Small Area Estimation Using Multi-Level Models

D Holt and F Moura, University of Southampton
D Holt, Department of Social Statistics, Southampton, S09 5NH, UK

Key words: Small area, mixed models, multi-level, BLUP

1. Introduction

Battese et al (1981, 1988) proposed and applied a nested error regression model to provide small area estimates. The model took the form:

$$Y_i = X_i \beta + v_{oi} 1_i + \varepsilon_i \qquad (1.1)$$

where Y_i, ε_i and $1_i=(1,...,1)'$ are vectors of length n_i for the sampled units in the ith small area, $i = 1,...,A$; and X_i is the $n_i \times (p+1)$ matrix of explanatory variables. The vector β is a set of (p+1) fixed regression parameters and v_{oi} is a scalar random effect for each small area where $E(v_{oi})=0$; $V(v_{oi})=\sigma_0^2$; cov $(v_{oi}, v_{oi'})=0$ $i\neq i'$. The ε_i are assumed independent ($E(\varepsilon_i) = 0$; $V(\varepsilon_i)=\sigma_\varepsilon^2 I_i$) and v_{oi} and ε_i are assumed independent. For the whole population (1.1) applies with n_i replaced by N_i the small area population sizes. When the components of variance are assumed to be known, the Best Linear Unbiased Predictor (BLUP) estimator of the ith small area mean is obtained from (1.1):

$$\hat{\bar{Y}}_{i(RI)} = \bar{X}_i'\hat{\beta} + \hat{v}_{oi} \qquad (1.2)$$

Where $\bar{X}_i$ is the (p+1) vector of population means for the auxiliary variables including a constant term for the ith small area; $\hat{\beta}$ is the BLUP estimator for β and $\hat{v}_{oi}$ is a predictor for v_{oi}. The label RI is taken to imply a random intercept model.

Battese et al (1988) estimated β, σ_0 and σ_ε from real data and Prasad and Rao (1990) used these in a simulation study to evaluate the efficiency of their estimators for each small area, together with the accuracy of their approximation to the MSE and the relative bias of the estimator of the MSE. The numerical results show that when the model in (1.1) holds, the two stage estimator (as they called it) is considerably more efficient than the regression synthetic estimator and the approximately unbiased regression estimator. The relative bias of the MSE estimator is small (< 7%) under normally distributed random effects.

2. A more general model

2.1 Introduction

In model (1.1) the differences between small areas are represented by v_{oi}, a random intercept term. In practice the regression coefficients, β, can vary across the small areas too. Thus a more general approach should allow differences between slopes for each small area as a set of random terms. A second generalization is to extend the framework to a multi-level model by the introduction of explanatory variables at the small area level so helping to explain differences between small areas.

We consider the following generalization of model (1.1) for predicting the small area means:

$$Y_i = X_i\beta_i + \varepsilon_i$$
$$\beta_i = Z_i\gamma + v_i \qquad (2.1)$$

where:

Z_i is the $(p + 1) \times q$ design matrix of small area level variables, γ is the vector of length q of fixed coefficients, and $v_i = (v_{i0},\ldots, v_{ip})'$ is a vector of length p+1 of random effects for the ith small area. We assume the following about the distribution of the random vectors:

(a) The v_i are independent between small areas and have a joint distribution within each small area with $E(v_i)=0$ and $V(v_i)=\Omega$

$$\Omega = \begin{bmatrix} \sigma_0^2, \sigma_{01}, \ldots, \sigma_{0p} \\ \sigma_{10}, \sigma_1^2, \ldots, \sigma_{1p} \\ \vdots \quad \vdots \quad \vdots \\ \sigma_{p0}, \sigma_{p1}, \ldots, \sigma_p^2 \end{bmatrix} \quad (2.2)$$

(b) The ε_i's and v_i's are independent and $V(\varepsilon_i) = \sigma_\varepsilon^2 I_i$.

(c) Although this is unnecessary for point prediction multivariate normality is assumed for v_i and ε_i

A special case of (2.2) is when Ω is diagonal so that the small area regression coefficients are random but uncorrelated between covariates. We refer to this as the Diagonal model and to the models with correlated random effects as the General model. Other intermediate models exist with some covariance terms constrained to zero.

2.2 The estimator of the small area mean

Let $\theta = ([\text{vech}(\Omega)]', \sigma_\varepsilon^2)' = (\theta_1, \ldots, \theta_s)'$ be the random parameters in the model (2.1).

From Henderson (1975), the best linear unbiased predictor for $u_i = \bar{X}_i' \beta_i$ when θ is known is given by:

$$\hat{u}_i = \bar{X}_i'\hat{\beta}_i = \bar{X}_i' Z_i\hat{\gamma} + \bar{X}_i' \hat{v}_i \quad (2.3)$$

where $\hat{\gamma}$ is the generalized least squares estimator of γ ; that is:

$$\hat{\gamma} = \left(\sum_{i=1}^{t} Z_i'X_i'V_i^{-1}X_iZ_i\right)^{-1}\left(\sum_{i=1}^{t} Z_i'X_i'V_i^{-1}Y_i\right) \quad (2.4)$$

$$V_i = \sigma_\varepsilon^2 I_i + X_i\Omega X_i' \quad (2.5)$$

The estimator of the predicted values of the residual v_i is:

$$\hat{v}_i = \Omega X_i' V_i^{-1} (Y_i - X_iZ_i\hat{\gamma}) \quad (2.6)$$

When θ is unknown, there are iterative methods for estimating θ and γ. See for example, Longford, N.T. (1987) and Goldstein (1986 and 1989). We confine ourselves to the Restricted Maximum Likelihood Estimator.

Let $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_s)'$ and $\hat{\gamma}*$ be the restricted maximum likelihood estimator of θ and γ respectively. Replacing θ with $\hat{\theta}$ and $\hat{\gamma}$ with $\hat{\gamma}*$ in (2.2) and (2.6), we obtain the following predictor of u_i:

$$\hat{u}_i^* = \bar{X}_i'Z_i\hat{\gamma}* + \bar{X}_i'\hat{v}_i^* \quad (2.7)$$

where $\hat{v}_i^* = \hat{\Omega} X_i' \hat{V}_i^{-1} (Y_i - X_iZ_i\hat{\gamma}*)$ (2.8)

$$\hat{V}_i = \hat{\sigma}_\varepsilon^2 I_i + X_i\hat{\Omega} \hat{X}_i' \quad (2.9)$$

2.3 Approximation to the Mean Square Error (MSE)

Kackar and Harville (1984) show that the mean square error of $\hat{u}_i^*$ can be approximated by:

$$\text{MSE}(\hat{u}_i^*) = E[\hat{u}_i - u_i]^2 + E[\hat{u}_i^* - \hat{u}_i]^2 \quad (2.10)$$

It may be shown that an approximation to the MSE is:

$$MSE(\hat{u}_i^*) = \overline{X}_i'(G_i')^{-1}\Omega\overline{X}_i + \sigma_\varepsilon^2\, \overline{X}_i'\, (G_i')^{-1} Z_i \left[\sum_{i=1}^{t} Z_i'\, G_i^{-1} X_i'\, X_i Z_i\right]^{-1} Z_i' G_i^{-1} \overline{X}_i + \text{trace}\,(AB) \quad (2.11)$$

where G_i is the (p+1) × (p+1) matrix:

$$G_i = I_{p+1} + \sigma_\varepsilon^{-2}\, X_i'\, X_i\, \Omega \quad (2.12)$$

The elements of A are given by:

$$A_{\ell,s} = -\overline{X}_i' \left[(G_i')^{-1} \frac{\partial\Omega}{\partial\theta_\ell} R_i \Omega\right] \overline{X}_i$$

$$\ell = 1, \ldots, s\text{-}1 \quad (2.13)$$

$$A_{\ell,\ell'} = \overline{X}_i' \left[(G_i')^{-1} \frac{\partial\Omega}{\partial\theta_\ell} C_i \left(\frac{\partial\Omega}{\partial\theta_{\ell'}}\right) G_i^{-1} \right] \overline{X}_i \quad (2.14)$$

$\ell,\ell' = 1, \ldots, s\text{-}1$ and $\ell \leq \ell'$.

$$A_{s,s} = \overline{X}_i'\, \Omega\, S_i \Omega \overline{X}_i \quad (2.15)$$

where $C_i = \sigma_\varepsilon^{-2}\, G_i^{-1}\, X_i'\, X_i$;

$R_i = \sigma_\varepsilon^{-4}\, G_i^{-2}\, X_i'\, X_i$

and $S_i = \sigma_\varepsilon^{-4}\, G_i^{-2}\, C_i$.

The elements of B^{-1} are given by:-

and $$B^{-1}_{\ell,\ell'} = \frac{1}{2} \sum_{i=1}^{t} \text{tr}\left(\frac{\partial\Omega}{\partial\theta_\ell} C_i \frac{\partial\Omega}{\partial\theta_{\ell'}} C_i\right)$$

$$\ell, \ell' = 1, \ldots s\text{-}1; \ell \leq \ell' \quad (2.16)$$

$$B^{-1}_{\ell,s} = \frac{1}{2} \sum_{i=1}^{t} \text{tr}\left(\frac{\partial\Omega}{\partial\theta_\ell}\right) C_i$$

$$\ell = 1, \ldots, s\text{-}1 \quad (2.17)$$

$$B^{-1}_{s,s} = \frac{1}{2} \sum_{i=1}^{t} \text{tr}\, V_i^{-2} . \quad (2.18)$$

Each of the three terms in (2.11) can be associated with a particular source of error. The first term is the prediction variance for the situation in which all parameters are known. The second term is the increase of variance due to estimating the fixed parameters. The last term of (2.11) comes from estimating the random parameters.

Since prediction is only required for unobserved units in each small area, we obtain estimators of the small area mean and its mean square error as follows:

$$\hat{\overline{Y}}_{i(G)} = f_i\overline{y}_i + (\overline{X}_i - f_i\overline{x}_i)'(z_i\hat{\gamma}^* + \hat{v}_i^*) \quad (2.19)$$

$$MSE(\hat{\overline{Y}}_{i(G)}) = (1 - f_i)^2 \left[M\hat{S}E^*(\hat{\overline{Y}}_{i(G)}) + N_i^{-1}\, (1-f_i)^{-1}\sigma_e^2\right] \quad (2.20)$$

where $\overline{x}_i$ is the (p+1) vector of sample means and $M\hat{S}E^*\,(\overline{Y}_{i(G)})$ is equation (2.11) with $\overline{X}_i$ replaced by the vector of non-sampled means $\overline{X}_i^c = (\overline{X}_i - f_i\overline{x}_i)/(1-f_i)$.

3. A Numerical Framework

Holt and Moura (1992) used data from a sample of 951 retails stores in Southern Brazil to estimate the parameters of the model given in section 2.1. The data were classified into 73 small areas. They used a single auxiliary variable (X) which was highly correlated with the variable of interest and no small area level (Z) variables. The parameter estimates were used to generate simulated samples, under Normality, using the same fixed covariates (X) values for each sample. The parameter values were $\gamma_0 = 5.515$; $\gamma_1 = 1.046$; $\sigma_\varepsilon^2 = 0.0433$; $\sigma_0^2 = 0.0100$; $\sigma_1^2 = 0.0350$ and

σ_{01} = -0.0085. These results have been extended to include data generated under the diagonal model (assuming σ_{01} = 0) and the random intercept model (σ_1^2 = 0; σ_{01} = 0). For a single covariate (X) the random terms are σ_ε^2, σ_0^2, σ_1^2 and σ_{01} and the estimators compared were:

i) $\hat{\bar{Y}}_{i(RI)}$ the Random Intercept estimator which is the BLUP under the model (2.1) with σ_1^2 and σ_{01} taken as zero

ii) $\hat{\bar{Y}}_{i,(D)}$ the Diagonal Covariance estimator which is the BLUP under the model (2.1) with σ_{01} taken as zero

iii) $\hat{\bar{Y}}_{i(G)}$ the General estimator which is the BLUP under model (2.1).

For each estimator for each small area the MSE is obtained from the simulations by comparing the small area estimator with the corresponding predicted small area mean $\tilde{\bar{Y}}_i$ using the simulated random effects $\tilde{v}_{oi}$ and $\tilde{v}_{1i}$. For example for the Random Intercept estimator and the ith small area.

$$MSE_i(RI) = \frac{1}{R} \sum_r \left(\hat{\bar{Y}}_{i,r(RI)} - \tilde{\bar{Y}}_{i,r} \right)^2$$

where r = 1,...., R indexes the simulations. In this investigation R=10,000 simulations were generated.

Table 1 contains a summary of the ratio of the average mean square error of each estimator compared to the Random Intercept estimator. In addition, rb, the ratio of the relative absolute bias is presented. Specifically the relative absolute bias for a particular small area, i, for the Random Intercept estimator for example is defined as:

$$b_i(RI) = \frac{1}{R} \sum_{r=1} \frac{|\hat{\bar{Y}}_{i,r(RI)} - \bar{Y}_i|}{\bar{Y}_i}$$

and the average for all small areas is:

$$b(RI) = \frac{1}{A} \sum_{i=1}^{A} b_i(RI)$$

Hence the ratio of these measures for the General estimator and the Random Intercept estimator is defined as

$$rb(Gen) = b(Gen)/b(RI) .$$

We note that for this situation the random components are important both compared to the fixed parameters (σ_0/γ_0 = 0.02; σ_1/γ_1 = 0.18) and compared to the unit level variance ($\sigma_o^2/\sigma_\varepsilon^2$ = 0.23).

As one might expect when the data are generated under the General model the General estimator is best although the Diagonal estimator is almost as good. Both estimators are about 10% more efficient than the Random Intercept estimator. Even when the data is generated under simpler models (ie the Random Intercept or Diagonal models) the General Estimators and Diagonal still perform well and are robust. The ratios of the relative absolute bias show the same pattern.

Table 1: Properties of Diagonal and General Estimators compared to Random Intercept estimator

Estimator	Data Generation Model		
	General	Diagonal	Random Intercept
Diagonal	90 (95)	91 (95)	100 (100)
General	87 (93)	91 (95)	100 (100)

N.B. First entry is ratio(%) of Average Mean Square Error for estimator compared to Random Intercept estimator.

Second entry is ratio(%) of relative absolute bias compared to Random Intercept estimator.

There are several comments and criticisms that might be made of this numerical investigation:

i) The simulation is carried out using artificial data rather than 'true' Y values from real data even though the model parameters are determined by analysing real data.

ii) The data were generated under the assumed model framework and using Normally distributed errors.

iii) No area level variables (Z) used.

iv) The covariate values X are held constant for every simulation so that, as in the model framework, all the properties of the estimators are conditional on the fixed covariate values. Survey practitioners would prefer to see properties of estimators with respect to a repeated sampling framework where the X variables changed for each simulation according to a random selection of cases.

4. Numerical Investigations

In order to meet these challenges a different data set was obtained from a test census in the urban part of a county in Brazil which was divided into 140 enumeration districts. The data comprises 38740 household records and the variable of interest is the Head of Household's income (Y). After some preliminary modelling on the entire data set the unit level covariates (X) are taken to be the educational attainment of the Head of Household (ordinal scale of 0-5) and the number of rooms in the household (1-11+). Initially no area level (Z) variables were used.

It must be stressed that this data set is far from ideal insofar as the underlying model assumptions are concerned. One unit level covariate is ordinal although a linear regression model for the two covariates appears to fit the data set approximately. However, the explanatory power of the model is poor (i.e. $R^2 = .37$) and, as one might expect, the distribution of the residuals from the fitted model is highly skewed. In practice one would want to explore the data much more. The investigation presented here should be viewed as a test of the robustness of the predictive power of the multi-level model for making small area estimates in adverse circumstances.

Thus the assumed model is:-

$$Y_{ij} = \beta_{0i} + \beta_{1i}(x_{1ij} - \bar{x}_1) + \beta_{2i}(x_{2ij} - \bar{x}_2) + \varepsilon_{ij}$$

$$i=1 \ldots A;\ j=1 \ldots N_i$$

$$\beta_{01} = \gamma_0 + \nu_{0i}$$

$$\beta_{1i} = \gamma_1 + \nu_{1i}$$

$$\beta_{2i} = \gamma_2 + \nu_{2i}$$

where x_1 = number of rooms

and x_2 = educational attainment of Head of Household.

The records comprise a complete population from each of the 140 small areas so that the true population mean $\bar{Y}_i$ for each small area is known. A small number of records with extreme incomes was excluded from the population. Each simulation consists of a 10% simple random sample from each small area so that the sample size in each area is fixed but the choice of records and the corresponding covariate values is random. For each sample data set the random intercept, diagonal and general models were fitted. For comparative purposes the ordinary regression estimator (with σ^2, σ_1^2 and σ_{01} all zero) was also calculated. No diagnostic checks or tests were made for each separate simulation to see whether the sample appeared to be consistent with the assumed model. Each result in the tables is based upon 500 simulations.

The parameter values for the model are given in Table 2. It will be noted that compared to the parameter values that led to Table 1, the components of variance in Ω are much smaller than the unit level variance σ^2_ε. ($\sigma^2_0/\sigma^2_\varepsilon$ = 0.03) but are large compared to the fixed effects (σ_0/γ_0 = 0.14; σ_1/γ_1 = 0.4; σ_2/γ_2 = 0.4).

Table 2: Population Parameters (No Small Area Level Covariate)

Fixed Parameters	**Est**	se
constant (β_0)	8.46	0.11
Rooms (β_1)	1.22	0.05
Education (β_2)	2.60	0.09

Random Parameters (standard errors in parenthesis)

$$\Omega = \begin{pmatrix} \underset{(0.19)}{1.385} & \underset{(0.07)}{0.354} & \underset{(0.12)}{0.492} \\ & \underset{(0.04)}{0.234} & \underset{(0.05)}{0.333} \\ & & \underset{(0.12)}{0.926} \end{pmatrix}$$

$$\sigma^2_\varepsilon = 47.74 \ (0.35)$$

In order to investigate the impact on small area estimates of various features of the data a series of populations was created from which the simulations were carried out. These were:

P4 The actual Y and corresponding X values (the real data)

P3 The actual X values, but with Y generated artificially under the general model with Normally distributed errors. The 'true' population small area means $\bar{Y}_i$ were obtained from the 'population' created.

P2 The general model was fitted to the whole population and predicted values $\hat{Y}_{ij}$ obtained for each case. The observed Y_{ij} values were then replaced by

$$\tilde{Y}_{ij} = \hat{Y}_{ij} + \frac{(Y_{ij} - \hat{Y}_{ij})}{2}$$

This provides a data set from the 'true' data with the same X values and the same skewed distribution of residuals but reduces the unit level variance to $\sigma^2_\varepsilon/4$ and so makes the variances and covariances of the random terms relatively more important.

P1 As for P3 but with the artificial data generated under a model with the same parameter values except that the unit level variance is replaced by $\sigma^2_\varepsilon/4$.

The rationale behind the choices is that P1 yields a Normally distributed population with the area level random terms of relatively high importance although still not as large as in Table 1. P2 provides a population with a set of fixed and random parameter values corresponding to P1 but with the skewed distribution of residuals representative of the true data. P3 provides an artificial Normally distributed population with the relatively small but realistic components of variance and covariance. P4 is, of course, the original population with small random effects and a skewed set of residuals.

The important difference between Table 1 and the other results will be that the simulations for P1 to P4 are under a repeated sampling framework.

Table 3 contains the ratios of average Mean Square Error and Relative Absolute Bias of the Diagonal and General estimators compared to the random Intercept estimator for the four populations. For both of the artificial (Normally distributed residuals) Populations, P1 and P3, it is clear that the Diagonal and General estimators provide a clear gain in efficiency and reduction in bias compared to the Random Intercept estimator. For population P2 only the Diagonal estimator shows an improvement and for the real data (population P4) the Random Intercept estimator is best. The overall conclusion is that the Diagonal estimator is robust and can provide smaller average relative bias and MSE. For data generated under the model with Normal residuals the general estimator is preferred.

Table 3: Properties of Various estimators compared to the Random Intercept estimator

Estimator	Population			
	P1	P2	P3	P4
Diagonal	81	91	88	103
	(92)	(93)	(95)	(99)
General	73	99	79	105
	(88)	(97)	(90)	(106)
Regression	430	337	188	157
	(196)	(181)	(131)	(127)

Note: First entry is ratio of MSE of each estimator to MSE of Random Intercept estimator (%).

Second entry is ratio of Relative Absolute Bias (b) for each estimator compared to Random Intercept Estimator (%).

The Regression Estimator is included for comparative purposes and demonstrates the loss of efficiency if no components of variance are included in the estimation process at all.

Figure 1 shows the plot of the MSE for each small area for the Random Intercept and Diagonal Estimators for population P2 where the Relative Efficiency for the Diagonal Estimator is 91%. The diagonal line on the plot represents equal MSE for the two estimators. It will be seen that there are small areas with small or large MSE for which one or the other estimator is preferred. Thus whichever estimator is chosen there will be small areas which consequently achieve smaller mean square error and others that suffer a loss of precision.

The essential feature of the Random Intercept estimator is that all regression slopes are assumed to be constant and the only area specific random effect is the random intercept. When the small area specific regression slope is different from the average for all small areas then one might conjecture that the General or Diagonal estimator would be more efficient. However even if the regression slope is different in a particular small area, this will have no effect on the small area prediction if the sample covariate mean $\bar{x}_i$ and the population mean $\bar{X}_i$ are approximately equal. Hence gains in efficiency for the General Estimator for example, compared to the Random Intercept estimator are likely to occur for small areas where

i) the small area regression slope β_i is different from the overall average β (ie v_i is large).

ii) the within area variance for the auxiliary variables is large

and

iii) the sample size is small

To illustrate this point. Figure 2 shows the plot of relative efficiency:-

$$e_i = \left(\frac{MSE_i(RI)}{MSE_i(G)} - 1 \right) \times 100\%$$

against a measure of the small area difference in regression slopes which was derived from the population data:

$$d_i = n_i^{-1}(v_{1i}, v_{2i})\ S_{xx,i}(v_{1i}, v_{2i})'$$

where $S_{xx,i}$ is the variance-covariance matrix for the auxiliary variables in the ith small area. In accordance with the conjecture, it will be seen that the large efficiency gains for the General Estimator generally correspond to large values of d_i.

5. Introducing An Area Level Covariate

The model in section 4 was modified by introducing an area level variable (Z_i: the number of cars per household in each small area). Analyses using the complete population data indicated that this was a useful explanatory variable for the two regression coefficients for the individual level variables 'rooms' and 'education' but not for the constant term. That is

$$Y_{ij} = \beta_{oi} + \beta_{1i}(x_{1ij}-\bar{x}) + \beta_{2i}(x_{xij}-\bar{x}_2) + \varepsilon_{ij}$$

$$\beta_{oi} = \gamma_0 + v_{oi}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} z_i + v_{1i}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21} z_i + v_{2i}$$

The effect of introducing an area level variable is to leave the unit level variance σ^2_ε unchanged but to reduce the area level components of variance by converting some of the between area differences to fixed effects related to z_i. Thus the covariance structure was:

$$\Omega = \begin{pmatrix} 1.364 & .239 & .242 \\ (.189) & (.045) & (.069) \\ & .081 & .017 \\ & (.017) & (.020) \\ & & .279 \\ & & (.047 \end{pmatrix}$$

$$\sigma^2_\varepsilon = 47.74 \quad (.345)$$

As might be expected the effect of introducing an area level explanatory variable is to improve the MSE of all the estimators but to reduce the difference between the Random Effect, the Diagonal and General Estimators.

Table 4: Properties of various estimators which include an arealevel covariate to the Random Intercept estimator with no area level covariate

Estimator	Population P2	P4
Diagonal (with z)	87 (92)	98 (100)
General (with z)	105 (105)	- (-)
General(2) (with z)	88 (93)	100 (101)
Random Intercept (with z)	87 (93)	92 (98)

N.B. First entry is ratio of Mean Square Error (%).

Second entry is ratio of Average Relative Bias (%).

Table 4 compares the various estimators using z with the Random Intercept estimator with no area level covariate. There are two General estimators since the full General estimator could not always be fitted to each simulated sample for the real data (P4). Consequently a second estimator (General(2)) was used in which the covariance terms σ_{01} and σ_{02} were set to zero. For the real data set (P4) it may be seen that the Diagonal and General(2) estimators are comparable to the Random Intercept estimator with no z but that the Random Intercept estimator (with z) is the most efficient. For the other population (P2) the Diagonal, General(2) and Random Intercept estimators (with z) are equally efficient and all are superior to the Random Intercept estimator without z.

6. Discussion

It is clear from many studies that there is no single approach to the small area estimation problem which will prove satisfactory for all situations. The size and homogeneity of the areas, the availability of auxiliary information, whether the variable of interest is continuous or discrete and the explanatory power of any assumed relationships will all affect the properties of any estimator. Approaches which prove suitable in some situations may be unsuitable in others.

The estimators considered are model based and are extensions to the random intercept regression model. The extension takes two forms: the introduction of random coefficient regression and the introduction of area level variables to model the between area components of variance.

Such approaches need a measure of precision (MSE) which is specific to each small area, and ideally also an estimator for this MSE which can be calculated as part of the analysis. These are provided for the class of multi-level models described here. However one also needs an understanding of when the MSE for small areas is likely to be improved by more complex estimators and confidence that the estimators are robust to departures from the model assumptions.

It is these last points which have been the main subject of this paper. The results are inconclusive. Some small area estimates will be improved by use of new estimators but others will be made worse. In some circumstances the overall performance will deteriorate and simpler estimators are to be preferred. This is particularly true of models which use several correlated random effects because the estimation of these terms can increase Mean Square Error. The introduction of area level covariates (z) on the other hand should be fruitful since it extends the estimation of small area means to areas for which no simple data is available. Diagnostic measures are needed which will help to guide the use of models for estimation purposes.

Acknowledgement

Mr Moura is on study leave from the Brazilian Government Statistical Agency (IBGE). Supported by CAPES (Brazilian Federal Agency for Post-Grad Education).

References

Battese, GE and Fuller, WA (1981). "Prediction of County Group Areas Using Survey and Satellite Data", in Proc Section on Survey Research Methods, Amer Stat Assoc pp500-505.

Battese, GE, Harter, RM and Fuller, WA (1988). "An Error Components Model for Prediction of County Group Areas Using Survey and Satellite Data", Amer Stat Assoc 83, pp28-36.

Goldstein, H (1986). "Multilevel mixed linear model analysis using iterative generalized least squares", Biometrika, 73, pp43-56.

Goldstein, H (1989). "Restricted Unbiased Iterative Generalized Least Squares Estimation", Biometrika, 76, pp622-623.

Henderson, CR (1975). "Best Linear Unbiased Estimation and Prediction Under a Selection Model", Biometrics, 31, p.423-447.

Holt, D and Moura, F (1993). Mixed Models for Making Small Area Estimates. In Kalton, G, Kordos, J and Platek, R, eds: Small Area Statistics and Survey Designs, 1, 221-231. Central Statistical Office, Warsaw.

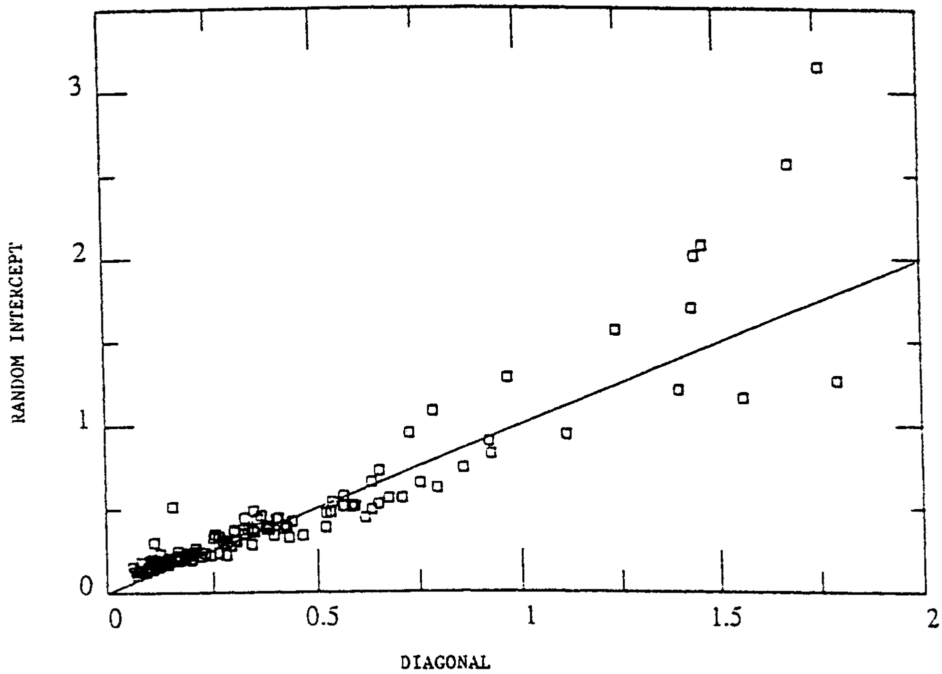
Kackar, RN and Harville, DA (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models", Amer Stat Assoc, 79, pp853-862.

Longford, N (1987). "A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Effects", Biometrika, 74, pp817-827.

ML3 Software for Three-level Analysis, Users' Guide for V.2 (1991). Institute of Education, University of London.

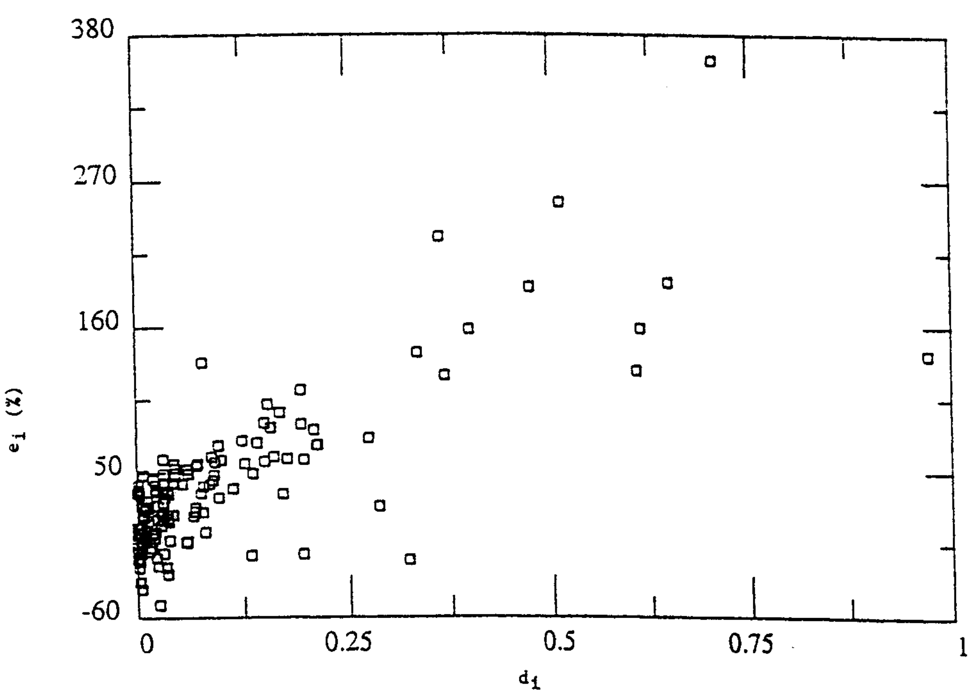
Prasad, NGN and Rao, JNK (1990). "The Estimation of the Mean Squared Error of Small-Area Estimators", J. Amer. Stat. Assoc., 85, pp163-171.

FIGURE 1: COMPARISON OF MSE FOR RANDOM INTERCEPT AND DIAGONAL ESTIMATOR FOR POPULATION P2



Note: One point omitted at (6.3, 6.8)

FIGURE 2: PLOT OF e_i VERSUS SMALL AREA DISPERSION d_i



Note: One point omitted at (3.2, 95.3)