A Hierarchical Bayes Generalization of the Fay-Herriot Method to Unit Level Nonlinear Mixed Models for Small Area Estimation

A. C. Singh, R. E. Folsom, and A. K. Vaish*
RTI International, Research Triangle Park, NC 27709
***avaish@rti.org**

KEY WORDS: Estimating Functions, Pseudo Score Functions, Survey Weighted Hierarchical Bayes, and MCMC

1. Introduction

It is often the case that the sampling design cannot be ignored when estimating the model parameters governing the finite population. In other words, the model cannot be assumed to hold for the sampled data due to selection bias. This is due to the fact that the design covariates (these are the covariates that drive the sample selection probabilities) cannot all be accounted for in the model for reasons such as the desire to have a parsimonious model, the desire to avoid instability in parameter estimates due to over-fitting, and above all the desire to meet the analyst's goals. Also, some covariates may have to be excluded from the model for practical reasons such as unavailability of finite population totals that are needed in constructing estimates of small area totals.

A natural way to overcome the presence of non-ignorable sample designs in the joint design-model based estimation is to work with transformed or aggregate-level data such as the direct survey estimates for small areas. With aggregate level data, one can take account of the sampling design in specifying the likelihood of model parameters by appealing to the central limit theorem for large samples; here it is assumed that variances of the direct estimates can be treated as known which, in practice, amounts to smoothing them via generalized variance functions. If the preferred model is unit-level which is often the case in practice, there is clearly a loss of efficiency by using aggregate level data. However, this is the price one pays for not knowing the likelihood for unit level sample data from complex designs. Such an approach was used by Fay-Herriot (FH) in their pioneering paper on small area estimation in 1979.

The FH method uses an aggregated level mixed model for small area estimation (SAE). It is commonly used in practice and has several advantages: (i) it is simple and easy to implement, (ii) SAEs produced using the FH method are design consistent, and (iii) they are robust against non-ignorable sample design.

However, the FH approach does have some limitations: (i) there is a loss of information due to area level data aggregation; (ii) unit-level covariate information cannot be exploited, which in turn give rise to wider prediction intervals; (iii) model depends on the level of aggregation e.g. state level model is different from county level model so SAEs are not internally consistent and this problem becomes more acute in the case of nonlinear models; (iv) nonlinear extension suffers from further data loss, e.g. when modeling a low prevalence outcome variable at a lower level of geography such as prevalence of heroin usage in a county, then many counties will be discarded from the model since log (0) is undefined; (v) the large sample assumption required to validate the Gaussian approximation is not reasonable for direct small area estimates; and (vi) smoothing of estimates of variances of direct estimates may not be adequate or possible for areas with few or no observations.

A new approach representing a generalization of the FH method to unit-level nonlinear mixed models is presented here which, like FH, employs data aggregation but through survey-weighted estimating functions (EFs) rather than estimators. Working with EFs helps to alleviate the problems associated with the FH method because EFs can be better approximated by a Gaussian distribution even for the modest sample sizes, and can always be collapsed, if necessary, to improve the Gaussian approximation and the precision of variance estimates. Also, EFs can be based on unit-level covariate information, and can be specified at the lowest level of aggregation to avoid the problem of internal inconsistency. For hierarchical Bayes (HB) SAE, the proposed approach simply replaces the likelihood (computed under the assumption of ignorable design) with the estimating function based Gaussian likelihood which does not require ignorability of the design.

The method is illustrated by means of a simple example of fitting a HB linear mixed model with one covariate to data obtained from a non-ignorable sample design. Both fixed and random parameters are estimated to construct SAEs and Markov Chain Monte Carlo (MCMC) technique is used for HB parameter estimation.

2. Recent Advances in SAE Modeling for Nonlinear Mixed Models

In an innovative attempt to account for the sample design, Prasad and Rao (1999) derived an aggregate level model for direct estimates from the unit level linear model using survey weights, and obtained pseudo-optimal SAEs. It is pseudo in the sense that the design was assumed to be ignorable and only the effect of unequal selection probabilities (i.e., sampling weights) was accounted in estimation of the joint design-model variance. Also, for estimating variance components and the mean square error (MSE) of SAEs, the unequal weighting effect and the sample design was ignored. You and Rao (2001) used a similar framework for developing pseudo HB estimates. There method is also not robust against non-ignorable sample design. However, it uses unit level covariate information and is applicable to nonlinear mixed models.

Folsom, Shah, and Vaish (1999) developed Survey Weighted Hierarchical Bayes (SWHB) SAE methodology for non-linear unit level mixed models. The SAEs obtained by SWHB methodology are design consistent and internally consistent. Moreover, the SWHB methodology allows use of unit level covariates in SAE modeling. However, the SWHB methodology is not robust against non-ignorable sample designs and would lead to biased posterior distributions. We remark that the problem of SWHB-SAE arose in the context of the National Household Survey on Drug Abuse (now known as the National Survey on Drug Use and Health), where it was desired to fit a mixed logistic model. This was a daunting SAE application task with a very large data set and many covariates for which no existing software was applicable, and this task was addressed by Folsom et al. (1999) and Shah et al. (2000).

Our goal is to attempt to take full account of the survey design in unit-level modeling and to develop methods that apply to both linear and nonlinear models. In the next section, we describe the propose SAE methodology.

3. Proposed Methodology

The proposed methodology is based on survey weighted EFs. Use of survey weighted EFs has been implicitly invoked by survey statisticians for a long time in ratio and regression type estimators, see e.g., Fuller (1975), Cassel, Särndal, and Wretman (1976). The pioneering work of Binder (1983) explicitly introduced a general theoretical framework of survey weighted EFs for deriving estimators of super population parameters, and their asymptotic properties under a given sample design. The optimality of survey-weighted EFs under joint design-model randomization was, however, established by Godambe and Thompson (1986) using the optimality framework of Godambe (1960). To illustrate our methodology, we first consider the framework (at the census level) for a linear mixed model with one covariate. Let

$$y_{ij} = \beta_0 + \beta_1\, x_{ij} + \eta_i + \varepsilon_{ij}$$

where

$$\eta_i \sim N(0,\sigma_\eta^2),\ (\beta_0,\beta_1) \sim U(R^2),\ and\ \sigma_\eta^2 \sim IG(v_0/2, \sigma_{\eta_0}^2/2)$$

$$y_{ij}\,\big|(\beta_0,\beta_1,\eta_i,\sigma_\varepsilon^2) \sim N(\beta_0+\beta_1 x_{ij}+\eta_i, \sigma_\varepsilon^2), \quad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2)$$

$$i = 1,\ldots,M\,(strata) \quad and \quad j = 1,\ldots,N_i\,.$$

Note, for simplicity, we have assumed that σ_ε^2 is known. Otherwise, the prior for σ_ε^2 can be assumed to be improper like that of the mean parameters β's. In that case, we can introduce a separate EF, $\phi_{\sigma_\varepsilon^2}$, for σ_ε^2 which treats σ_ε^2 as a mean parameter. It turns out, as expected and as in the case of FH, that σ_ε^2 is not a part of the variance-covariance matrix of ϕ's (defined below) when a suitable design-based estimate is substituted. So we need to add an extra EF if the estimation of σ_ε^2 is also of interest.

With non-ignorable sample design (small areas as strata), define survey weighted EFs for β_1 *and* $\eta_1,\ldots,\eta_M$ using the sample data as

$$\phi_{\eta_i} = \sum_{j=1}^{n_i} (y_{ij} - \beta_0 - \beta_1\, x_{ij} - \eta_i) w_{ij}, \quad i = 1,\ldots,M \quad and$$

$$\phi_{\beta_1} = \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}(y_{ij} - \beta_0 - \beta_1 x_{ij} - \eta_i) w_{ij}$$

where w_{ij} are the design weights or inverse of the first order probabilities. The EF for β_0 is simply the sum of ϕ_{η_i}'s. We propose to use the above set of EFs as the starting point for Bayes or HB estimation, i.e., the likelihood would be defined by the distribution of these EFs. Clearly, EFs use unit-level information and they use it efficiently in view of their optimality properties. It is also known that EFs can be better approximated as a Gaussian distribution even for the modest sample sizes (McCullagh and Nelder, 1989) because by their very nature, they are simple sums of elementary zero functions, although the elementary functions could be complex by themselves. Moreover, EFs can be easily collapsed to improve the Gaussian approximation as well as the precision of variance estimates. Note that the serious problem of internal inconsistency can be avoided by defining the EFs at the lowest level of aggregation. Thus, parameters at the higher levels of aggregation can be obtained from the lowest level parameter estimates which serve as building blocks. It should also be noted that, typically in practice, the joint inclusion probabilities $(\pi_{i(jk)})$ of units j and k in stratum i are not available and therefore, survey weighted EFs can't be constructed if they involve cross-product terms, e.g., if they involve double sums within a stratum i. It is, therefore, desirable to specify the above model so that the error terms ε_i's are i.i.d. which, in turn, gives rise to single sums within the strata for survey weighting.

Now, the vector of EFs serves as the condensed input data which after collapsing, if necessary, gives rise to an approximate Gaussian likelihood, L(y^*|β, η, ·) where y^* denotes the implicit condensing of information in y via EFs. Thus, for the unit-level HB analysis, the original likelihood L(y |·) (which would have been based on the ignorable design assumption) is replaced by the estimating function based Gaussian likelihood (EFGL), L(y^*|·) which does not assume ignorability of the design.

Let $\Phi' = (\phi_{\eta_1},\ldots,\phi_{\eta_M},\phi_{\beta_1}),\quad and \quad V_\Phi = Cov(\Phi)$ where V_Φ is design based variance-covariance matrix of Φ. Further suppose, $\Phi \sim N_{M+1}(0, V_\Phi)$. Due to this assumption, the proposed method is henceforth referred to as Estimating Function Based Gaussian Likelihood (EFGL). The EFG log-likelihood is given by $l(\Phi\,|\,data) = const - \frac{1}{2}(\Phi' V_\Phi^{-1} \Phi)$. It may be noted that there is, in fact, a second component involving σ_ε^2 when the variance covariance matrix of ϕ's is computed under joint design-model randomization. However, it is negligible in comparison to the first term, V_Φ, under the usual assumption of $n_i << N_i$. It should also be emphasized that, in practice, some collapsing of ϕ_{η_i}'s may often be required because the corresponding n_i 's may be small. We may need this collapsing to improve the Gaussian approximation, as well as to improve the precision of the estimate V_Φ. With the specification of EFGL, estimation of parameters $[\eta=(\eta_1,\ldots,\eta_M),\beta=(\beta_0,\beta_1),\sigma_\eta^2]$ can proceed in the HB setup using MCMC steps. The next section gives details of full conditional posterior distributions needed for MCMC.

4. MCMC for the Proposed HB-SAE

For implementing Gibbs sampling, the full conditional posterior distributions are obtained below. Note that,

$$\Phi=\begin{bmatrix} \sum_{j=1}^{n_1} y_{1j}w_{1j}-\beta_0\sum_{j=1}^{n_1} w_{1j}-\beta_1\sum_{j=1}^{n_1} x_{1j}w_{1j}-\eta_1\sum_{j=1}^{n_1} w_{1j} \\ \vdots \\ \sum_{j=1}^{n_M} y_{Mj}w_{Mj}-\beta_0\sum_{j=1}^{n_M} w_{Mj}-\beta_1\sum_{j=1}^{n_M} x_{Mj}w_{Mj}-\eta_M\sum_{j=1}^{n_M} w_{Mj} \\ \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}y_{ij}w_{ij}-\beta_0\sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}w_{ij}-\beta_1\sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}^2w_{ij}-\sum_{i=1}^{M}\eta_i\sum_{j=1}^{n_i} x_{ij}w_{ij} \end{bmatrix}$$

Let

$$Y'_{1\times(M+1)}=(\sum_{j=1}^{n_1} y_{1j}w_{1j},\ldots,\sum_{j=1}^{n_M} y_{Mj}w_{Mj},\sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}y_{ij}w_{ij}),$$

$$W'=(\sum_{j=1}^{n_1} w_{1j},\ldots,\sum_{j=1}^{n_M} w_{Mj}) \quad , \quad X'=(\sum_{j=1}^{n_1} x_{1j}w_{1j},\ldots,\sum_{j=1}^{n_M} x_{Mj}w_{Mj}),$$

$$X_1=\begin{pmatrix} W_{M\times1} & X_{M\times1} \\ \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}w_{ij} & \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}^2w_{ij}\end{pmatrix}, \text{ and } \quad X_2=\begin{pmatrix} diag(W) \\ X' \end{pmatrix}_{(M+1)\times M}.$$

Then we have

$$\Phi=Y-X_1\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}-X_2\begin{pmatrix}\eta_1\\ \vdots \\ \eta_M\end{pmatrix}=Y-X_1\beta-X_2\eta.$$

$$Let\ Y_1=Y-X_2\eta \ \ and \ \ Y_2=Y-X_1\beta.$$

Then full conditional posterior distributions are given below

$$\left[\beta|\eta,data\right]\sim N[(X_1'V_\Phi^{-1}X_1)^{-1}X_1'V_\Phi^{-1}Y_1,(X_1'V_\Phi^{-1}X_1)^{-1}],$$

$$\left[\eta\,|\,\beta,\sigma_\eta^2,data\right]\sim$$

$$N\left[(X_2'V_\Phi^{-1}X_2+\frac{I}{\sigma_\eta^2})^{-1}X_2'V_\Phi^{-1}Y_2,(X_2'V_\Phi^{-1}X_2+\frac{I}{\sigma_\eta^2})^{-1}\right],$$

$$\left[\sigma_\eta^2|\eta\right]\sim IG[\nu_1,\nu_2]\ where$$

$$\nu_1=(\nu_0+M)/2\ and\ \nu_2=(\sigma_{\eta_0}^2+\sum_{i=1}^{M}\eta_i^2)/2.$$

Using the full conditional posterior distributions given above, we can now use Gibbs sampling technique and generate MCMC samples from the joint posterior distribution of the model parameters as described below.

Step 0: Fix $\beta^{(0)}=(\beta_0^{(0)},\beta_1^{(0)})$, $\eta^{(0)}=(\eta_1^{(0)},\ldots\eta_M^{(0)})$, $\sigma_\eta^{2^{(0)}}$, and obtain $V_\Phi^{(0)}$.

Step 1: Generate a

$\beta^{(1)}$ *from* $\left[\beta|\eta,data\right]$

$\eta^{(1)}$ *from* $\left[\eta|\beta,\sigma_\eta^2,data\right]$ *and obtain* $V_\Phi^{(1)}$

$\sigma_\eta^{2^{(1)}}$ *from* $\left[\sigma_\eta^2|\eta\right]$

Step 2: Repeat Step 1 for the required number of MCMC cycles.

5. Simulation Experiment

Consider a universe of $i=1,\cdots,M$ strata (small areas) where $M=100$ and let N_i denote the number of population members in stratum-i. In this simulation experiment, we set $N_i=N_0\,(1+\exp(u_i^*))$ where N_0 is a constant and u_i^* is obtained by truncating $u_i\sim N(0,0.2)$ at $\pm(0.2)^{.5}$. For simplicity, we consider a single covariate super population linear mixed model $y_{ij}=\beta_0+x_{ij}\ \beta_1+\eta_i+\varepsilon_{ij}$ where $\beta_0=0.5$, $\beta_1=1$, $\eta_i\sim N(0,0.2)$, $\varepsilon_{ij}\sim N(0,4)$, and $j=1,\cdots,N_i$. The covariate $x_{ij}=\upsilon_i+\delta_{ij}$ where $\upsilon_i\sim N(0,0.1)$ and $\delta_{ij}\sim N(0,1)$. We generate $K=150$ population level data sets with common x_{ij} and N_i where N_i 's are generated using N_0 =3000. For each of these population data sets, we further stratify the stratum-i population into two substrata Ω_{i+} with $\varepsilon_{ij}>0$ and Ω_{i-} with $\varepsilon_{ij}\le 0$. Let N_{i+}, N_{i-} denote the sizes of these substrata and n_{i+}, n_{i-} denote the sizes of the simple random samples selected without replacement from these strata, respectively. Note that the substratum sizes vary across populations. Let $N=\sum_{i=1}^{100} N_i$ and $n=\sum_{i=1}^{100} n_i$ where $n_i=n_{i-}+n_{i+}$. We generate 150 populations and corresponding 150 samples. In our simulation experiment, $N=628897$, $n_{i-}=60$ and $n_{i+}=20$.

For each sample ($s=1,\cdots,150$), using Gibbs sampling technique, we generate 10,000 MCMC samples for each of the model parameters, namely $\beta_0,\beta_1,\eta_1,\ldots\eta_{\rm M}$, and σ_η^2. These MCMC samples are tested for convergence criterion using CODA (Convergence Output Data Analysis software). First 1000 MCMC samples are deleted for "burn-in" period and from the rest of the 9000 MCMC samples we selected every ninth sample to minimize any auto-correlation among samples, yielding a final MCMC sample of size 1000.

Let $\theta_{sc}=(\beta_{0sc},\beta_{1sc},\eta_{isc},\sigma_{\eta sc}^2)$ denote the parameter values from the c-th MCMC cycle corresponding to the s-th sample. In Table 1, the average posterior mean of θ is defined as $(\sum_{s=1}^{150}\sum_{c=1}^{1000}\theta_{sc}) \div (1000\times150)$ and the average posterior standard deviation of each element of θ_{sc} is defined as the square root of $(\sum_{s=1}^{150}\sum_{c=1}^{1000}(\theta_{sc}-\bar{\theta}_s)^2)\div(1000\times150)$ where $\bar{\theta}_s = (\sum_{c=1}^{1000}\theta_{sc})\div 1000$. Let $\Theta_{isc} = \beta_{0sc} + \bar{X}_i\ \beta_{1sc} + \eta_{isc}$ denotes the small area estimate from the s-th sample for the i-th area using the c-th MCMC cycle where $\bar{X}_i = (\sum_{j=1}^{N_i} x_{ij}) \div N_i$. Also, define $\Theta_{is}^* = \beta_0 + \bar{X}_i\ \beta_1 + \eta_{is}$ where η_{is} is the true value of η_i for the s-th population. Let L_{is} and U_{is} denote 2.5 and 97.5 percentiles of the posterior distribution of Θ_{is} obtained from 1000 MCMC samples of Θ_{isc}. Define $\psi_{is} = \begin{cases} 1 & \text{if} \quad \Theta_{is}^* \in [L_{is}, U_{is}] \\ 0 & \textit{otherwise.} \end{cases}$

The coverage probability distribution characteristics given in Tables 2 are obtained from the distribution of 100 area-i specific values of $(\sum_{s=1}^{150}\psi_{is})\div 150$.

6. Simulation Results

Tables 1 and 2 given at the end of Section 10 summarize the simulation results. In Table 1, average posterior means and standard deviations for the EFGL method are compared with solutions from the HB version of the FH model and with the unweighted solution. The average posterior means from the FH and EFGL solutions are very close to each other. The average posterior mean for β_0 from the unweighted solution is negatively biased (-0.2953) due to the fact that we over sample the $\Omega_{i\_}$ substrata. The average posterior standard deviations from the EFGL and FH solutions for β_0 and σ_η^2 are also very close to each other whereas the average posterior standard deviation for β_1 from the EFGL model is more than 6 times smaller than the solution from the FH model. The average posterior standard deviations from the EFGL model are very close to the average posterior standard deviation from the unweighted solution.

In Table 2, coverage probabilities for the EFGL solution are compared with the FH solution coverage probabilities. The coverage probabilities for both solutions are very close. However prediction intervals for the EFGL solution are 20% narrower than the FH solution, which is expected, since the EFGL solution utilizes unit level covariate information whereas the FH solution uses aggregated level covariate information.

7. Unit Level Nonlinear Mixed Models

The EFGL method introduced in Section 3 for finding HB-SAEs in the context of mixed linear models can be easily applied to mixed nonlinear models, the only difference being that full conditional posteriors of the β's and η's no longer have analytic solutions. Therefore, as expected, the method gets more computationally intensive. To illustrate the ideas, we consider the following mixed logistic model:

$y_{ij} = \mu_{ij} + \varepsilon_{ij}$ where y_{ij} ~Bernoulli(μ_{ij}),

$\text{logit}(\mu_{ij}) = x_{ij}'\beta + \eta_i$,

$\eta_i \sim_{i.i.d.} N(0, \sigma_\eta^2)$, $\beta \sim U(R^p)$, and $\sigma_\eta^2 \sim IG(\nu_0/2, \sigma_{\eta_0}^2/2)$.

The EFs in this case remain similar to the linear case except that the elementary zero functions (or the residuals) $y_{ij} - \mu_{ij}$, are complex due to the nonlinear form of μ_{ij}'s. Observe that the EFs continue to be simple linear functions of elementary zero functions, and hence they behave well in terms of Gaussian approximation. The EFs for the logistic case (when small areas are strata) are given by

$$\varphi_{\eta_i} = \sum_{j=1}^{n_i}(y_{ij} - \mu_{ij})w_{ij} \overset{\sim}{\sim} N(0, V_{\eta_i})$$

$$\varphi_\beta = \sum_{i=1}^{M}\sum_{j=1}^{n_i} x_{ij}(y_{ij} - \mu_{ij})w_{ij} \overset{\sim}{\sim} N(0, V_\beta).$$

We can orthogonalize φ_β with respect to φ_{η_i}'s. Also, when β includes an intercept, φ_β element corresponding to the intercept should be dropped because of its linear dependence on the φ_{η_i}'s. Now, the likelihood, $L(y^* \mid \beta, \eta)$ can be approximately specified as before, but the MCMC steps are modified as follows:

Step 1. $[\beta \mid y^*, \eta]$

Since the full sample is typically very large, the full conditional posterior can be well approximated by

$$[\beta \mid y^*, \eta] \overset{\sim}{\sim} N(\hat{\beta}_{\text{mode}}, \Sigma_{\psi_\beta}^{-1}) \quad (1)$$

where $\hat{\beta}_{\text{mode}}$ solves $\psi_\beta = 0$, and $\psi_\beta = (\partial/\partial\beta)\log L(y^* \mid \beta, \eta)$, $\Sigma_{\psi_\beta} = -E[(\partial\psi_\beta/\partial\beta)]$.

Note that unlike the linear case, $\hat{\beta}_{\text{mode}}$ does not have an analytic form. Also note that instead of the approximate posterior distribution (1), one can get realizations from an exact posterior distribution by using the Metropolis-Hastings (MH) step within MCMC in which (1) can be used as a proposal distribution.

Step 2. $[\eta_i \mid \eta_{i'}, \beta, y^*, \sigma_\eta^2], i \neq i'$ and $i = 1, ..., M$.

As mentioned earlier, this again does not have an analytic solution. We could use MH with mle/prior for the proposed distribution. In other words, solve $\psi_{\eta_i} - \sigma_\eta^{-2}\eta_i = 0$ to get $\hat{\eta}_{i,\text{mle-adj}}$, where $\psi_{\eta_i} = (\partial/\partial\eta_i)\log L(y^* \mid \beta, \eta)$, and use $N\left(\hat{\eta}_{i,\text{mle-adj}}, \left(\sigma_{\psi_{\eta_i}}^2 + \sigma_\eta^{-2}\right)^{-1}\right)$ as the proposal distribution where $\sigma_{\psi_{\eta_i}}^2 = -E[\partial\psi_{\eta_i}/\partial\eta_i]$.

Step 3. $\left[\sigma_{\eta}^{2} \mid \eta\right]$

We get the same result as in the linear case. Note that Step 4 for $\left[\sigma_{\varepsilon}^{2} \mid \cdot\right]$ is not needed because σ_{ε}^{2} is a known function of μ_{ij} in the logistic case.

8. Conclusion

The EFGL method was developed to exploit unit-level covariate information, to take full account of the survey design, and to have a valid (approximate) likelihood for HB-SAE methodology applied to generalized linear/nonlinear mixed models. It generalizes the aggregate-level model of Fay-Herriot (1979), and the pseudo-likelihood approach of Folsom, Shah and Vaish (1999). There are essentially two main ideas in EFGL, namely, the data aggregation via EFs and EF-collapsing. The main reason for EF-collapsing is to improve Gaussian approximation, and the secondary purpose is to improve the variance estimate's precision. In practice, it may be preferable to use separate modeling to make variance estimates more stable. However, even if variance estimates are not precise, it is often of interest, in practice, to see how much variance reduction can be realized through SAE modeling.

The other idea of data aggregation in EFGL is somewhat similar to that of the FH method except it tries to take advantage of the unit-level information as much as possible. Hence, the resulting estimates from the EFGL method are expected to be more efficient than those from the FH method. In particular, for the case of simple linear mixed models with known variance components, it can be easily shown analytically that precision of the estimates of fixed effects (β) can be improved substantially in the case of unit level models if the covariates (x_{ij}) have sufficient variability within areas. There is also some gain in efficiency of random effect (η_i's) estimates. However, if η_i's are also defined as coefficients of suitable covariates (z_{ij}'s) as in the case of random regression coefficients, then high efficiency gains in estimating random effects can also be realized if there is sufficient variability in z_{ij}'s within areas.

The ideas underlying the proposed method of EFGL are quite general, and the method is applicable to general nonlinear mixed models for survey data. However, it does have some limitations which the user should keep in mind. Some loss of efficiency is inevitable due to data aggregation and EF-collapsing. This is the price we pay for not having enough information about the likelihood of the sampled data, and by not being able to ignore the sample design. EF-collapsing may be needed for the Gaussian approximation. In practice, it is better to avoid it if possible as it doesn't distinguish much between the areas involved in collapsing. At the design stage, one can take measures to ensure a sufficient number of observations in each small area in order to avoid EF-collapsing.

9. Acknowledgements

This research was supported in part by grants from the U.S. Census Bureau (through NSF) and from CDC to the University of North Carolina at Chapel Hill. The earlier versions of this paper were presented at the International Conference on Small Area Estimation and Related Topics, Potomac, MD, April 11-14, 2001 and at the FCSM Conference, Arlington, VA, Nov 14-16, 2001. The authors would like to thank Jon Rao, Bob Fay, William Bell, Don Malec, Malay Ghosh, and Nat Schenker for their helpful and encouraging comments.

10. References

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

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 615-620.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Folsom, R. E., Shah, B. V., and Vaish, A. K. (1999). Substance abuse in states: A methodological report on model based estimates from the 1994-96 NHSDAs. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 371-375.

Fuller, W. A. (1975). Regression analysis for sample surveys. *Sankhya*, Ser C, 37, 117-132.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31, 1208-1212.

Godambe, V. P. and Thompson, M. E, (1986). Parameters of super population and survey population, their relationship and estimation. *International Statistical Review*, 54, 127-38.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed.

Prasad, N. G. N. and Rao, J. N. K. (1999). On robust small area estimates using a simple random effects model. *Survey Methodology*, 25, 67-72.

Shah, B. V., Barnwell, B., Folsom, R.E., and Vaish, A. K. (2000). Design consistent small area estimates using Gibbs algorithm for logistic models. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 106-111.

You, Y. and Rao, J. N. K. (2001). Pseudo hierarchical Bayes small area estimation combining unit level models and survey weights. *Journal of Statistical Planning and Inference*.

Table 1: Average Posterior Mean and Standard Deviation for Model Parameters

Parameter	True Value	Average Posterior Mean			Average Posterior Standard Deviation		
		Fay-Herriot	EFGL	Unweighted	Fay-Herriot	EFGL	Unweighted
β_0	0.5	0.5024	0.5007	-0.2953	0.0488	0.0469	0.0460
β_1	1.0	1.0012	1.0011	0.9999	0.1696	0.0258	0.0206
σ_η^2	0.2	0.1971	0.1976	0.1703	0.0341	0.0317	0.0304

Table 2: Coverage Probability and Ratio of Predication Interval Widths

Distributional Characteristics	Coverage Probability		Ratio of CI Width (EFGL/Fay-Herriot)
	Fay-Herriot	EFGL	
95%	0.973	0.973	0.808
75%	0.953	0.953	0.802
50%	0.940	0.933	0.798
Mean	0.941	0.934	0.797
25%	0.927	0.913	0.792
5%	0.910	0.900	0.785