Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2003: Challenges in Survey Taking for the Next Decade

2003

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada Symposium 2003
Challenges in Survey Taking for the Next Decade

ESTIMATING FIXED EFFECTS AND VARIANCE COMPONENTS IN A RANDOM INTERCEPT MODEL USING SURVEY DATA

Yong You, J.N.K. Rao and Milorad Kovacevic[1]

ABSTRACT

The random intercept model is often used in model-based survey data analysis. In this paper, an iterative weighted estimating equations (IWEE) approach is developed to estimate the fixed effects and the variance components in the random intercept model using sampling weights. The proposed IWEE method is an extension of the method proposed by You and Rao (2002) for the fixed effects estimation and the method proposed by Waclawiw and Liang (1993) for the variance components estimation. A simulation study and an example based on a real survey data are presented to illustrate the proposed method. Our study has shown that the proposed IWEE performs very well and converges quickly. An advantage of the proposed method is that only final sampling weights are needed in the estimation procedure, unlike other methods available in the literature requiring higher level sampling weights such as primary sampling unit weights, which may not be available to users in practice.

KEYWORDS: Estimating Equations; Nested Error Regression Model; Sampling Weigthts.

1. INTRODUCTION

The random intercept model is often used in model-based survey data analysis including small area estimation. Conventional methods of estimating the regression parameters (fixed effects) and the variance components in the model ignore the sampling weights; for example, the method of fitting-of-constants, maximum likelihood (ML) and residual maximum likelihood (REML). In this paper, we propose an iterative weighted estimating equations (IWEE) approach to estimate the fixed effects and the variance components in the random intercept model, taking account of sampling weights. This method is an extension of the method proposed by You and Rao (2002) for the fixed effects estimation and the method proposed by Waclawiw and Liang (1993) for the variance components estimation. It updates estimates of the fixed effects and variance components alternatively until convergence is achieved. Thus, we can simultaneously obtain the sampling weighted estimates of the fixed effects and variance components. A small simulation study and an example based on a real survey data are also presented to illustrate the proposed method.

Let y_{ij} be the value of variable of interest for the j-th unit in the i-th group (area), and $x_{ij} = (x_{ij1},...,x_{ijp})'$ with $x_{ij1} = 1$ be the vector of auxiliary variables associated with y_{ij} ($i = 1,...,m; j = 1,..., N_i$). A random intercept population model based on $\{y_{ij}, x_{ij}\}$ is given by

$$y_{ij} = x'_{ij}\beta + v_i + e_{ij}, \quad j = 1,..., N_i, i = 1,...,m, \tag{1}$$

where $\beta = (\beta_0,...,\beta_{p-1})'$ is the $p \times 1$ vector of fixed effects or regression parameters, v_i is a random effect associated with the i-th group, and N_i is the number of population units in the i-th group. The random effects v_i are assumed to be iid $N(0,\sigma_v^2)$ and independent of the unit level errors e_{ij}, which are assumed to be iid $N(0,\sigma_e^2)$.

We assume that samples are drawn independently across groups according to a specified sampling design. The sample data $\{y_{ij}, x_{ij}, j = 1,..., n_i; i = 1,..., m\}$ is assumed to obey the population model, i.e.,

[1] Yong You, Household Survey Methods Division, Statistics Canada, Ottawa, Canada, K1A 0T6; J.N.K. Rao, School of mathematics and Statistics, Carleton University, Ottawa, K1S 5B6; Milorad Kovacevic, Social Survey Methods Division, Statistics Canada, Ottawa, Canada, K1A 0T6.

$$y_{ij} = x'_{ij}\beta + v_i + e_{ij}, \quad j = 1,...,n_i, i = 1,...,m, \tag{2}$$

where n_i is the sample size in the i-th group. This implies that the sampling design within groups is ignorable or selection bias is absent. Model (2) is also known as the nested error regression model (Battese, Harter and Fuller, 1988). The main purpose of this paper is to estimate the regression parameters β and the variance components σ_e^2 and σ_v^2, using model (2) and the sampling weights.

The unit level model (2) may be aggregated to group level by using direct survey estimators. Let $\tilde{w}_{ij}$ be the basic sampling weight attached to y_{ij}. A direct design-based estimator of the group level population mean is given by

$$\bar{y}_{iw} = \frac{\sum_{j=1}^{n_i} \tilde{w}_{ij} y_{ij}}{\sum_{j=1}^{n_i} \tilde{w}_{ij}} = \sum_{j=1}^{n_i} w_{ij} y_{ij} ,$$

where $w_{ij} = \tilde{w}_{ij} / \sum_{j=1}^{n_i} \tilde{w}_{ij} = \tilde{w}_{ij} / \tilde{w}_{i\cdot}$ and $\sum_{j=1}^{n_i} w_{ij} = 1$. Following You and Rao (2002, 2003), we obtain the following aggregated model from the unit level model (2):

$$\bar{y}_{iw} = \bar{x}'_{iw}\beta + v_i + \bar{e}_{iw}, \quad i = 1,...,m, \tag{3}$$

where $\bar{e}_{iw} = \sum_{j=1}^{n_i} w_{ij} e_{ij}$ with $E(\bar{e}_{iw}) = 0$ and $\text{var}(\bar{e}_{iw}) = \sigma_e^2 \sum_{j=1}^{n_i} w_{ij}^2 \equiv \sigma_e^2 \delta_i^2$, and $\bar{x}_{iw} = \sum_{j=1}^{n_i} w_{ij} x_{ij}$.

Remaining sections of the paper are organized as follows. Section 2 presents some standard methods for estimating β, σ_e^2 and σ_v^2, ignoring the sampling weights $\tilde{w}_{ij}$. An iterative weighted estimating equations (IWEE) method using sampling weights is given in section 3. A small simulation study is conducted in section 4 to evaluate the proposed method. We then compare the proposed method with the standard methods using real data in section 5. And finally in section 6, we give some conclusion and remarks.

2. ESTIMATION WITHOUT SAMPLING WEIGHTS

To estimate the regression parameters β, we first assume that the variance components σ_e^2 and σ_v^2 are known in the unit level model (2). Then β is estimated by the generalized least square (GLS) estimator

$$\tilde{\beta}_{GLS} = \left(\sum_{i=1}^{m} x'_i V_i^{-1} x_i \right)^{-1} \left(\sum_{i=1}^{m} x'_i V_i^{-1} y_i \right) \equiv \tilde{\beta}(\sigma_e^2, \sigma_v^2) , \tag{4}$$

where $x_i = (x_{i1},...,x_{in_i})'$, $y_i = (y_{i1},...,y_{in_i})'$, and $V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 1_{n_i} 1'_{n_i}$ with 1_{n_i} and I_{n_i} denoting the unit vector and identity matrix of order n_i, respectively. The estimator $\tilde{\beta}_{GLS}$ depends on the variance components σ_e^2 and σ_v^2. The variance of $\tilde{\beta}_{GLS}$ is $\text{var}(\tilde{\beta}_{GLS}) = (\sum_{i=1}^{m} x'_i V_i^{-1} x_i)^{-1}$.

A simple method of estimating the variance components σ_e^2 and σ_v^2 involves performing two ordinary least squares regression and then using the method of moments to get unbiased estimators of σ_e^2 and σ_v^2 (Fuller and Battese, 1973; Stukel and Rao, 1997). An unbiased estimator of σ_e^2, denoted as $\hat{\sigma}_{eM}^2$, is given by

$$\hat{\sigma}_{eM}^2 = (n - m - p + 1)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^2 , \tag{5}$$

where $\{\hat{\varepsilon}_{ij}\}$ are residuals from the ordinary least squares (OLS) regression of $y_{ij}-\bar{y}_i$ on $\{x_{ij1}-\bar{x}_{i\cdot 1},...,x_{ijp}-\bar{x}_{i\cdot p}\}$ and $(\bar{y}_i,\bar{x}_{i\cdot 1},...,\bar{x}_{i\cdot p})$ are the sample means in the i-th group. An unbiased estimator of σ_v^2 is given by

$$\tilde{\sigma}_{vM}^2 = n_*^{-1}\left[\sum_{i=1}^{m}\sum_{j=1}^{n_i}\hat{u}_{ij}^2-(n-p)\hat{\sigma}_e^2\right], \tag{6}$$

where $n_*^{-1}=n-tr[(X'X)^{-1}\sum_{i=1}^{m}n_i^2\bar{x}_i\bar{x}_i']$ with $X'=(x_1',...,x_m')$, and the $\{\hat{u}_{ij}\}$ are the residuals from the OLS regression of y_{ij} on $\{x_{ij1},...,x_{ijp}\}$. Since $\tilde{\sigma}_{vM}^2$ can take negative values, a truncated estimator of σ_v^2 is obtained as $\hat{\sigma}_{vM}^2=\max(\tilde{\sigma}_{vM}^2,0)$. Note that $\hat{\sigma}_{vM}^2$ is no longer unbiased, but it is consistent as m, the number of groups, increases. The estimators $\hat{\sigma}_{eM}^2$ and $\hat{\sigma}_{vM}^2$ are equivalent to those found by using the well-known method of fitting-of-constants (F-C) due to Henderson (1953). The moment estimators $\hat{\sigma}_{eM}^2$ and $\tilde{\sigma}_{vM}^2$ are, therefore, also referred to as fitting-of-constants estimators. Once σ_e^2 and σ_v^2 are estimated, then using the GLS estimator (4), we get $\hat{\beta}_{GLS}=\tilde{\beta}(\hat{\sigma}_{eM}^2,\hat{\sigma}_{vM}^2)$ as the estimate of β.

Goldstein (1995) proposed an iterative generalized least squares (IGLS) method to estimate the fixed regression parameter β and the variance components σ_e^2 and σ_v^2. The IGLS methods involves two applications of the GLS. The first step is to obtain the GLS estimate $\tilde{\beta}_{GLS}$ of β assuming σ_e^2 and σ_v^2 are known. The second step is to use the GLS estimate $\tilde{\beta}_{GLS}$, given by (4), to form the "raw" residuals $\tilde{y}_{ij}=y_{ij}-x_{ij}'\tilde{\beta}_{GLS}$. Then the estimation of σ_e^2 and σ_v^2 involves an application of GLS on the vector form of the cross-product matrix of the residuals $\tilde{y}_{ij}$, assuming normality. The IGLS method involves iterative updating between the GLS estimate of β and the GLS estimates of σ_e^2 and σ_v^2 until the procedure converges.

3. ESTIMATION WITH SAMPLING WEIGHTS

We now present an iterative weighted estimating equations (IWEE) method to estimate β and variance components σ_e^2 and σ_v^2, using the sampling weights $\tilde{w}_{ij}$.

3.1 Estimating β

First, we obtain the best linear unbiased prediction (BLUP) estimator of the random effect v_i, given the parameters β, σ_e^2 and σ_v^2, from the aggregated model (3) as

$$\tilde{v}_{iw}(\beta,\sigma_e^2,\sigma_v^2)=\gamma_{iw}(\bar{y}_{iw}-\bar{x}_{iw}'\beta)\,. \tag{7}$$

Then, following You and Rao (2002, 2003), we solve a survey-weighted estimating equation for β:

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\tilde{w}_{ij}x_{ij}[y_{ij}-x_{ij}'\beta-\tilde{v}_{iw}(\beta,\sigma_e^2,\sigma_v^2)]=0\,. \tag{8}$$

A model-unbiased estimator of β is obtained from (8) as

$$\tilde{\beta}_w=\left[\sum_{i=1}^{m}\sum_{j=1}^{n_i}\tilde{w}_{ij}x_{ij}(x_{ij}-r_{iw}\bar{x}_{iw})'\right]^{-1}\left[\sum_{i=1}^{m}\sum_{j=1}^{n_i}\tilde{w}_{ij}(x_{ij}-r_{iw}\bar{x}_{iw})y_{ij}\right]\equiv\tilde{\beta}_w(\sigma_e^2,\sigma_v^2)\,. \tag{9}$$

Note that $\tilde{\beta}_w$ is obtained by using the unit level model (2), the aggregated model (3) and the survey weights $\tilde{w}_{ij}$. Based on (9) and the unit level model (2), we note that $\tilde{\beta}_w \mid \beta, \sigma_e^2, \sigma_v^2 \sim N(\beta, \Phi_w)$, where the covariance matrix Φ_w is given in You and Rao (2002, 2003) as

$$\Phi_w = \sigma_e^2 (\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} z_{ij}')^{-1} (\sum_{i=1}^{m}\sum_{j=1}^{n_i} z_{ij} z_{ij}')[(\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} z_{ij}')^{-1}]' + \sigma_v^2 (\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} z_{ij}')^{-1} [\sum_{i=1}^{m} (\sum_{j=1}^{n_i} z_{ij})(\sum_{j=1}^{n_i} z_{ij})'][(\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} z_{ij}')^{-1}]'$$

with $z_{ij} = \tilde{w}_{ij}(x_{ij} - \gamma_{iw}\bar{x}_{iw})$. The covariance matrix Φ_w depends on σ_e^2 and σ_v^2.

3.2 Estimating σ_e^2

To estimate σ_e^2, we first obtain the within group residuals as $e_{ij} - \bar{e}_{iw} = y_{ij} - \bar{y}_{iw} - (x_{ij} - \bar{x}_{iw})'\beta$. Then we take the expectation of the sampling-weighted residual sum of squares

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i} \tilde{w}_{ij}(e_{ij} - \bar{e}_{iw})^2$$

with respect to the model. This leads to the following model-unbiased estimator of σ_e^2:

$$\tilde{\sigma}_{ew}^2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} \tilde{w}_{ij}\left[y_{ij} - \bar{y}_{iw} - (x_{ij} - \bar{x}_{iw})'\beta\right]^2}{\sum_{i=1}^{m}\left[(1-\delta_i^2)\sum_{j=1}^{n_i}\tilde{w}_{ij}\right]} \equiv \tilde{\sigma}_{ew}^2(\beta), \quad (10)$$

where $\delta_i^2 = \sum_{j=1}^{n_i} w_{ij}^2$. Note that the estimator $\tilde{\sigma}_{ew}^2$ depends on β.

3.3 Estimating σ_v^2

To estimate σ_v^2, we first note that the BLUP estimator $\tilde{v}_{iw}$ given by (7) is also the posterior mean of v_i given $\bar{y}_{iw}$ based on the aggregated model (3), i.e., $\tilde{v}_{iw} = E(v_i \mid \bar{y}_{iw})$, assuming the parameters β, σ_e^2 and σ_v^2 are known (You and Rao, 2003). Also note that $E(\tilde{v}_{iw}) = E(E(v_i \mid \bar{y}_{iw})) = E(v_i) = 0$, and

$$E(V(v_i \mid \bar{y}_{iw})) = E(E(\tilde{v}_{iw} - v_i)^2 \mid \bar{y}_{iw}) = E(\tilde{v}_{iw} - v_i)^2,$$

further more,

$$E(\tilde{v}_{iw} - v_i)^2 = \sigma_v^2(\gamma_{iw} - 1)^2 + \sigma_e^2\delta_i^2\gamma_{iw}^2,$$

where $\gamma_{iw} = \sigma_v^2/(\sigma_v^2 + \sigma_e^2\delta_i^2)$. Then, using the decomposition $V(v_i) = V(E(v_i \mid \bar{y}_{iw})) + E(V(v_i \mid \bar{y}_{iw}))$, the variance component σ_v^2 may be expressed as

$$\sigma_v^2 = V(\tilde{v}_{iw}) + E(\tilde{v}_{iw} - v_i)^2 = E(\tilde{v}_{iw}^2) + E(\tilde{v}_{iw} - v_i)^2. \quad (12)$$

Now we average (12) over the groups i and use (11) to get the following estimator of σ_v^2:

$$\tilde{\sigma}_{vw}^2 = \frac{1}{m}\sum_{i=1}^{m}\tilde{v}_{iw}^2 + \frac{\sigma_v^2}{m}\sum_{i=1}^{m}(\gamma_{iw}-1)^2 + \frac{\sigma_e^2}{m}\sum_{i=1}^{m}\delta_i^2\gamma_{iw}^2 = \frac{1}{m}\sum_{i=1}^{m}\tilde{v}_{iw}^2 + \frac{1}{m}\sum_{i=1}^{m}\frac{\sigma_e^2\sigma_v^2\delta_i^2}{\sigma_v^2+\sigma_e^2\delta_i^2} \equiv \tilde{\sigma}_{vw}^2(\tilde{v}_w, \sigma_e^2, \sigma_v^2)\,, \qquad (13)$$

where $\tilde{v}_w = (\tilde{v}_{1w},...,\tilde{v}_{mw})'$. Note that $\tilde{\sigma}_{vw}^2$ depends on σ_e^2 and σ_v^2. Under simple random sampling (SRS) with $w_{ij} = 1/n_i$, the estimator (13) reduces to the estimator of σ_v^2 given by Waclawiw and Liang (1993):

$$\tilde{\sigma}_{vw}^2 = \frac{1}{m}\sum_{i=1}^{m}\tilde{v}_i^2 + \frac{1}{m}\sum_{i=1}^{m}\frac{\sigma_e^2\sigma_v^2}{n_i\sigma_v^2+\sigma_e^2}$$

with $\tilde{v}_i = \gamma_i(\bar{y}_i - \bar{x}_i'\beta)$ and $\gamma_i = \sigma_v^2/(\sigma_v^2 + \sigma_e^2/n_i)$.

3.4 An iterative procedure

We now propose iterative updating steps to obtain survey-weighted estimates for β, σ_e^2 and σ_v^2 simultaneously. Starting with initial values $\hat{\sigma}_e^{2(0)}$ and $\hat{\sigma}_v^{2(0)}$ and for $k = 0,1,2,...$, we update the parameters as follows:

(1) Calculate $\hat{\beta}^{(k+1)} = \tilde{\beta}_w(\hat{\sigma}_e^{2(k)}, \hat{\sigma}_v^{2(k)})$, where $\tilde{\beta}_w$ is given by (9);

(2) Calculate $\hat{\sigma}_e^{2(k+1)} = \tilde{\sigma}_{ew}^2(\hat{\beta}^{(k+1)})$, where $\tilde{\sigma}_{ew}^2$ is given by (10);

(3) Calculate $\hat{v}_{iw}^{(k+1)} = \tilde{v}_{iw}(\hat{\beta}^{(k+1)}, \hat{\sigma}_e^{2(k+1)}, \hat{\sigma}_v^{2(k)})$, where $\tilde{v}_{iw}$ is given by (7);

(4) Calculate $\hat{\sigma}_v^{2(k+1)} = \tilde{\sigma}_{vw}^2(\hat{v}_w^{(k+1)}, \hat{\sigma}_e^{2(k+1)}, \hat{\sigma}_v^{2(k)})$, where $\tilde{\sigma}_{vw}^2$ is given by (13) and $\tilde{v}_w^{(k+1)} = (\tilde{v}_{1w}^{(k+1)},...,\tilde{v}_{mw}^{(k+1)})'$.

Steps (1)-(4) complete one cycle. The iterative cycles are continued until convergence to get the IWEE estimates of β, σ_e^2 and σ_v^2. We can use the fitting-of-constants estimates $\hat{\sigma}_{eM}^2$ and $\hat{\sigma}_{vM}^2$ as the starting values $\hat{\sigma}_e^{2(0)}$ and $\hat{\sigma}_v^{2(0)}$, respectively.

4. SIMULATION STUDY

To evaluate the proposed IWEE method, we conducted a small simulation study. We constructed a synthetic finite populations with $m = 30$ groups (clusters). Each group consisted of $N_i = 500$ population units, and the synthetic population was generated from the unit level full models (1) by taking $x_{ij} = (1, x_{1ij})'$, $\beta_0 = 50$, $\beta_1 = 10$, $\sigma_e^2 = 225$ and $\sigma_v^2 = 100$. The auxiliary variable x_{1ij} was generated from an exponential distribution with mean 200. From the synthetic population, PPS (probability proportional to size) samples within each group were drawn independently. To implement PPS sampling, we used x_{1ij} as the size measures for each y_{ij}. Using these x values, we computed selection probabilities $p_{ij} = x_{1ij}/\sum_j x_{1ij}$ for each unit y_{ij} and then used them to select PPS with replacement samples of equal size, $n_i = n$, within each group, by taking n = 5 and 20, respectively. The basic design weights are given by $\tilde{w}_{ij} = n^{-1}p_{ij}^{-1}$ so that $w_{ij} = p_{ij}^{-1}/\sum_j p_{ij}^{-1}$.

The whole process was repeated R = 500 times, and for each run $r\,(r = 1,...,R)$, we computed the GLS estimate of β and the F-C estimates of σ_e^2 and σ_v^2. Using these estimates as starting values in the IWEE algorithm, we obtained the IWEE estimates of β, σ_e^2 and σ_v^2, respectively. The convergence was very fast and only a few iterations were needed.

Our purpose is to evaluate the performance of the proposed IWEE method and compare it with the GLS estimate of β and the F-C estimates of σ_e^2 and σ_v^2. For this purpose, we computed IWEE estimates under simple random sampling (SRS) and PPS unequal selection probabilities. We calculated the absolute relative bias (ARB) and the relative error (RE) of the estimators. For example, the ARB of an estimator $\hat{\sigma}_e^2$ of σ_e^2 is calculated as $ARB(\hat{\sigma}_e^2) = | E^*(\hat{\sigma}_e^2)/\sigma_e^2 - 1 |$, where E^* denotes the average over the R = 500 runs; that is, $E^*(\hat{\sigma}_e^2) = \sum_{r=1}^{R} \hat{\sigma}_e^2(r)/R$, where $\hat{\sigma}_e^2(r)$ is the estimate based on the r-th simulation run. The RE of $\hat{\sigma}_e^2$ is computed as $RE(\hat{\sigma}_e^2) = \sqrt{\sum_{r=1}^{R} (\hat{\sigma}_e^2(r) - \sigma_e^2)^2 / R} / \sigma_e^2$. Table 1 presents the simulation result of ARB. For fixed effects β_0 and β_1, both GLS and IWEE perform very well in terms of ARB. The ARB is less than 1% for β_0 and less than 0.02% for β_1, indicating unbiasedness. For variance components estimation, the ARB is less than 2% for σ_e^2 and less than 4% for σ_v^2 under both F-C and IWEE approaches, indicating that both F-C and IWEE lead to unbiased estimates of the variance components. Also, estimators of σ_e^2 have smaller bias than the estimators of σ_v^2. Table 2 presents the comparison of RE. For β_0 and β_1, IWEE(SRS) leads to smaller RE than the GLS method. However, the general IWEE leads to larger RE than GLS due to the use of unequal sampling weights. As the sample size, *n*, increases, the RE decreases. For σ_e^2, IWEE(SRS) and F-C have the same RE. For σ_v^2, IWEE(SRS) has smaller RE than F-C. Under unequal probability sampling, IWEE has larger RE than F-C for both σ_e^2 and σ_v^2, as expected. Also, estimators of σ_v^2 have larger RE than the estimators of σ_e^2. We conclude that the proposed IWEE method leads to unbiased estimators of the fixed effects and variance components. Under SRS, IWEE method is very efficient. In general, IWEE has larger RE than the usual methods because it takes account of the unequal sampling weights in the estimation procedure.

Table 1: Comparison of Absolute Relative Bias (ARB)%.

	n = 5			n = 20		
	GLS	IWEE(SRS)	IWEE	GLS	IWEE(SRS)	IWEE
β_0	0.64	0.55	0.24	0.71	0.51	0.32
β_1	0.005	0.007	0.013	0.007	0.007	0.008
	n = 5			n = 20		
	F-C	IWEE(SRS)	IWEE	F-C	IWEE(SRS)	IWEE
σ_e^2	1.5	2.3	1.9	1.3	2.0	1.7
σ_v^2	3.2	3.6	3.8	3.3	3.3	3.5

Table 2: Comparison of Relative Error (RE)%.

	n = 5			n = 20		
	GLS	IWEE(SRS)	IWEE	GLS	IWEE(SRS)	IWEE
β_0	5.1	4.2	8.5	2.7	2.0	6.5
β_1	0.048	0.043	0.082	0.025	0.021	0.042
	n = 5			n = 20		
	F-C	IWEE(SRS)	IWEE	F-C	IWEE(SRS)	IWEE
σ_e^2	12.6	12.8	16.7	6.0	6.0	8.8
σ_v^2	30.7	28.2	38.2	15.0	13.3	28.8

5. APPLICATION TO REAL DATA

In this section, we consider a real data set studied by Battese, Harter and Fuller (1988) in the context of small area estimation. Battese et. al. (1988) considered the estimation of mean hectares of corn and soybeans per segment for 12 counties in north-central Iowa. The total number of sampled segments (total sample size) for the 12 counties is 36, and the sample size n_i $(i = 1,...,12)$ within each county ranged from 1 to 5. The total number of segments N_i (population size) within each county ranged from 402 to 965.

In our calculations, we assumed simple random sampling within areas, as in You and Rao (2002). Thus, the basic survey weight for each sampled unit in area i is $\tilde{w}_{ij} = N_i / n_i$ and $w_{ij} = n_i^{-1}$. The sampling model is

$$y_{ij} = \beta_0 + x_{1ij}\beta_1 + x_{2ij}\beta_2 + v_i + e_{ij}, \quad j = 1,...,n_i, i = 1,...,12,$$

where y_{ij} is the number of hectares of corn (or soybeans) in the j-th segment of the i-th county, x_{1ij} and x_{2ij} are the number of pixels classified as corn and soybeans, respectively, in the j-th segment of the i-th county. Our main interest here is to estimate the fixed effects β and the variance components σ_e^2 and σ_v^2.

To estimate the variance components, we considered five methods for the purpose of comparison, namely, the method of fitting constants (F-C), maximum likelihood (ML), restricted ML (REML), IGLS and the proposed IWEE. Table 3 presents the estimates of σ_e^2 and σ_v^2 for corn and soybeans data using these five methods. From Table 3, the F-C and REML give similar estimates, ML and IGLS give identical estimates, and IWEE gives reasonable estimates. The estimates under IWEE lie between the estimates under F-C (REML) and ML (IGLS), excepting σ_v^2 for soybeans. For IWEE, we used the F-C estimates as the starting values. The iterative algorithm of IWEE converged very quickly. In this example, only a few iterations were needed.

Table 3: Estimation of variance components σ_e^2 and σ_v^2.

	Parameter	F-C	REML	ML	IGLS	IWEE
Corn	σ_e^2	149.6	147.3	137.4	137.3	139.4
	σ_v^2	139.7	139.9	120.9	121.1	130.2
soybeans	σ_e^2	195.2	190.4	177.0	177.0	187.9
	σ_v^2	261.8	247.3	217.6	217.6	207.2

Table 4 presents the weighted estimates of fixed effects β using (9) based on the different estimators of σ_e^2 and σ_v^2 including F-C, REML, IGLS and IWEE. It is clear from Table 4 that the point estimates are very similar. In terms of standard errors, F-C and REML lead to slightly larger standard errors than IGLS and IWEE. The result has shown that under SRS, IWEE estimates are efficient. This is consistent with the simulation results given in section 4.

Table 4: Estimation of fixed effects β.

		Estimates				Standard errors			
	Coefficient	F-C	REML	IGLS	IWEE	F-C	REML	IGLS	IWEE
Corn	β_0	58.491	58.481	58.526	58.492	27.122	26.933	25.939	26.185
	β_1	0.316	0.316	0.316	0.316	0.054	0.054	0.052	0.052
	β_2	-0.160	-0.160	-0.159	-0.160	0.062	0.061	0.059	0.060
Soybeans	β_0	-14.483	-14.388	-14.227	-13.907	31.396	30.974	29.805	30.588
	β_1	0.005	0.005	0.004	0.003	0.062	0.061	0.059	0.060
	β_2	0.514	0.514	0.515	0.515	0.072	0.071	0.068	0.070

6. CONCLUSION

In this paper, we proposed a new method to estimate the fixed effects and variance components in the random intercept multilevel model (nested error regression model) by taking into account of the sampling weights. We compared the proposed IWEE method with some existing methods through a simulation study and application to a real data set. The proposed IWEE method is computationally simple and converges quickly. It makes use of both survey data and sampling weights. Under SRS, IWEE is very efficient. Our method has extended the work of You and Rao (2002) and Waclawiw and Liang (1993) for the fixed effects and variance components estimation. The proposed method can be used in model-based survey data analysis including model-based small area estimation.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988) An error components model for prediction of county crop area using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Fuller, W.A. and Battese, G.E. (1973) Transformation for estimation of linear models with nested error structure. *Journal of the American Statistical Association*, 68, 626-632.

Goldstein, H. (1995) Multilevel Statistical Models. London, Edward Arnold: New York, Wiley.

Henderson, C.R. (1953) Estimation of variance and covariance components. *Biometrics*, 9, 226-252.

Stukel, D. M. and Rao, J. N. K. (1997) Estimation of regression models with nested error structure and unequal error variances under two and three stage cluster sampling. *Statistics & Probability Letters*, 35, 401-407.

Waclawiw, M. A. and Liang, K. Y. (1993) Prediction of random effects in the generalized linear model. *Journal of the American Statistical Association*, 88, 171-178.

You, Y. and Rao, J.N.K. (2002) A pseudo empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30, 431-439.

You, Y. and Rao, J.N.K. (2003) Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*, 111, 197-208.