Estimation using Gaussian Replicates of the Pivotal based on Weighted Quasi-Score Vector

A.C. Singh[1] and C.J.J. Nadeau[2]
[1]Center for Excellence in Survey Research, NORC at the University of Chicago,
Chicago, IL 60603
[2]Statistics Canada, R.H. Coats, 16th Floor, Ottawa, ON K1T 4B3
singh-avi@norc.org

Abstract

The method of weighted quasi-likelihood (wql) is commonly used for point estimation in survey data analysis along with the Taylor method for variance estimation (VE) and Wald for interval estimation (IE). However, for small or moderate sample sizes, it is known that the above VE may be unstable and IE may have poor coverage properties. We consider ways to improve standard VE and IE by using Gaussian replicates of the pivotal estimating function (EF) derived from the wql-score vector using the method of randomly recentered estimating equations of Singh (2007). The basic idea is that the nonstudentized pivotal is expected to be closer to normal than the *wql*-estimator. The replicate estimates are obtained by solving the equation defined by setting the pivotal EF equal to random draws from the standard multivariate normal distribution. The computational problem of solving a multivariate estimating equation for each replicate may be quite cumbersome in the case of high dimensions. As an alternative, we propose an EF-based MCMC in a frequentist framework for this purpose which consists of solving repeatedly one-dimensional estimating equations and thus simplifies the computations considerably although it may be time consuming. The method is illustrated by an example of fitting a logit model to data from the Canadian Community Health Survey.

Key Words: Taylor Variance Estimate; Wald Interval Estimate; Estimating Function Based MCMC

1. Introduction

First consider simple surveys; i.e., the sample is simple random with replacement or it may be without replacement but with a negligible sampling fraction. For fitting a generalized linear model in a semiparametric framework (in the sense of up to first two moment assumptions), the method of quasi-likelihood based on estimating functions (EFs) is commonly used; see e.g., Wedderburn (1974), and McCullagh and Nelder (1989, Ch. 9). In particular, for a random sample of *n* binary observations *i*=1,2,…,*n*, conditional on covariates x_i 's, consider fitting a logit model given by

$$\begin{array}{l} y_i = \mu_i(\theta) + \varepsilon_i, \quad y_i \mid x_i \sim_{ind} Ber(\mu_i(\theta)) \\ \text{logit}(\mu_i(\theta)) = x_i'\theta_{p\times 1} \end{array}, \qquad (1.1)$$

where x_i is a *p*-vector of covariates, and the parameter θ is of dimension *p*. The *ql*-estimator of θ solves the equations obtained by setting the *ql*-score or EF vector $\psi_{ql(\theta)}$ to 0 where

$$\psi_{ql(\theta)} = (-\partial\mu(\theta)/\partial\theta')'V_{\varepsilon}^{-1}(y-\mu(\theta)) = \sum\nolimits_{i=1}^{n} x_i(y_i-\mu_i(\theta)), \qquad (1.2)$$

V_{ε} being the covariance matrix $\mathrm{diag}(\mu_i(\theta)(1-\mu_i(\theta))_{1\le i\le n}$ of the observation errors ε_i's. It turns out that in this case, the *ql*-estimator coincides with the maximum likelihood (*ml)* estimator. Note that if we allow for cluster-correlation in the data as is often the case in practice, then the *ql*-score vector under multiplicative overdispersion continues to be given by (1.2). However, now it is not possible to specify the joint likelihood and hence the *ml*-estimator without making any further assumptions about the intra-cluster dependence. For the example under consideration, the cluster correlation is assumed to be absent for simplicity. The point estimate (PE) $\hat{\theta}^{ql}$ can be obtained by solving (1.2) using an iterative method such as Newton-Raphson. The estimated covariance $\hat{\Sigma}_{ql(\theta)}$ of $\hat{\theta}^{ql}$ after Taylor linearization is obtained in a sandwich form as

$$\hat{\Sigma}_{ql(\theta)} = J_{ql(\theta)}^{-1}V_{ql(\psi)}J_{ql(\theta)}'^{-1}\Big|_{\hat{\theta}^{ql}}\ ; \quad J_{ql(\theta)} = (-\partial\psi_{ql(\theta)}/\partial\theta')', \qquad (1.3)$$

where $J_{ql(\theta)}$ is the $p\times p$ observed *ql*-information matrix computed as $\sum\nolimits_{1}^{n} u_i(\theta)x_ix_i'$ where $u_i(\theta)=\mu_i(\theta)(1-\mu_i(\theta))$ and $V_{ql(\psi)}$ is the $p\times p$ covariance matrix of the *ql*-EF vector $\psi_{ql(\theta)}$. Here $V_{ql(\psi)}$ coincides with $J_{ql(\theta)}$ because the link function is canonical, and the EF is optimal for the model in the sense that $\psi_{ql(\theta)}$ is closest to the conceptual (which may be unknown) *ml*-score vector by being the projection of the score vector on the elementary EFs $(y_i-\mu_i(\theta))_{1\le i\le n}$ under the covariance norm; see Godambe and Thompson (1989); also Singh and Rao (1997). In the above problem with binary data, the two score vectors (*ql- and ml-)* do of course coincide with each other.

Under regularity conditions, the estimator $\hat{\theta}^{ql}$ is consistent, and gives rise to a consistent estimator $\mu_i(\hat{\theta}^{ql})$ of the prevalence or the conditional mean given the covariate value. In this paper we will mainly be concerned with estimating such prevalence parameters for given covariate levels. Using Taylor, a consistent variance estimator (VE) $v(\mu_i(\hat{\theta}^{ql}))$ for $\mu_i(\hat{\theta}^{ql})$ can be obtained as

$$v(\mu_i(\hat{\theta}^{ql})) = (\partial\mu_i(\theta)/\partial\theta')\Sigma_{ql(\theta)}(\partial\mu_i(\theta)/\partial\theta')'\Big|_{\hat{\theta}^{ql}}\ ; \partial\mu_i(\theta)/\partial\theta' = u_i(\theta)x_i' \ .(1.4)$$

The VE $v(\mu_i(\hat{\theta}^{ql}))$ tends to be unstable (i.e., with high relative variance) for small *n* or for low or high prevalence outcomes. For interval estimate (IE) of $\mu_i(\theta)$ under *ql*-estimation, one can use the Wald IE -- a normality based symmetric interval given by

$$\mu_i(\hat{\theta}^{ql}) \pm z_{\alpha/2}\sqrt{v(\mu_i(\hat{\theta}^{ql}))} \qquad (1.5)$$

which also tends to show undercoverage for small n or for low or high prevalence outcomes due to inadequate (i.e., unbalanced tails) normal approximation caused by nonlinearity of $\mu_i(\hat{\theta}^{ql})$, the main reason being that the skewness in the distribution of estimated prevalence may be marked. Moreover, the interval boundaries may not lie in the admissible range of (0,1); the latter problem can, however, be addressed by using logit-Wald (cf. Singh and Nadeau, 2008) resulting in asymmetric IE which, although conservative, tends to improve coverage.

For complex survey data, we can use the optimal weighted quasi-likelihood (*wql*) method of Godambe and Thompson (1986). For the logit model (1.1) with two parameters of intercept θ_0 and slope θ_1, say, the *wql*-EFs for the parameters are given by

$$\psi_{wq(\theta_0)} = \sum_{i=1}^{n} w_i(y_i - \mu_i(\theta)), \quad \psi_{wq(\theta_1)} = \sum_{i=1}^{n} w_i x_i(y_i - \mu_i(\theta)), \qquad (1.6)$$

where w_i's denote the design weights. Note that we could have also used the pseudo-maximum likelihood (*pml*) estimator as described in Binder (1983). The two estimators (*wql* and *pml*) are generally identical except that *wql*-method does not require a likelihood for the superpopulation model. A consistent estimate $\hat{V}_{wq(\psi)}$ of the covariance matrix $V_{wq(\psi)}$ of the EF-vector $\psi_{wq(\theta)}$ can be obtained using standard survey sampling methods where the unknown θ-parameter is evaluated at the *wql*-estimator $\hat{\theta}^{wql}$ computed as a solution of (1.6). Analogous to (1.3), the estimated covariance matrix $\hat{\Sigma}_{wq(\theta)}$ of $\hat{\theta}^{wql}$ is obtained after Taylor linearization in a sandwich form (Binder, 1983), and is given by

$$\hat{\Sigma}_{wq(\theta)} = J^{-1}_{wq(\theta)} V_{wq(\psi)} J'^{-1}_{wq(\theta)} \Big|_{\hat{\theta}^{wql}}; \quad J_{wq(\theta)} = (-\partial \psi_{wq(\theta)} / \partial \theta')', \qquad (1.7)$$

where the observed *wq*-information matrix $J_{wq(\theta)}$ is similar to the matrix in (1.3) except that sampling weights w_i's are appropriately inserted. Now the *wql*-Wald or logit Wald method for IE can be applied using the approximate distribution of $\hat{\theta}^{wql}$ as $N_p(0, \hat{\Sigma}_{wq(\theta)})$.

As an alternative to *ql*- or *wql*- estimation, the method of randomly recentered estimating equations (RREE) of Singh (2007) (see also Singh and Nadeau, 2008, for the case of multidimensional parameters) may be considered with the goal of improving the performance of VE and IE for finite samples. The RREE method consists of creating parameter estimates by solving the standardized (but nonstudentized) EF vector centered at the random vector of values drawn from the pivotal normal distribution. The basic idea is that the nonstudentized pivotal is expected to be closer to normal than the *wql*-

estimator. For survey data, the covariance matrix $\hat{V}_{wq(\theta)}$ is smoothed using design-effect (deff)-type ideas so that it can be expressed analytically as a function of the parameters. The Monte Carlo distribution of parameter estimates or the empirical confidence distribution so obtained is used to compute new PE, VE, and IE.

In the multi-parameter case, the computation for implementing RREE could, however, be quite demanding as it requires in general solving a set of nonlinear equations (number of equations correspond to the dimension of the parameter) repeatedly for a large number of recenters, and this is in addition to the iterations required for solving each equation. Even for the usual *ql*-estimation in the case of a high dimensional parameter, although we need to solve for only one value (i.e., 0) of the center, the computation can be tedious and methods such as modified Gauss-Seidel are proposed for computational simplicity so that only one equation is solved at a time; see Jiang (2000). In this paper, we propose a new frequentist application of Markov Chain Monte Carlo (MCMC) to EFs so that for each value of the recenter vector (i.e., the right hand side of the estimating equation), only one equation is solved at a time while holding other parameters fixed at their current values. The resulting chain or the sequence of parameter estimates after a large number of cycles yields the empirical joint stationary distribution or the confidence distribution of the parameter vector from which PE, VE, and IE can be obtained. Note that unlike the usual MCMC applications, here the joint frequentist distribution of estimated parameters is in fact approximately known to be normal and we don’t have the integration problem for finding marginals. Instead, we want to obtain an alternative (empirical) approximation to the joint distribution using RREE for improved finite sample performance, and in doing so we want to work with one EF at a time conditional on other parameters which is somewhat analogous to drawing realizations from full conditionals in GIBBS sampling under MCMC. This is the motivation of the proposed method termed EF-MCMC.

The organization of this paper is as follows. Section 2 contains a background review for RREE for simple surveys, while the proposed method of EF-MCMC is described in Section 3. EF-MCMC for the problem of complex surveys is considered in Section 4 along with an illustrative application to the data from the 2001 Canadian Community Health Survey (CCHS). Empirical results based on a simulation study to compare RREE with and without MCMC are presented in Section 5. Section 6 contains concluding remarks.

2. Background Review: RREE for Simple Surveys

In RREE, we start with a pivotal $H_{ql(\psi)}^{-1}\psi_{ql(\theta)}$ based on the *ql*-EF $\psi_{ql(\theta)}$ such that even for moderate *n*,

$$H_{ql(\psi)}^{-1}\psi_{ql(\theta)} \sim_{approx} N(0, I_{p\times p}) , \qquad (2.1)$$

where $H_{ql(\psi)}$ is the Cholesky root of the covariance matrix $V_{ql(\psi)}$; i.e., $V_{ql(\psi)} = H_{ql(\psi)}H'_{ql(\psi)}$. Under *ql*-estimation, $\hat{\theta}^{ql}$ is obtained by solving $H_{ql(\psi)}^{-1}\psi_{ql(\theta)} = 0$, or $\psi_{ql(\theta)} = 0$, while under RREE, a large number *R* of replicate parameter estimates $\{\tilde{\theta}_r\}_{1\le r\le R}$ are obtained by randomly recentering the estimating equations; i.e., by solving the *p* equations for each *r*,

$$H_{ql(\psi)}^{-1}\psi_{ql(\theta)} = \varepsilon_r ; \quad \varepsilon_r \sim_{iid} N_p(0, I) \quad . \qquad (2.2)$$

The empirical distribution $\{\tilde{\theta}_r\}_{1\le r\le R}$ so obtained gives rise to new PE, VE, and IE for θ, and any function of it; see Singh (2007) for theoretical details. In fact, in certain special cases, one can show analytically that although both PE and VE based on RREE may have higher relative biases than that under *ql*-estimation, but rather interestingly they have smaller relative MSE (mean square error); the bias decreasing with larger sample sizes as expected. However, it is IE where RREE has in general much improved finite sample property in terms of central and tail coverage probabilities; the reason being that the (nonstudentized) pivotal $H_{ql(\psi)}^{-1}\psi_{ql(\theta)}$ is expected to be closer to normal than the commonly used pivotal $H_{ql(\psi)}^{*-1}\psi_{ql(\theta)}$ under *ql*-estimation where $H_{ql(\psi)}^{*}$ is $H_{ql(\psi)}$ evaluated at $\hat{\theta}^{ql}$; see e.g., McCullagh (1991), and Godambe and Thompson (1999). For example, in the case of estimating a proportion θ, the superior performance of the Wilson IE based on the pivotal $\sqrt{n}(\bar{y}-\theta)\big/\sqrt{\theta(1-\theta)}$ over the commonly used Wald IE based on the pivotal $\sqrt{n}(\bar{y}-\theta)\big/\sqrt{\bar{y}(1-\bar{y})}$ is well documented in Cai, Brown and Dasgupta (2001).

With RREE, it is possible to truncate the empirical distribution of $\hat{\theta}^{ql}$ (although the theoretical distribution is approximately normal) by discarding those recenters ε_r's that give rise to infeasible or nonexistent solutions to the estimating equations. Moreover, before computing PE and VE, it may be desirable in practice to trim extreme replicate values in $\{\tilde{\theta}_r\}_{1\le r\le R}$ for which at least one element of the *p*-vector lies outside the interval given by $\text{median} \pm 2.5(IQR)$; IQR denoting the inter-quartile range. Such trimming helps to robustify PE and VE against extreme values. However, trimming is not needed for IE as it is generally not affected by extreme values. The empirical distribution $\{\tilde{\theta}_r\}_{1\le r\le R}$ gives rise to empirical distributions of other functions of $\hat{\theta}^{ql}$ such as the conditional predictive mean or the prevalence $\mu_i(\hat{\theta}^{ql})$ as $\{\mu_i(\tilde{\theta}_r)\}_{1\le r\le R}$ at the covariate value x_i of the logit model mentioned earlier.

For a simple example of a multi-parameter RREE, consider the logit model with a single covariate and the intercept. Then

$$\mu_i(\theta) = \exp(\theta_0 + \theta_1 x_i)[1 + \exp(\theta_0 + \theta_1 x_i)]^{-1} \qquad (2.3)$$

and the corresponding *ql*- EFs based on a random sample of *n* observations are

$$\psi_{ql(\theta_0)} = \sum_{i=1}^{n} (y_i - \mu_i(\theta)), \quad \psi_{ql(\theta_1)} = \sum_{i=1}^{n} x_i (y_i - \mu_i(\theta)) \, . \qquad (2.4)$$

The covariance matrix $V_{ql(\psi)}$ is given by

$$V_{ql(\psi)} = \begin{pmatrix} \sum_1^n u_i(\theta) & \sum_1^n x_i u_i(\theta) \\ \sum_1^n x_i u_i(\theta) & \sum_1^n x_i^2 u_i(\theta) \end{pmatrix}, \quad u_i(\theta) = \mu_i(\theta)(1 - \mu_i(\theta) \qquad (2.5)$$

Now, the replicate values of the vector parameter estimate $\hat{\theta}^{ql}$ are obtained by solving iteratively

$$\begin{pmatrix} \psi_{ql(\theta_0)} \\ \psi_{ql(\theta_1)} \end{pmatrix} = H_{ql(\psi)} \begin{pmatrix} \varepsilon_{1,r} \\ \varepsilon_{2,r} \end{pmatrix}, \qquad (2.6)$$

where $\{\varepsilon_{1,r}, \varepsilon_{2,r} : 1 \le r \le R\}$ are independent standard normal deviates. For computational convenience, one can first evaluate $H_{ql(\psi)}$ at the initial value $\hat{\theta}^{ql}$ to compute the next iterative value of θ serving as the current value. Next $H_{ql(\psi)}$ is evaluated at the current value of θ, and the process is repeated until convergence to obtain $\tilde{\theta}_r$.

3. Proposed Method of EF-MCMC for Implementing Multi-dimensional RREE

We propose a new application of MCMC for large sample frequentist estimation problems using RREE. First consider simple surveys for the sake of simplicity. The case of complex surveys follows in an analogous manner as described in the following Section 4. Before we consider MCMC for EFs, it is useful to note the approximate equivalence of confidence distributions of the parameter θ obtained from the EF-based pivotal and the estimator-based pivotal. The confidence distribution of θ (a frequentist concept) needed for RREE can be viewed as being analogous to an empirical posterior distribution of θ in a Bayesian framework. Although the estimator-based pivotal is not of interest, it would be useful to consider RREE based on this pivotal to motivate the proposed method. Since the joint confidence distribution of θ obtained from $\hat{\theta}^{ql}$ is proper, it follows that it can be determined jointly by all full conditional confidence distributions. Now, partitioning θ as $(\theta_1, \theta_2{}')'$ where $\theta_1$denotes any one element of θ, and θ_2 the remainder, the full conditional confidence distribution of θ_1 given θ_2 is easily obtained as a univariate normal $N(\hat{\theta}_{1.2}, \Sigma_{11.2})$ where

$$\hat{\theta}_{1.2} = \hat{\theta}_1^{ql} + \Sigma_{12}\Sigma_{22}^{-1}(\theta_2 - \hat{\theta}_2^{ql}),\ \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \qquad (3.1)$$

and the matrices $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ in customary notation denote the partitioned submatrices of the covariance matrix $\Sigma_{ql(\theta)}$of (1.3). Consider the sequence of values of elements of the θ-vector drawn one at a time from (3.1) while keeping other elements fixed at their current values; here $\hat{\theta}^{ql}$ could serve as the initial starting value. Following Casella, Ravine, and Robert (2001), since the above sequence forms an aperiodic and irreducible Markov Chain and the joint distribution is proper, the observed chain is convergent to its stationary distribution. Thus we have a GIBBS version of MCMC based on the estimator-based pivotal.

However, our goal is to obtain the confidence distribution of θ using MCMC based on the EF-based pivotal instead. In the parametric case, the optimal EF for θ_1 given θ_2 is the conditional score function which is asymptotically sufficient for θ_1. So heuristically, we would expect that the pivotal based on the conditional score function for $\theta_1$would give rise to a confidence distribution of $\theta_1$that is approximately equivalent to the corresponding conditional confidence distribution based on the maximum likelihood estimator. In the semi-parametric case, this suggests that the conditional confidence distribution of $\theta_1$given θ_2 based on the optimal *ql*-score function of θ_1 given θ_2 would be approximately equivalent to $N(\hat{\theta}_{1.2}, \Sigma_{11.2})$ obtained from the *ql*-estimator. This is indeed true and can be seen as follows.

First observe that the conditional confidence distribution of θ_1 given θ_2 based on the estimator $\hat{\theta}^{ql}$ can be obtained from the corresponding marginal of $\Sigma_{ql(\theta)}^{-1}(\hat{\theta}^{ql} - \theta)$ in view of (3.1) because

$$\Sigma_{ql(\theta)}^{-1}(\hat{\theta}^{ql} - \theta) = \begin{pmatrix} \Sigma_{11.2}^{-1}(\hat{\theta}_1^{ql} - \theta_1) - \Sigma_{11.2}^{-1}\Sigma_{12}\Sigma_{22}^{-1}(\hat{\theta}_2^{ql} - \theta_2) \\ \Sigma_{22.1}^{-1}(\hat{\theta}_2^{ql} - \theta_2) - \Sigma_{22.1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}(\hat{\theta}_1^{ql} - \theta_1) \end{pmatrix} = \begin{pmatrix} \Sigma_{11.2}^{-1}(\hat{\theta}_{1.2} - \theta_1) \\ \Sigma_{22.1}^{-1}(\hat{\theta}_{2.1} - \theta_2) \end{pmatrix}, \qquad (3.2)$$

Now using the relation $\left(\hat{\theta}^{ql}-\theta\right) \approx \mathcal{J}^{-1}_{ql(\theta)}\psi_{ql(\theta)}$ where $\mathcal{J}_{ql(\theta)}$ is the expected information matrix $E(\mathrm{J}_{ql(\theta)})$, and (1.3), we have

$$\Sigma^{-1}_{ql(\theta)}\left(\hat{\theta}^{ql}-\theta\right) \approx \Sigma^{-1}_{ql(\theta)}\mathcal{J}^{-1}_{ql(\theta)}\psi_{ql(\theta)} = \mathcal{J}'_{ql(\theta)}\, V^{-1}_{ql(\psi)}\psi_{ql(\theta)} \quad . \qquad (3.3)$$

From (3.3) it follows that for optimal *ql*-score functions or EFs $\psi_{ql(\theta)}$, the full conditional confidence distributions of parameters based on the EF-pivotal can be obtained from the corresponding marginals of $\psi_{ql(\theta)}$ because of information unbiasedness $V_{ql(\psi)} = \mathcal{J}_{ql(\theta)}$. It also follows that for nonoptimal *ql*-score functions $\psi_{ql(\theta)}$, the corresponding marginals of the transformed EF-vector $\mathcal{J}'_{ql(\theta)}\, V^{-1}_{ql(\psi)}\psi_{ql(\theta)}$ provide the desired full conditional distributions as the information unbiasedness property is now satisfied ; see Godambe and Thompson (1989). It is also observed from (3.3) that it is important to match the order of the parameter with the corresponding order of the EF to get the correct marginal distributions.

Thus for the two-parameter logit model (2.3), the *ql*-score functions are given by (2.4) which happen to be optimal and the corresponding covariance matrix by (2.5). So starting with the initial value $\hat{\theta}^{ql}$, the steps of MCMC for cycle $r+1$ consist of the following two steps where $(\theta_{0r\,,}\,\theta_{1r})$ denote the realized values of θ after cycle r.

Step I: Solve $\psi_{ql(\theta_0)} = \varepsilon_{1r}\sqrt{\sum_{i=1}^{n} u_i(\theta)\,|_{\theta_r}}$ iteratively to obtain $\theta_{0(r+1)}$.

Step II: Set $\theta_0 = \theta_{0(r+1)}$, and solve $\psi_{ql(\theta_1)} = \varepsilon_{2r}\sqrt{\sum_{i=1}^{n} x_i^2 u_i(\theta)\,|_{(\theta_{0(r+1)\,,}\,\theta_{1r})}}$ to obtain $\theta_{1(r+1)}$.

This completes one cycle. Repeat it many times allowing for a burn-in period and then take every 10th or so to obtain a set of *R* realizations from the confidence distribution of θ. The desired PE, VE, and IE can then be computed. Incidentally, in generating cycles of MCMC, we discard those values of ε that do not lead to convergence.

4. EF-MCMC for Complex Survey Data

We will describe EF-MCMC for complex survey data in terms of an example from CCHS where one may be interested in modeling the prevalence of healthy life style behavior outcomes such as smoking habit (a binary variable) using covariates such as age, gender, and education. Consider the data of cycle 1.1 of the Canadian Community Health Survey (CCHS) conducted in 2000-2001 whose goal was to collect general health information at the health Region level, a sub-provincial level of geography (Béland, 2002). The target population is all persons aged 12 years or older living in private dwellings in the ten provinces and three territories. The sample design is fairly complex involving stratified multi-stage cluster sampling. For the RREE application, suppose the parameter of interest is the proportion of smokers among 18 years old for the Yukon territory. For Yukon, the sample size was 809 while the population size was 24937. For fitting a two parameter logit model, the *wql*-EFs were given earlier in Section 1 by (1.6) and VE under *wql*-estimation given by (1.7).

To apply RREE to survey data, we need to express $V_{wq(\psi)}$ as a function of all model parameters under consideration. In practice, it may be computationally tedious because of the need to have a design-based estimate of $V_{wq(\psi)}$ iteratively for finding replicate parameter estimates. To alleviate this problem, we can use a smoothed version $\bar{V}_{wq(\psi)}$ of $V_{wq(\psi)}$ based on a working covariance matrix whose variances are adjusted by variance-deffs and correlations by correlation-deffs after using Fisher's z-transformation on correlations where deff , denoting design effect, is defined in the usual manner as a multiplicative adjustment factor for variance, but is defined as an additive adjustment for the transformed correlation. This is explained below. Under the assumption of design ignorability for the model (1.1), a working covariance $V^*_{wq(\psi)}$ as an alternative to $V_{wq(\psi)}$, can be obtained under the model while conditioning on the selected units in the sample as

$$V^*_{wq(\psi)} = \begin{pmatrix} \sum_1^n w_i^2 u_i(\theta) & \sum_1^n w_i^2 x_i u_i(\theta) \\ \sum_1^n w_i^2 x_i u_i(\theta) & \sum_1^n w_i^2 x_i^2 u_i(\theta) \end{pmatrix}, \qquad (4.1)$$

Now the diagonals of the above matrix involve variances and off-diagonals involve correlations and square-roots of variances. We multiply all variances by corresponding variance-deffs which are defined as ratios of the variances $v_{wq(\psi_j)}$ and $v^*_{wq(\psi_j)}$ for corresponding *wq*-score functions; the variance-deffs are evaluated only once at the consistent estimator $\hat{\theta}^{wql}$ and not for each replicate. Similarly, denoting by $F(\rho)$ the Fisher's z-transformation $(1/2)\log((1-\rho)^{-1}(1+\rho))$, we adjust the function $F(\rho^*)$ by adding the term $\left(F(\rho)-F(\rho^*)\right)$ evaluated at $\hat{\theta}^{wql}$ where ρ denotes one of the possible correlations between *wq*-score functions ψ_j and $\psi_{j'}$; the adjusted function is transformed back using inverse Fisher's z. Now RREE without MCMC can be applied as before using the normal approximation $N_p(0,\bar{V}_{wq(\theta)})$ for the distribution of $\psi_{wq(\theta)}$. RRRE with MCMC also follows along the same lines as described in Section 3 for simple surveys except that the *wql*-EFs are not optimal (as they do not use second moment assumptions) and so the EF-vector will need to be transformed first to make it information unbiased before defining the pivotal.

Table 1 illustrates the results obtained under *wql*-estimation and under RREE both with and without MCMC. Note this is only an illustration of the EF-MCMC methodology because the dimension of the parameter here is not high. The number of cycles used in MCMC was 21000 with first 1000 discarded for burn-in and 1 in 20 realizations were retained for the empirical joint confidence distribution. However, in the case of RREE without MCMC, 2000 replicates were used which were the same as reported in Singh and Nadeau (2008). It is of interest to compare results under the wrong assumption of simple designs. As expected estimates appear more precise than they really are assuming simple designs. The results for different methods are very similar because the prevalence parameter of interest is not low and the deff-adjusted sample size is still quite large.

Table 1: PE, VE, and IE for $\boldsymbol{\mu(\theta, age = 18)}$
(2001 CCHS Yukon Data)

Method	**Simple Design Assumed**			**Complex Design of CCHS**		
	PE in %	**VE x10^4**	**IE in %**	**PE in %**	**VE x10^4**	**IE in %**
QL(Wald)	28.95	6.91	(23.79, 34.10)	24.93	10.17	(18.67, 31.18)
QL(Logit Wald)	28.95	6.91	(24.08, 34.36)	24.93	10.17	(19.21, 31.68)
RREE	29.01	7.11	(23.97, 34.32)	25.05	10.34	(19.09, 31.64)
RREE (MCMC)	28.87	7.45	(23.76, 34.60)	24.85	10.66	(18.99, 31.94)

5. Simulation Results

A simulation study was conducted for a four parameter logit model for binary data where the first parameter corresponds to the intercept, second for the covariate x for i=1, 2, …,n was defined as the centered version of $x_i = \min\{1, (\text{mod}(i,10)+0.5)/10\}$; i.e., it takes values from -.45 to .45 in increments of .10 and then repeats itself, the third for the covariate being a dummy variable taking the value of 1 for the odd observation and 0 otherwise, and fourth for the interaction between the second and third covariates. The θ-vector was chosen as (-0.75, 3, 0,0) and the covariate values as (1,-.03, 0,0) which define the value of the prevalence parameter of interest. The sample size n was set at 10, 20, 30, 50, and 100, the number M of simulation runs at about 2500, and the number R of recenters for each simulation at 1000. For MCMC version of RREE, 21000 cycles were run for each simulation and after discarding the first 1000 for the burn-in period, every 20th cycle was retained. Table 2 compares results for five methods, QL (quasi-likelihood consisting of the solution of *ql*-EF at 0 for PE, and Taylor linearization for VE), RREE, RREE/MCMC, QL-RREE and QL-RREE/MCMC. The composite methods QL-RREE and QL-RREE/MCMC are composite in the sense that while the PE is based on QL, the VE is based on RREE and RREE/MCMC replicates. Extreme replicate estimates were trimmed using the rule of median ± 2.5(IQR) for all elements of the parameter vector for RREE methods. The abbreviation ME in Table 2 stands for MSE estimator which is the same as VE but the RB and RRMSE are computed by treating it as ME.

Table 2: % Relative Bias and Relative Root MSE of PE and ME for $\boldsymbol{\mu}(\boldsymbol{\theta}, \boldsymbol{x})$

	n=10				**n=30**			
	PE		**ME**		**PE**		**ME**	
Method	**RB**	**RRMSE**	**RB**	**RRMSE**	**RB**	**RRMSE**	**RB**	**RRMSE**
QL	81.39	158.11	23.63	74.81	8.23	84.28	7.54	72.57
RREE	98.25	141.80	-9.78	23.59	18.59	76.61	6.95	47.65
RREE/ MCMC	95.67	155.41	-16.78	35.53	15.55	83.51	-7.74	54.11
QL-RREE			-27.44	32.42			-11.63	40.65
QL-RREE/ MCMC			-19.59	36.05			-9.43	53.41

The main points to observe is that the performance of RREE and RREE/MCMC are generally similar except for RB for ME for both sample sizes where RREE/MCMC tends to provide smaller values than those for RREE. This needs to be investigated further. However, as expected RREE provides more stable estimate of ME than QL. In fact, it follows from the behavior of composite methods that RREE provides considerably more stable estimate of MSE of the usual PE (based on QL) compared to the Taylor method. More detailed results including those for IE are currently under investigation.

6. Concluding Remarks

In this paper we presented a computationally simpler alternative based on EF-MCMC to implement the method of RREE when the parameter dimension is not low. It is a new application of MCMC and different from the usual one where numerical integration to obtain marginals of the posterior is computationally complex. Here, instead the problem is to find marginals of the empirical confidence distribution as solutions of EFs. The EF-MCMC consists of solving one-dimensional equations at each step within MCMC cycles.

It is remarked that the EF-MCMC method may not be economical in time, but it does involve much simpler computational steps which may be attractive in practice.

References

Béland, Y. (2002). Canadian Community Health Survey – Methodological overview. *Health Reports*, Vol 13, No.3, 9-14 (Statistics Canada, Catalogue no. 82-003).

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

Brown, L.D., Cai, T and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci.*, Vol 16, No.2, 101-133.

Casella, G.C., Lavine, M., and Robert, C.P. (2001). Explaining the perfect sampler. *Amer. Statist*., 55, 299-305.

Godambe, V.P. and Thompson, M.E. (1986) Parameters of superpopulation and finite population: their relationship and estimation. *Int. Statist. Rev.*, 54, 37-59.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasi-likelihood estimation (with discussion). *J. Statist. Plan. Inf.*, 12, 137-72.

Godambe, V.P. and Thompson, M.E. (1999). A new look at confidence intervals in survey sampling. *Survey Methodology*, Vol. 25, No. 2, 161-174.

Jiang, J. (2000). A nonlinear Gauss-Seidel algorithm for inference about GLMM, *Computational Statistics*, 15, 229-241.

McCullagh, P. (1991). Quasi-likelihood and estimating functions. *Statistical Theory and Modeling:* In honor of Sir David Cox, FRS, ed. D.V. Hinkley, N. Reid, and E.J. Snell, Chapman and hall, London, 265-286.

McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models*, 2nd Ed., London: Chapman and Hall.

Singh, A.C. (2007). A new application of estimating functions to point, variance, and interval estimation for simple and complex surveys. *Proc. Fed. Comm. Statist. Meth*., Washington, DC, www.fcsm.gov.

Singh, A.C. and R.P. Rao (1997). Optimal Instrumental Variable Estimation for Linear Models with Stochastic Regressors Using Estimating Functions. In Basawa, I.V., V.P. Godambe, and R.L. Taylor (Eds.), *Selected Proceedings of the Symposium on Estimating Functions,* IMS Lecture Notes-Monograph Series, Vol. 32, pp. 177-192.

Singh, A.C. and Nadeau, C.J.J. (2008). An Alternative to the Logit-Wald Method for Inference under Models for Proportions, *ASA Proc. Surv. Res. Meth. Sec*.