_{hik}b_{gj}$ where $b_{gj} = 0$ if $(hi) = (gj)$; $= n_g/(n_g - 1)$ if $h = g$ and $i \neq j$; $= 1$ if $h \neq g$. Replacing w_{hik}^* by $w_{hik(gj)}^*$, we get $\hat{\mathbf{S}}_{(gj)}(\beta)$, $\tilde{\beta}_{(gj)}$ and $\tilde{\mathbf{S}}_{2(gj)}(\tilde{\beta}_{(gj)})$. Using $\tilde{\mathbf{S}}_{2(gj)}$ we get

$$\mathbf{v}_J(\tilde{\mathbf{S}}_2) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} \left(\tilde{\mathbf{S}}_{2(gj)} - \tilde{\mathbf{S}}_2\right)\left(\tilde{\mathbf{S}}_{2(gj)} - \tilde{\mathbf{S}}_2\right)^T. \quad (4.5)$$

Computation of $\tilde{\beta}_{(gj)} = (\tilde{\beta}_{1(gj)}^T, \beta_{20}^T)^T$ can be simplified by performing only a single Newton-Raphson iteration for the solution of $\hat{\mathbf{S}}_{1(gj)}(\beta_1^T, \beta_{20}^T) = 0$, using $\tilde{\beta}$ as the starting value. The jackknife quasi-score test (4.1) is invariant to a one-to-one reparametrization of β with non-singular Jacobian, unlike the Wald test X_W^2.

A Taylor linearization estimator of cov $(\tilde{\mathbf{S}}_2)$, denoted as $\mathbf{v}_L(\tilde{\mathbf{S}}_2)$, can be obtained using the asymptotic representation (4.4) of $\tilde{\mathbf{S}}_2$ as a vector of estimated totals. We replace $\tilde{\mathbf{u}}_{2hik}(\beta^*)$ by $\tilde{\mathbf{u}}_{2hik}(\tilde{\beta}) = \mathbf{u}_{2hik}(\tilde{\beta}) - \hat{\mathbf{A}}\mathbf{u}_{1hik}(\tilde{\beta})$, where $\hat{\mathbf{A}}$ is an estimator of $\mathbf{A}^*$, and then use (3.8) with $\mathbf{u}_{hik}(\hat{\beta})$ changed to $\tilde{\mathbf{u}}_{2hik}(\tilde{\beta})$. There are several possible choices for $\hat{\mathbf{A}}$. It might seem natural to use $\hat{\mathbf{J}}(\tilde{\beta})$ in place of $\mathbf{I}^*$, where $\hat{\mathbf{J}}(\beta)$ is given by (3.7). For the special case of scalar β_2 (i.e., $r = 1$) and one time point (i.e., $T_i = 1$), Binder and Patak (1994) used this form of quasi-score test to construct confidence intervals for β_2, although their approach is different from that given here. This choice, however, does not have the desired invariance property in general. We can get an invariant quasi-score test by taking the expectation of $\hat{\mathbf{J}}(\beta)$ under the mean specification defined by (2.1), i.e., by using

$$\hat{\mathbf{I}}(\tilde{\beta}) = \sum_{hik \in s_r} \sum_t w_{hik}^* \mathbf{D}_{hikt}(\tilde{\beta})\mathbf{D}_{hikt}(\tilde{\beta})^T / V_0(\tilde{\mu}_{hikt}) \quad (4.6)$$

where $\mathbf{D}_{hikt}(\beta) = \partial\mu_{hikt}/\partial\beta$ with $\tilde{\mu}_{hikt}$ denoting the value of the mean μ_{hikt} at $\beta = \tilde{\beta}$. Moreover, the resulting test is likely to be more stable. Of course, for the binary response logistic regression case $\hat{\mathbf{I}}(\beta) = \hat{\mathbf{J}}(\beta)$.

Under the stratified multistage sampling setup it can be shown that $\mathbf{v}_J(\tilde{\mathbf{S}}_2) = \mathbf{v}_L(\tilde{\mathbf{S}}_2)$, so that the jackknife and Taylor linearization quasi-score tests are asymptotically equivalent.

More general hypotheses of the form $H_0 : \psi = \mathbf{h}(\beta) = \mathbf{0}$ can also be tested using the quasi-score method. The estimate $\tilde{\beta}$ under H_0 is obtained by solving

$$\hat{\mathbf{S}}(\beta) - \mathbf{H}(\beta)^T\boldsymbol{\lambda} = 0; \quad \mathbf{h}(\beta) = 0 \quad (4.7)$$

for β and $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ is the $r \times 1$ vector of Lagrange multipliers. Let $\tilde{\mathbf{S}}_h = \mathbf{H}(\tilde{\beta})[\hat{\mathbf{I}}(\tilde{\beta})]^{-1}\hat{\mathbf{S}}(\tilde{\beta})$, then the jackknife quasi-score is given by

$$X_S^2 = \tilde{\mathbf{S}}_h^T[\mathbf{v}_J(\tilde{\mathbf{S}}_h)]^{-1}\tilde{\mathbf{S}}_h, \quad (4.8)$$

where $\mathbf{v}_J(\tilde{\mathbf{S}}_h)$ is the jackknife estimator of cov $(\tilde{\mathbf{S}}_h)$ which is obtained in a straightforward manner from (4.5) by changing $\tilde{\mathbf{S}}_2$ to $\tilde{\mathbf{S}}_h$ and $\tilde{\mathbf{S}}_{2(gj)}$ to $\tilde{\mathbf{S}}_{h(gj)}$.

A Taylor linearization estimator of $\mathrm{cov}(\tilde{\mathbf{S}}_h)$, denoted as $\mathbf{v}_L(\tilde{\mathbf{S}}_h)$, can also be obtained using the following asymptotic representation of $\tilde{\mathbf{S}}_h$:

$$\tilde{\mathbf{S}}_h \approx \mathbf{H}^*\mathbf{I}^{*-1}\hat{\mathbf{S}}(\beta^*) - \mathbf{H}^*(\tilde{\beta} - \beta) = \mathbf{H}^*\mathbf{I}^{*-1}\hat{\mathbf{S}}(\beta^*), \quad (4.9)$$

noting that $\mathbf{0} = \mathbf{h}(\tilde{\boldsymbol{\beta}}) \approx \mathbf{h}(\boldsymbol{\beta}^*) + \mathbf{H}^*(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ so that $\mathbf{H}^*(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \approx 0$ under H_0,where $\mathbf{H}^* = \mathbf{H}(\boldsymbol{\beta}^*)$. Now letting $\tilde{\mathbf{u}}_{hik}(\tilde{\boldsymbol{\beta}}) = \mathbf{H}(\tilde{\boldsymbol{\beta}})\hat{\mathbf{I}}(\tilde{\boldsymbol{\beta}})^{-1}\mathbf{u}_{hik}(\tilde{\boldsymbol{\beta}})$, it follows from (4.9) that $\mathbf{v}_L(\tilde{\mathbf{S}}_h)$ is given by (3.8) with $\mathbf{u}_{hik}(\hat{\boldsymbol{\beta}})$ changed to $\tilde{\mathbf{u}}_{hik}(\tilde{\boldsymbol{\beta}})$, assuming complete response.

5. Working Correlation Structure

In this section we generalize the previous results on quasi-score tests to the case of a "working" correlation matrix of $\mathbf{y}_i$, assuming $T_i = T$. The working covariance matrix of $\mathbf{y}_{hik} = (y_{hik1}, \dots, y_{hikT})^T$ is assumed to be of the form $\mathbf{V}_{0hik} = A_{hik}^{\frac{1}{2}}\mathbf{R}\mathbf{A}_{hik}^{\frac{1}{2}}$ with common correlation structure across units (hik), i.e., $\mathbf{R}_{hik} = \mathbf{R}$, where $\mathbf{A}_{hik} = \text{diag}(V_{0hik1}, \dots, V_{0hikT})$ and $V_{0hikt} = \text{var}(y_{hikt})$.

We use $\tilde{\boldsymbol{\beta}}$, obtained under working independence and H_0, to get an estimator of $\mathbf{R}$:

$$\hat{\mathbf{R}}(\tilde{\boldsymbol{\beta}}) = \sum_{s_r} w_{hik}^* \tilde{\mathbf{R}}_{hik} / \sum_{s_r} w_{hik}^*, \qquad (5.1)$$

where $\tilde{\mathbf{R}})_{hik} = \mathbf{R}_{hik}(\tilde{\boldsymbol{\beta}})$ with

$$\mathbf{R}_{hik}(\boldsymbol{\beta}) \qquad (5.2)$$
$$= \mathbf{A}_{hik}^{\frac{1}{2}}(\mathbf{y}_{hik} - \boldsymbol{\mu}_{hik}(\boldsymbol{\beta}))(\mathbf{y}_{hik} - \boldsymbol{\mu}_{hik}(\boldsymbol{\beta}))^T \mathbf{A}_{hik}^{-\frac{1}{2}}.$$

Note that $\hat{\mathbf{R}}(\tilde{\boldsymbol{\beta}})$ is a design consistent estimator of the census parameter $\mathbf{R}_M = \sum \mathbf{R}_{hik}(\boldsymbol{\beta})/M$. Now using $\hat{\mathbf{R}}(\tilde{\boldsymbol{\beta}})$, we get

$$\mathbf{u}_{hik}^*(\boldsymbol{\beta}) = (\partial \boldsymbol{\mu}_{hik}^T / \partial \boldsymbol{\beta}) \tilde{\mathbf{V}}_{0hik}^{-1}(\mathbf{y}_{hik} - \boldsymbol{\mu}_{hik}), \qquad (5.3)$$

where $\tilde{\mathbf{V}}_{0hik} = \tilde{\mathbf{A}}_{hik}^{-\frac{1}{2}}\hat{\mathbf{R}}(\tilde{\boldsymbol{\beta}})\tilde{\mathbf{A}}_{hik}^{\frac{1}{2}}$. The results of Section 3 under working independence can be extended by changing $\mathbf{u}_{hik}(\boldsymbol{\beta})$ to $\mathbf{u}_{hik}^*(\boldsymbol{\beta})$ given by (5.3). The information matrix $\mathbf{I}(\boldsymbol{\beta})$ now changes to

$$\hat{\mathbf{I}}(\tilde{\boldsymbol{\beta}}) = \sum_{hik \in s_r} w_{hik} \mathbf{D}_{hik}(\tilde{\boldsymbol{\beta}}) \tilde{\mathbf{V}}_{0hik}^{-1} \mathbf{D}_{hik}(\tilde{\boldsymbol{\beta}})^T \qquad (5.4)$$

where

$$\mathbf{D}_{hik}(\boldsymbol{\beta}) = \partial \mathbf{u}_{hik}^T / \partial \boldsymbol{\beta}.$$

Properties of the resulting score tests are under investigation.

Liang and Zeger (1986) consider the case of general T_i, assuming working exchangeable correlation structure, moving average process (MA-1) or autoregressive process (AR-1). However, this approach can lead to inefficient estimators of $\boldsymbol{\beta}$ under misspecification of the correlation structure, as demonstrated by Sutradhar and Das (1998).

6. Concluding Remarks

The Wald and quasi-score tests become unstable if the effective degrees of freedom is small. In the context of stratified multistage sampling, effective degrees of freedom, f, is usually taken as the total number of sample primary units minus the number of strata. For a subgroup (or domain), f can be much less if the subgroup is not uniformly distributed across the samples primary units. If f is not large, an F-version of the Wald or quasi-score tests might perform better in controlling the size of the test. An F-version of the quasi-score test treats

$$F_S = [(f - r + 1)/(fr)]X_D^2 \qquad (6.1)$$

as an F-variable with r and $f - r + 1$ degrees of freedom.

Alternative Rao-Scott (1984) corrected tests or Bonferroni-t tests (Korn and Graubard, 1990) might perform better than X_S^2 or F_S when f is small. Rotnitzky and Jewell (1990) proposed Rao-Scott corrected score tests in the case of a simple random sample of subjects. We plan to study the properties of these alternative tests.

References

Binder, D.A. (1983). On the variance os asymptotically normal estimators from complex surveys. *Int. Statist. Rev.* **51**, 279-292.

Binder, D.A. (1998). Longitudinal Surveys: Why are these surveys different from all other surveys? *Survey Method*, **24**, in press.

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

Boos, D.D. (1992). On generalized score tests. *American Statistician*, **46**, 327-333.

Korn, E.L. and Graubard, B.I. (1990). Simultaneous testing of regression coefficients with complex survey data: use of Bonferroni t statistics. *Amer. Statist.*, **44**, 270-276.

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

Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proc. Comb. Phil. Soc.*, **44**, 50-57.

Rao, J.N.K. and Scott, A.J. (1984). On chi-squared tests for multiway contigency tables with cell proportios estimated from survey data. *Ann. Statist.*, **12**, 46-60.

Rao, J.N.K. and Scott, A.J. (1996). Quasi-score tests with survey data. *Proc. Survey Sec.*, Statistical Society of Canada Annual Meetings, 1996, 33-38.

Rotnitzky, A. and Jewell, N.P. (1990). Hypothesis testing in semi-parametric generalized linear models for cluster correlated data. *Biometrika*, **77**, 485-497.

Sutradhar, B.C. and Das, K. (1998). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika*, **85** (in press).

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