QÜESTIIÓ, Vol 20, 2 pp. 187-212, 1996

SMALL-AREA ESTIMATION USING ADJUSTMENT BY COVARIATES

N.T. LONGFORD*

Linear regression models with random effects are applied to estimating the population means of indirectly measured variables in small areas. The proposed method, a hybrid with design- and model-based elements, takes account of the area-level variation and of the uncertainty about the fitted regression model and the area-level population means of the covariates. The method is illustrated on data from the U. S. Department of Labor Literacy Surveys and is informally validated on two states, Mississippi and Oregon, for which statewide surveys have been conducted.

Keywords: Effective sample size, linear regression, random effect, sampling variation.

1. INTRODUCTION

Large-scale educational surveys, such as the U. S. Department of Labor and National Adult Literacy Survey, the National Education Longitudinal Survey, and the National Assessment of Educational Progress in the United States, often provide abundant information about their target populations and certain subpopulations, but cannot be used directly for inference about small areas, such as states, counties, or census tracts. States or smaller administrative units often contract out smaller-scale surveys for their jurisdictions. Such surveys often do not utilize any information from the national surveys. As a result, some duplication in collection of information takes

* Nicholas T. Longford. Department of Medical Statistics, De Montfort University, The Gateway, Leicester LE1 9RH. England. Email: ntl@dmu.ac.uk

– Article rebut el maig de 1995.

– Acceptat el juny de 1996.

place. Considerable savings could be achieved and more information extracted if information about a small area contained in the national and small-area surveys could be combined, or indeed, if inference about a small area, based on the national sample, would use information ('borrow strength') from the other small areas.

This paper explores the extent to which information from national surveys can be used for inference about smaller units. We discuss in detail inference for states but the approach is equally applicable to other jurisdictions.

In a typical setup, a regression equation is fitted to the outcome variable y in the survey data, using an appropriately selected vector of covariates $\mathbf{x}$, yielding an estimate $\hat{\beta}$ of the vector of regression parameters β. Let the mean vector of the covariates $\mathbf{x}$ for small area j be $\mathbf{x}^{(j)}$, and let $\hat{\mathbf{x}}^{(j)}$ be its estimate, obtained not necessarily from the same survey. When no values of the outcome variable in area j are available,

$$\hat{y}^{(j)} = \hat{\mathbf{x}}^{(j)}\hat{\beta} \tag{1}$$

is the obvious estimator (predictor) of the mean of the outcome variable y in area j. When area j is represented in the survey, an estimator of the area mean, $\bar{y}^{(j)}$, based solely on the data from the area, can be combined with the synthetic estimator $\hat{y}^{(j)}$,

$$\hat{y}_c^{(j)} = a_j\bar{y}^{(j)} + (1 - a_j)\hat{y}^{(j)},$$

with the area-specific coefficient a_j chosen so as to minimize the mean squared error of the combined estimator $\hat{y}_c^{(j)}$.

This paper gives details of the prediction procedure outlined above and describes an application to the Job Training Partnership Act (JTPA) and the U. S. Employment Service and Unemployment Insurance (ES/UI) surveys administered by the U. S. Department of Labor in 1989–90. For details of these surveys, see Kirsch and Jungeblut (1992). The prediction procedure is an application of the approach of Battese, Harter, and Fuller (1988), and is here extended to account for sampling weights and uncertainty about the population mean of the covariates.

Section 2 summarizes the two-level (random-effects) regression model on which the predictions are based. Section 3 describes the adaptation of this model and its model fitting algorithms for sampling weights. Sections 4 and 5 give the minimal details of the datasets and the variables used. The methods are illustrated in Section 6 on examples that compare the prediction for Mississippi and Oregon based on the national surveys, with the estimates of the population means from the statewide sample surveys. Section 7 summarizes the paper, outlines a way of assessing the information about a small area in the national sample, and discusses how estimators for a small area can be combined.

2. LINEAR REGRESSION AND ITS USE IN PREDICTION

For an outcome variable y consider the random-effects regression model

$$y_{ij} = \mathbf{x}_{ij}\beta + \delta_j + \varepsilon_{ij}, \qquad (2)$$

where the subscripts i and j denote the elementary unit (subject) $i = 1, \ldots, n_j$ within area (e.g., state) $j = 1, \ldots N_2$; the random terms δ_j and ε_{ij} are mutually independent random variables with centered normal distributions and respective variances σ_2^2 and σ_1^2. Zero expectation of δ_j is not a restrictive assumption because a non-zero mean would be confounded in the regression $\mathbf{x}\beta$. Let p be the number of regression parameters (i.e., the length of the vector of covariates $\mathbf{x}_{ij}$). It is assumed throughout that the first component of $\mathbf{x}$ is equal to unity for each subject. The choice of the variables in $\mathbf{x}$ is a well-appreciated problem involving balancing the requirements of model parsimony and adequacy.

The random term δ_j can be interpreted as the deviation of area j from the national mean, after an adjustment for the covariates. The area-level variance σ_2^2 is a summary measure of the (adjusted) differences among the areas and it plays an important role in prediction for a state. To illustrate this, suppose the regression parameters β are known exactly. The synthetic predictor for area j with the known population mean vector of $\mathbf{x}$ equal to $\mathbf{x}^{(j)}$, is $\mathbf{x}^{(j)}\beta$. This predictor is not exact, though, because the 'true' value of the mean $\bar{y}^{(j)}$ is

$$\mathbf{x}^{(j)}\beta + \delta_j + \frac{1}{m_j}\sum_i^{m_j}\varepsilon_{ij}, \qquad (3)$$

where m_j is the population size of area j ($m_j > n_j$, unless a full census is taken in area j). When the area is not represented in the data ($n_j = 0$), no information about δ_j is available. Then the mean squared error of the predictor in (3) is $\delta_j^2 + \sigma_1^2/m_j$, and so its lower bound (approximate value for large m_j) is δ_j^2. The expectation of this lower bound over the small areas, σ_2^2, represents a component of uncertainty about the prediction for each area. If prediction based on a random-effects model is to be used for areas not represented in the data, the variables $\mathbf{x}$ should be selected so that the adjusted area-level variance σ_2^2 is as small as possible. When an area is represented in the survey, the data for area j, but also the data for the other areas, contain information about δ_j.

2.1. Sampling variation of the prediction

Suppose the sampling variance matrix of $\hat{\beta}$ is Σ_b, and it is estimated by $\hat{\Sigma}_b$. Details of estimating β and Σ_b are given in Section 3.

Let $\hat{\mathbf{x}}^{(j)}$ be an estimator of the mean vector of the covariates $\mathbf{x}$ for area j, with its sampling variance matrix Σ_S estimated by $\hat{\Sigma}_S$. In the illustration in Section , $\hat{\mathbf{x}}^{(j)}$ is the ratio estimator and its sampling variance matrix is derived assuming a weighted random sampling design. The predictor $\hat{y}^{(j)}$ given in (1) involves products of random variables (parameter estimates and sample means). If $\hat{\mathbf{x}}^{(j)}$ were subject to no error, that is, $\Sigma_S = \mathbf{0}$, the sampling variance of the predictor would be

$$\text{var}\left(\hat{y}^{(j)}\right) = \mathbf{x}^{(j)}\Sigma_b\left(\mathbf{x}^{(j)}\right)^{\top}. \tag{4}$$

When the mean $\hat{\mathbf{x}}^{(j)}$ is based on data not used in estimating β, $\hat{\mathbf{x}}^{(j)}$ and $\hat{\beta}$ are independent. Then

$$\text{var}\left(\hat{y}^{(j)}\right) = \text{tr}(\Sigma_S\Sigma_b) + \mathbf{x}^{(j)}\Sigma_b\left(\mathbf{x}^{(j)}\right)^{\top} + \beta^{\top}\Sigma_S\beta, \tag{5}$$

which may be much greater than the variance in (4); poorer information about the covariates causes poorer prediction.

Next, suppose $\hat{\beta}$ and $\hat{\mathbf{x}}^{(j)}$ are both normally distributed, and let Σ_{bS} be their covariance matrix. Then

$$\begin{aligned} \text{var}\left(\hat{y}^{(j)}\right) &= \text{tr}(\Sigma_S\Sigma_b) + \mathbf{x}^{(j)}\Sigma_b\left(\mathbf{x}^{(j)}\right)^{\top} + \beta^{\top}\Sigma_S\beta \\ &\quad + \text{tr}\left(\Sigma_{bS}^2\right) + 2\mathbf{x}^{(j)}\Sigma_{bS}\beta. \end{aligned} \tag{6}$$

See Appendix for proof. Equation (5) is obtained from (6) by setting $\Sigma_{bS} = \mathbf{0}$. Estimating the covariance matrix Σ_{bS} is a problem, especially when the data for the small area is used for estimating β and the estimator $\hat{\beta}$ has a complex form. Typically, no area constitutes a large proportion of the data, and so the correlations of $\hat{\beta}$ and $\hat{\mathbf{x}}^{(j)}$ are small. Then the terms involving Σ_{bS} in (6) can be ignored and (5) applies.

When $\hat{\mathbf{x}}^{(j)}$ and $\hat{\beta}$ are unbiased the expectation of the estimator $\hat{y}^{(j)}$ is $\mathbf{x}^{(j)}\beta + \text{tr}(\Sigma_{bS})$, and so the conditional mean squared error of $\hat{y}^{(j)}$, given δ_j, is

$$\mathbf{E}\left\{\left(\hat{y}^{(j)} - y^{(j)}\right)^2 \mid \delta_j\right\} = \text{var}\left(\hat{y}^{(j)}\right) + \left\{\delta_j - \text{tr}(\Sigma_{bS})\right\}^2. \tag{7}$$

When Σ_{bS} is not known the size of the bias and mean squared error can be inferred by substituting a range of plausible values for the matrix Σ_{bS} and for the deviation δ_j. When the small area is not represented in the survey sample $\Sigma_{bS} = \mathbf{0}$, and there is no information about δ_j. Then the typical mean squared error is obtained by averaging over the marginal distribution of δ_j:

$$\mathbf{E}\left[\mathbf{E}\left\{\left(\hat{y}^{(j)} - y^{(j)}\right)^2 \mid \delta_j\right\}\right] = \text{var}\left(\hat{y}^{(j)}\right) + \sigma_2^2. \tag{8}$$

For a cluster represented in the survey, the deviation δ_j can be estimated as its estimated conditional expectation given the data, see Section 3.1.

3. SAMPLING WEIGHTS

The sampling weights are an important feature of the sampling design of a survey. Inferences based on data from such a survey have to take account of the weights. For instance, the commonly used estimator of the mean of a simple random sample $y_i, i = 1, \ldots N$, is $\bar{y} = N^{-1} \sum_i y_i$. Its counterpart for independent data with sampling weights w_i is the 'weighted' mean, or the ratio estimator,

$$\bar{y}_w = \frac{\sum_i w_i y_i}{\sum_i w_i}.$$

Similarly, the regression parameter vector β, estimated for independent observations with equal weights as

$$\hat{\beta} = \left(\sum_i \mathbf{x}_i \mathbf{x}_i^{\top} \right)^{-1} \sum_i y_i \mathbf{x}_i ,$$

has the 'weighted' version

$$\hat{\beta}_w = \left(\sum_i w_i \mathbf{x}_i \mathbf{x}_i^{\top} \right)^{-1} \sum_i w_i y_i \mathbf{x}_i .$$

For estimation of the variances σ_2^2 and σ_1^2, as well as of the sampling variance matrix for $\hat{\beta}_w$, it is essential to use an appropriate normalization of the weights. Potthoff, Woodbury, and Manton (1992) show that the normalization in which the total of weights and the total of their squares are equal is appropriate. That is, the weights w_i are replaced by

$$w_i^* = w_i \frac{\sum_{i'} w_{i'}}{\sum_{i'} w_{i'}^2}.$$

It is assumed throughout that the weights have been normalized in this fashion, and the asterisk * on w is dropped. The total of the normalized weights, $N_w = \sum_i w_i$, can be interpreted as the *effective sample size*, that is, the size of a simple random sample that would contain the same amount of information about the population mean as the sample at hand. It can be shown that $N_w \leq N$, and equality holds only when the weights are constant.

When $\sigma_2^2 = 0$, the linear regression model in (2) simplifies to the ordinary regression. Then the estimators of the residual variance and the sampling variance of $\hat{\beta}$,

$$\hat{\sigma}^2 = \frac{1}{N_w - p} \sum_i (y_i - \mathbf{x}_i \hat{\beta})^2$$

$$\text{var}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_i w_i \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1},$$

are *approximately* unbiased.

Strictly speaking, the assumptions about the sampling weights listed in Potthoff *et al.* (1992) are not satisfied because the weights are random variables (owing to poststratification). However, we concur with Potthoff *et al.* that the consequences of randomness are not severe. Although the poststratified weights are derived by a complex process of adjustment for a number of background variables aggregated at various levels, the absolute changes of the weights are insubstantial relative to variation of the design weights.

3.1. Fitting random-effects models

The random-effects model in (2) can be fitted by the Fisher scoring algorithm. We list the relevant equations for the case of equal weights, and then describe the adaptation for unequal sampling weights.

Let $\mathbf{y}$ be the $N \times 1$ vector of outcomes, $\mathbf{X}$ the $N \times p$ matrix of regressors, $\mathbf{e} = \mathbf{y} - \mathbf{X}\beta$ the vector of residuals, and $\mathbf{V} = \text{var}(\mathbf{y})$ the variance matrix of the outcomes.

The first and second-order partial derivatives of the log-likelihood

$$l = -\frac{1}{2} \left\{ N \log(2\pi) + \log(\det \mathbf{V}) + \mathbf{e}^\top \mathbf{V}^{-1} \mathbf{e} \right\} \tag{9}$$

with respect to the regression parameters are

$$\frac{\partial l}{\partial \beta} = \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{e}$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta^\top} = -\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}.$$

An iteration of the Fisher scoring algorithm updates a current estimate $\hat{\beta}_{old}$ to obtain the 'new' estimate

$$\hat{\beta}_{new} = \hat{\beta}_{old} + \left(\mathbf{X}^\top \mathbf{V}_{old}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^\top \mathbf{V}_{old}^{-1} \mathbf{e}_{old},$$

where $\mathbf{V}_{old}$ and $\mathbf{e}_{old}$ are equal to $\mathbf{V}$ and $\mathbf{e}$ evaluated for the current values of the parameter estimates. Substitution for $\mathbf{e}_{old}$ and elementary algebra yield

$$\hat{\beta}_{new} = \left(\mathbf{X}^\top \mathbf{V}_{old}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top \mathbf{V}_{old}^{-1}\mathbf{y}; \tag{10}$$

the Fisher scoring and Newton-Raphson algorithms coincide with the generalized least squares.

In evaluating (10), inversion of large matrices is avoided by exploiting the pattern of the variance matrix $\mathbf{V}$. First, $\mathbf{V}$ is block-diagonal, with blocks

$$\mathbf{V}_j = \sigma_1^2\mathbf{I}_{n_j} + \sigma_2^2\mathbf{J}_{n_j}$$

correspoding to areas ($\mathbf{I}_m$ and is the $m\times m$ unit matrix and $\mathbf{J}_m = \mathbf{1}_m\mathbf{1}_m^\top$ the $m\times m$ matrix of ones). Next,

$$\begin{aligned} \det(\mathbf{V}_j) &= \sigma_1^{2n_j}(1+n_j\tau) \\ \mathbf{V}_j^{-1} &= \sigma_1^{-2}\left(\mathbf{I}_{n_j} - \frac{\tau}{1+n_j\tau}\mathbf{J}_{n_j}\right), \end{aligned}$$

where $\tau = \sigma_2^2/\sigma_1^2$.

Some advantage is gained by using the variance ratio τ instead of σ_2^2. Letting $\mathbf{W} = \sigma_1^{-2}\mathbf{V}$, the variance σ_1^2 can be separated out in the log-likelihood l,

$$l = -\frac{1}{2}\left\{N\log(2\pi\sigma_1^2) + \log(\det\mathbf{W}) + \sigma_1^{-2}\mathbf{e}^\top\mathbf{W}^{-1}\mathbf{e}\right\}. \tag{11}$$

The first-order partial derivative with respect to σ_1^2 has the root

$$\hat{\sigma}_1^2 = \frac{\mathbf{e}_{old}^\top\mathbf{W}_{old}^{-1}\mathbf{e}_{old}}{N}. \tag{12}$$

Finally, noting that $\partial\mathbf{W}/\partial\tau = \mathrm{diag}_j\left\{\mathbf{J}_{n_j}\right\}$,

$$\begin{aligned} \frac{\partial l}{\partial\tau} &= -\frac{1}{2}\sum_j \mathbf{1}_{n_j}^\top\mathbf{W}_j^{-1}\mathbf{1}_{n_j} + \frac{1}{2\sigma_1^2}\sum_j\left(\mathbf{e}_j^\top\mathbf{W}_j^{-1}\mathbf{1}_{n_j}\right)^2 \\ -\mathbf{E}\left(\frac{\partial^2 l}{\partial\tau^2}\right) &= \frac{1}{2}\sum_j\left(\mathbf{1}_{n_j}^\top\mathbf{W}_j^{-1}\mathbf{1}_{n_j}\right)^2, \end{aligned}$$

where $\mathbf{e}_j$ is the subvector of $\mathbf{e}$ corresponding to area j and $\mathbf{W}_j = \sigma_1^{-2}\mathbf{V}_j$. At each iteration the current estimate of τ is updated as

$$\hat{\tau}_{new} = \hat{\tau}_{old} - \left\{ \mathbf{E}\left(\frac{\partial^2 l}{\partial \tau^2} \right) \right\}^{-1} \frac{\partial l}{\partial \tau}, \tag{13}$$

with the right-hand side evaluated at the current solution $(\hat{\beta}_{old}, \hat{\sigma}^2_{1,old}, \hat{\tau}_{old})$. The iterations, consisting of (10), (12), and (13), are terminated when a convergence criterion is satisfied. Such a criterion can be based on the change of the log-likelihood, $l_{new} - l_{old}$, on the size of the corrections for the estimates, the norm of the score vector, or a combination of these criteria.

The algorithm requires the following statistics: the totals of squares and cross-products, $(\mathbf{y},\mathbf{X})^\top(\mathbf{y},\mathbf{X})$, and within-area totals $(\mathbf{y}_j, \mathbf{X}_j)^\top \mathbf{1}_{n_j}$. The algorithm is adapted for sampling weights by replacing these summaries by their weighted versions, $(\mathbf{y}_j, \mathbf{X}_j)^\top \mathbf{w}_j$ and $(\mathbf{y},\mathbf{X})^\top \mathrm{diag}(\mathbf{w})(\mathbf{y},\mathbf{X})$, where $\mathbf{w}_j$ is the $n_j \times 1$ vector of sampling weights for area j, and $\mathbf{w}$ is the $N \times 1$ vector of weights for the entire sample. Note that the area-level sample size n_j is replaced by the total weight $\sum_i w_{ij}$.

An integral part of the algorithm is estimation of the realized values of δ_j as the conditional expectations and of their precision as the conditional variances of δ_j given the data and the parameter estimates:

$$\begin{aligned} \mathbf{E}(\delta_j \mid \mathbf{y}_j) &= \frac{\tau \mathbf{e}_j^\top \mathbf{1}_{n_j}}{1 + n_j \tau} \\ \mathrm{var}(\delta_j \mid \mathbf{y}_j) &= \frac{\sigma_2^2}{1 + n_j \tau}. \end{aligned}$$

The weighted versions of these equations are obtained by replacing $\mathbf{e}_j^\top \mathbf{1}_{n_j}$ with $\sum_i e_{ij} w_{ij}$ and n_j with $\sum_i w_{ij}$.

The Fisher scoring algorithm can be straightforwardly adjusted for restricted maximum likelihood estimation (REML), see Harville (1974). The log-likelihood in (11) is adjusted by the term

$$\Delta l_R = \frac{1}{2} \log \left\{ \det \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} \right) \right\},$$

and the scoring vector and information matrix by the corresponding partial derivatives. For instance, in the equation for $\hat{\sigma}_1^2$, (12), the denominator is reduced from N to $N - p$, taking account of the uncertainty about the p regression parameter estimates. The approach based on the best linear unbiased prediction (BLUP, Kackar and Harville, 1984) can also be adapted for sampling weights. These approaches rely on normality of the random terms and on linearity. For alternative approaches, see Beran and Hall (1992) and references therein.

4. DATA

Data from two national surveys of adult literacy, ES/UI and JTPA, administered by the U. S. Department of Labor, and data from surveys from the states of Mississippi and Oregon are available. The surveys are designed to assess the nature and extent of the literacy skills of the U. S. adult population aged 16 and over.

The survey instrument, common to all four surveys, consists of a 15-minute background questionnaire and a 45-minute set of exercises. The design of each survey is that of a stratified multi-stage clustered sample. The target population in the national surveys was stratified to seven geographical regions and U. S. states were drawn from each region with replacement as the primary sampling units. There were some minor differences in the definitions of the clusters between the two national surveys. The sampling weights were derived by adjustment of the design weights due to poststratification. For the purposes of illustration we treat the poststratified weights as if they were the design weights, and we ignore all levels of clustering except state level.

The sample sizes and the effective sample sizes (N_w) are given in Table 1. The coefficient of variation of the weights, given in the third row of the table, is defined as the ratio of the sample variance and the square of the sample mean of the weights,

$$\rho = \frac{\text{var}(w_.)}{\bar{w}^2}.$$

Table 1

Raw and effective sample sizes in the adult literacy surveys

	Survey			
	ES/UI	JTPA	Mississippi	Oregon
Sample size	3277	2501	1804	1993
Effective sample size	1403	1046	1629	1854
Coefficient of variation	1.34	1.39	0.11	0.08

It is a useful indicator of how much smaller the effective sample size is compared to the raw sample size. Constant weights correspond to $\rho = 0$. In the national surveys (ES/UI and JTPA) the weights vary considerably; for instance, the ratio of the largest and smallest weights is 86.5 and 53.5 in ES/UI and JTPA, respectively. In the statewide surveys, these ratios are only 10.6 (Mississippi) and 3.9 (Oregon). The histograms of the normalized sampling weights are drawn in Figure 1. Each national survey dataset contains less information, in the sense of the effective sample size, than either dataset from the state surveys, even though the former have greater (raw) sample sizes.

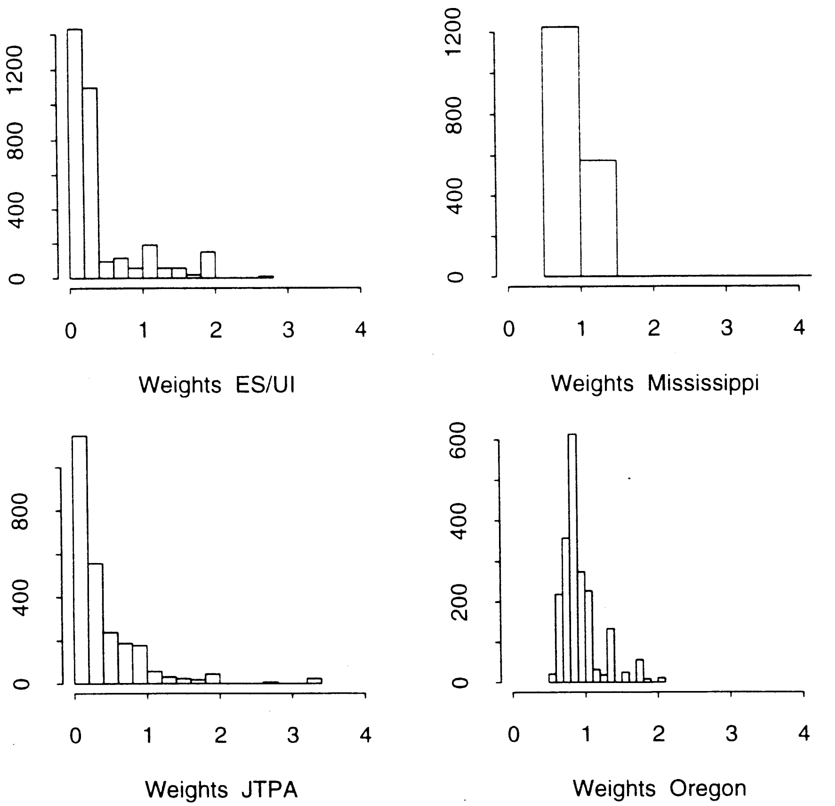

Figure 1. ***Histograms of the normalized sampling weights for the four surveys. The same scale for the horizontal axis is used in all four panels.***

The principal outcome variable, called the (literacy) *proficiency score*, is defined on a scale common to all four surveys. The probability of correct response of subject i with proficiency score θ_i to question k with item parameters (a_k, b_k, c_k) is modelled as

$$P(Z_{ik} = 1 \mid \theta_i; a_k, b_k, c_k) = c_k + (1 - c_k)\frac{\exp(a_k + b_k\theta_i)}{1 + \exp(a_k + b_k\theta_i)};$$

the parameters a_k, b_k, and c_k can be interpreted as the difficulty, the discrimination, and the probability of guessing, respectively. In the item response model applied, the responses Z_{ik} are assumed to be conditionally independent given θ_i, and a normal prior distribution is imposed for $\{\theta_i\}$. Note that the proficiency scores $\{\theta_i\}$ are confounded with the item parameters $\{a_k\}$ and $\{b_k\}$, and so the mean and variance of the prior distribution for $\{\theta_i\}$ are set merely to ensure identifiability. The proficiencies as well as their estimates are in the range $(-\infty, +\infty)$; the value of zero is of no special significance.

The survey subjects' scores on this scale are estimated using a marginal maximum likelihood approach (Bock and Aitkin, 1981, and Mislevy and Bock, 1983). Inferences about the proficiency scores derived by regarding the estimates $\hat{\theta}_i$ as the true values θ_i are likely to underestimate the precision because they ignore the substantial sampling variance of each estimate $\hat{\theta}_i$. To take account of the uncertainty associated with the *estimated* proficiency scores, a set of five *imputed values* are randomly drawn from the approximation to the posterior distribution of the proficiency scores. The decision to use five imputed values was based on extensive simulations. Any analysis of the proficiency scores (e.g., regression) involves identical analyses using each set of the five imputed values. Let $\hat{\beta}_h$, $h = 1, \ldots, 5$, be a quintet of such estimates. Then the estimate that refers to the proficiency scores is

$$\hat{\beta} = \frac{1}{5} \sum_h \hat{\beta}_h .$$

The standard errors for the parameters that refer to the proficiency scores are obtained similarly, but they have to be inflated by the variance of the estimates across the five analyses. Suppose s_h^2 is the estimated sampling variance of $\hat{\beta}_h$. Then the sampling variance of $\hat{\beta}$ is estimated as

$$s^2 = \frac{1}{5} \sum_h s_h^2 + \frac{1}{4} \sum_h \left(\hat{\beta}_h - \hat{\beta} \right)^2 .$$

The main purpose of the study is to validate the outlined method; improvement in the estimation of population means for the particular two states is of lesser importance. For this purpose, we estimate the population means for the two states based on the national survey data and the covariates for the state-wide surveys, treating the within-state data on y as a 'hold-out' dataset. These estimates and their standard errors are then compared with their counterparts based solely on the imputed values for the within-state samples. In the concluding section we discuss how such pairs of estimators can be combined.

The histograms of the first sets of imputed values are displayed in Figure 2. The mean of the sample for Oregon is somewhat higher than that for Mississippi, and the sample variances in the state surveys are smaller than their national counterparts. There is a perceptible difference in the distributions for the national surveys (JTPA has smaller variance than ES/UI), although it may be accounted for by the varying weights. These comparisons carry over from the imputed values to the proficiency scores.

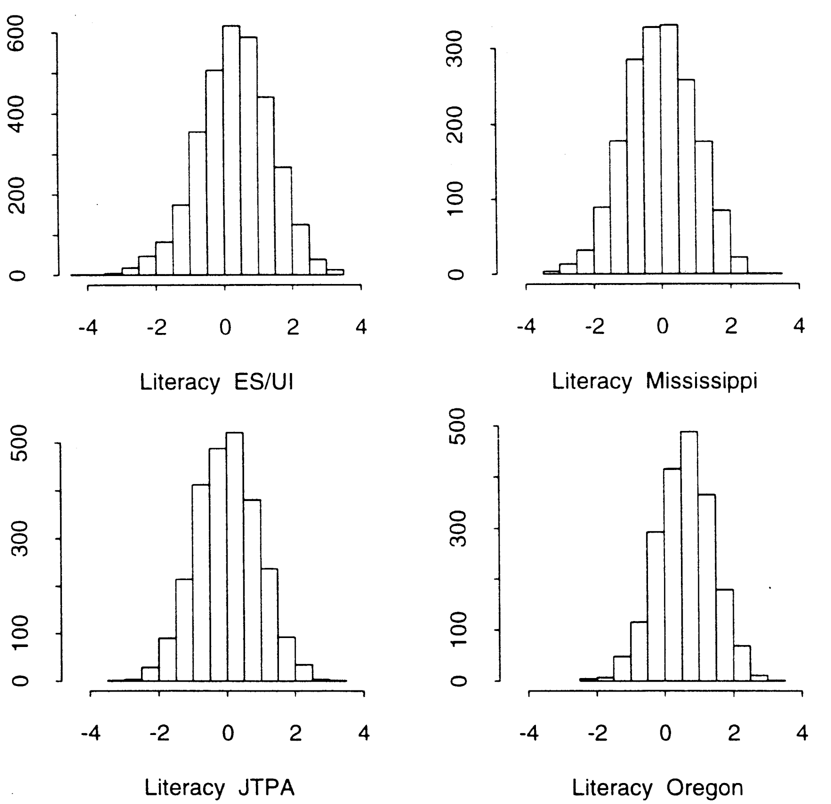

Figure 2. ***Histograms of the first sets of imputed values for the four surveys. The same scale for the horizontal axis is used in all four panels.***

5. ANALYSIS

For illustration, we describe the prediction for Mississippi and Oregon using the following set of predictor variables:

Sex (dichotomous);
Ethnicity (six nominal categories);

Educational level (five ordinal categories);
Age category (five ordinal categories);
More English (dichotomous);
More Mathematics (dichotomous);
Mother's education (quantitative);
Father's education (quantitative);
Personal income (quantitative);
Household income (quantitative);
C.Age (quantitative, in years);
Q.Age (quantitative, in years^2/1000).

The variable *Sex* is coded 0 for men and 1 for women. The other dichotomous variables have values 2 ('thinks more English/Mathematics would be useful') and 1 ('does not think so'). The variables related to income and parents' education are defined on the integer scale 1–11. See Kirsch and Jungeblut (1992) for details. The variable *C.Age* is the subject's age in years, and *Q.Age* is defined as $C.Age^2/1000$. For parsimony, they are used as quantitative variables, so that they are represented only by one parameter each.

The means and proportions (as applicable) for the covariates and for the four datasets are given in Table 2. The regression model fit for the ES/UI data using the first set of imputed values and the set of covariates listed above, is summarized in Table 3. For comparison, the ordinary least squares fit (OLS) and the maximum likelihood (MLE) fit are given. The likelihood ratio test statistic, equal to the difference of the OLS and MLE deviances (the values of -2 log-likelihood), can be used to assess the significance of the state-level variance σ_2^2. The value of this statistic is $3380.50 - 3354.89 = 25.61$; its approximate (asymptotic) null-distribution is χ_1^2. Thus, the state-level variation is highly significant. Note however, that the t-statistic for σ_2^2 is nominally not significant at the 5 per cent level.

To avoid problems with missing data, listwise deletion was applied; all records which contain a missing observation for at least one of the covariates were deleted. This reduced the sample size for the ES/UI data from 3277 to 3219, and the effective sample size from 1402.5 to 1378.4. Listwise deletion for the same covariates in the JTPA dataset resulted in reduction of the effective sample size from 1047.6 to 1023.9.

A model-based method for full use of incomplete data in random effects models is described in Longford (1995). Its implementation requires specification of a model for the process that gives rise to missing data. It may reduce the bias due to the informative nature of this process, but improvement in efficiency of the prediction is likely to be insubstantial because of the low proportion of missing data. The means and proportions in Table 2 refer to the entire sample; the corresponding figures for the listwise deleted samples differ insubstantially.

Table 2

Weighted sample means and proportions of the covariates

	Survey							
	ES/UI		JTPA		Mississippi		Oregon	
Variable	Proprt'n or mean	Count	Proprt'n or mean	Count	Proprt'n or mean	Count	Proprt'n or mean	Count
Sex 1	0.56	1756	0.41	1008	0.47	761	0.50	1064
2	0.44	1515	0.58	1484	0.53	1043	0.50	929
Ethnicity 1	0.63	2394	0.69	1556	0.66	1300	0.92	1845
2	0.12	375	0.21	663	0.32	473	0.01	16
3	0.20	384	0.06	159	0.01	20	0.03	57
4	0.02	40	0.00	17	0.00	2	0.02	28
5	0.01	48	0.03	76	0.00	5	0.02	28
6	0.02	36	0.01	30	0.00	4	0.01	19
Education 1	0.03	135	0.07	202	0.13	217	0.02	34
2	0.18	619	0.33	871	0.19	349	0.15	282
3	0.59	2006	0.55	1295	0.48	872	0.56	1104
4	0.19	513	0.06	130	0.19	362	0.27	570
5	0.00	4	0.00	3	0.00	1	0.00	1
Age categ. 1	0.10	314	0.17	489	0.09	140	0.09	138
2	0.18	616	0.19	485	0.11	173	0.12	183
3	0.22	727	0.21	505	0.13	238	0.17	301
4	0.32	1059	0.31	733	0.27	477	0.35	778
5	0.17	546	0.10	259	0.40	776	0.28	593
More Engl. 1	0.57	1792	0.66	1717	0.52	889	0.39	745
2	0.42	1461	0.33	762	0.47	901	0.61	1246
More Maths 1	0.69	2231	0.79	2002	0.61	1054	0.52	1003
2	0.30	1020	0.20	473	0.38	732	0.48	988
Mother's Ed.	3.77		3.67		4.38		4.31	
Father's Ed.	4.37		4.32		5.07		4.86	
Pers. Income	3.29		2.29		3.99		3.86	
Hous. Income	5.04		3.41		4.95		5.55	
Age (years)	33.84		30.67		42.64		37.91	
Quadr. age	1287.91		1061.96		2113.72		1612.45	

Note: For categorical variables the proportions are accompanied by the counts of subjects in the category. The counts of all the categories of a variable do not add up to the sample size because of missing data.

Table 3

Regression model fits for ES/UI using OLS and MLE

Parameter	OLS Estimate	OLS St. error	MLE Estimate	MLE St. error
Intercept	−1.148	(0.470)	−1.171	(0.465)
Sex 2–1	−0.199	(0.050)	−0.204	(0.052)
Ethnicity 2–1	−0.648	(0.080)	−0.641	(0.064)
3–1	−0.522	(0.067)	−0.449	(0.072)
4–1	−0.702	(0.174)	−0.679	(0.173)
5–1	−0.575	(0.212)	−0.592	(0.210)
6–1	−0.124	(0.200)	−0.102	(0.199)
Education 2–1	0.806	(0.155)	0.813	(0.153)
3–1	1.184	(0.148)	1.187	(0.147)
4–1	1.685	(0.158)	1.686	(0.157)
5–1	−0.111	(0.689)	−0.130	(0.681)
Age category 2–1	−0.287	(0.117)	−0.287	(0.116)
3–1	−0.257	(0.172)	−0.271	(0.170)
4–1	−0.373	(0.250)	−0.386	(0.247)
5–1	−0.198	(0.322)	−0.021	(0.318)
More English	0.547	(0.063)	0.539	(0.063)
More Maths	−0.055	(0.067)	−0.053	(0.067)
Mother's educ.	0.0190	(0.0093)	0.0187	(0.0092)
Father's educ.	0.0037	(0.0078)	0.0043	(0.0077)
Personal income	0.031	(0.014)	0.030	(0.014)
Household income	0.033	(0.011)	0.034	(0.011)
Age (years)	0.02236	(0.02810)	0.02288	(0.02777)
Quadratic age	−0.00038	(0.00029)	−0.00038	(0.00029)
σ_1^2	0.7414		0.7404	
τ			0.0098	(0.0083)
Deviance	3380.50		3354.89	

Table 4

Regression model fits for JTPA using OLS and MLE

Parameter	OLS Estimate	OLS St. error	MLE Estimate	MLE St. error
Intercept	−1.232	(0.470)	−1.244	(0.465)
Sex 2–1	−0.149	(0.052)	−0.148	(0.052)
Ethnicity 2–1	−0.620	(0.064)	−0.621	(0.064)
3–1	−0.699	(0.108)	−0.703	(0.107)
4–1	0.338	(0.380)	0.332	(0.375)
5–1	−0.0036	(0.160)	−0.0016	(0.159)
6–1	−0.548	(0.313)	−0.550	(0.310)
Education 2–1	0.374	(0.105)	0.374	(0.104)
3–1	0.839	(0.103)	0.832	(0.102)
4–1	1.257	(0.149)	1.254	(0.148)
5–1	0.545	(1.989)	0.547	(1.962)
Age category 2–1	0.047	(0.110)	0.049	(0.109)
3–1	−0.035	(0.172)	−0.034	(0.170)
4–1	0.032	(0.254)	0.029	(0.251)
5–1	0.077	(0.334)	0.073	(0.330)
More English	0.404	(0.062)	0.406	(0.062)
More Maths	−0.157	(0.072)	−0.154	(0.071)
Mother's educ.	0.0176	(0.0089)	0.0178	(0.0088)
Father's educ.	0.0018	(0.0074)	0.0018	(0.0073)
Personal income	−0.022	(0.011)	−0.021	(0.011)
Household income	0.003	(0.010)	0.0033	(0.0096)
Age (years)	0.03458	(0.03044)	0.03545	(0.03009)
Quadratic age	−0.00046	(0.00034)	−0.00047	(0.00033)
σ_1^2	0.5841		0.5838	
τ			0.0043	(0.0084)
Deviance	2297.88		2274.13	

Some of the estimated regression parameters in Table 3 are difficult to interpret. For instance, the parameters associated with the categories of *Education* are not in monotone order. However, this is of little importance since improved prediction is the sole purpose of the fitted regression model. Note that some of the variables (e.g., *Age category*) can be deleted from the model without substantial deterioration of the fit. Also, category 5 of *Education* is associated with very large standard error because it is represented by very few subjects. The category could be collapsed with category 4. For assessing the importance of a set of variables the likelihood ratio test is preferable to separate t-tests for each variable. Refinement of the model is dealt with in Section 6.1.

Instead of the state-level variance σ_2^2, the variance ratio $\tau = \sigma_2^2/\sigma_1^2$ and its standard error are estimated. Thus, the estimated state-level variance is $\hat{\sigma}_2^2 = 0.7404 \times 0.0098 = 0.00726$, and the corresponding standard deviation is $\hat{\sigma}_2 = \sqrt{0.00726} = 0.085$. This can be interpreted as the expected difference between the regression for a randomly drawn state and the 'average' regression given by the regression parameter vector β. The state-level variance is a source of uncertainty of the prediction for each state not represented in the survey.

The regression model fits for the JTPA data using the first set of imputed values are given in Table 4, in the same format as in Table 3. To conserve space, the regression model fits for the other sets of imputed values are not given. The estimated regression parameters for the proficiency scores are obtained by averaging over the five analyses. They are of little interest in the present context, because the predictions based on each set of imputed values will be averaged to obtain the prediction based on the proficiency scores. This way, uncertainty in estimation of the proficiency scores is allowed to permeate through all the stages of prediction.

In general, there is a lot of variation in the estimated parameters across the imputed values, as well as between the datasets. This may be of little consequence, though, because the substantially different regression parameter estimates may yield very similar predictions.

From each MLE model fit the inverse of the information matrix is stored, because it is used in estimation of the sampling variance of the prediction.

6. PREDICTION

In this section we describe prediction of the means of the proficiency scores for Mississippi and Oregon, using the national surveys and the covariate information from the surveys for these states. The equations for the predicted means and their standard errors, assuming accurately observed proficiency scores, are given by (1) and

(5). The five predictions (one for each set of imputed values) are then combined to obtain estimates which take into account the uncertainty in estimating the proficiency scores of the sampled individuals. The within-state means $\mathbf{x}^{(j)}$ were estimated by the ratio estimator, and their standard errors were obtained under the assumption of weighted random sampling design (we failed to obtain information that would identify the clusters in the state-wide surveys).

The two panels in Table 5 summarize this prediction for Mississippi and Oregon. Of principal interest is the right-most column, generated by the results based on the sets of imputed values. For comparison, the weighted sample mean ('observed' mean) and its sampling standard error from the respective statewide surveys are given in the last two lines of each panel.

Table 5

Prediction of the means of the proficiency scores for Mississippi and Oregon

		Imputed value					Prof-cy
Survey		1	2	3	4	5	score
Mississippi							
ES/UI	Mean	0.135	0.112	0.169	0.154	0.161	0.146
	St. error	0.065	0.068	0.062	0.060	0.058	0.067
JTPA	Mean	0.067	0.084	0.099	0.109	–0.006	0.071
	St. error	0.084	0.081	0.082	0.083	0.085	0.095
Miss.	Obs. mean	–0.117	–0.122	–0.108	–0.127	–0.124	–0.120
	St. error	0.026	0.026	0.026	0.027	0.026	0.027
Oregon							
ES/UI	Mean	0.624	0.603	0.625	0.629	0.641	0.624
	St. error	0.043	0.047	0.039	0.035	0.032	0.042
JTPA	Mean	0.486	0.516	0.539	0.495	0.467	0.501
	St. error	0.062	0.059	0.060	0.061	0.063	0.067
Oregon	Obs. mean	0.575	0.561	0.567	0.579	0.572	0.571
	St. error	0.019	0.019	0.019	0.019	0.019	0.020

The weighted sample means for both Mississippi and Oregon are at the extremes of the distribution for the within-state means of proficiency scores. The weighted within-state sample means in the ES/UI dataset are greater than that of Mississippi for all states that are represented by 25 or more subjects. In the JTPA dataset only Missouri (–0.29) has a lower weighted sample mean than Mississippi. All the within-state means in the JTPA dataset are lower than the mean for Oregon; in the ES/UI dataset Maryland (0.62) and Utah (0.63) have a slightly higher mean than Oregon, and the mean for Massachussets is 0.89. The weighted (national) sample means in ES/UI and JTPA are 0.30 and 0.04, respectively.

The predictions for Mississippi (0.146 and 0.071) are much higher than the observed weighted mean (–0.120), and the estimated standard errors for the prediction and the sample mean are too small to account for the discrepancy. For Oregon, the prediction appears to be much more successful; the discrepancy of the prediction from the observed mean is well within the estimated sampling error. Note that the standard errors for the prediction for Oregon are smaller that those for Mississippi. More detailed analysis of the sources of uncertainty in prediction can be carried out by comparing the three components of the sampling variance in (5).

The standard errors for prediction take no account of the (estimated) state-level variance. The estimates of the state-level variances for the respective surveys ES/UI and JTPA, averaged over the five analyses, are 0.00322 (standard error 0.00480) and 0.00095 (0.00414). These variances, if taken at face value, are by no means ignorable. Combined with the standard errors quoted in Table 5, on average, in the sense of (8), they inflate the standard errors for Mississippi from 0.067 to 0.087 (prediction based on ES/UI data), and from 0.095 to 0.099 (JTPA). The corresponding increases for Oregon are from 0.042 to 0.070 (ES/UI), and from 0.067 to 0.074 (JTPA). Note that the variance σ_2^2 is estimated with relatively little precision in both surveys.

6.1. Refinement of the model

The regression model given by the covariates $\mathbf{x}$ is a key element of the prediction procedure. Adequate model fit and small state-level variation are likely to be achieved by supplementing the covariates listed in Table 3 with further variables. Substantive information about the descriptive power and small reduction of the data by listwise deletion are two important criteria for selecting such variables.

First, we consider supplementing the regression model with the following covariates (abbreviated names, and for categorical variables the number of categories, are given in parentheses):

Enrolled in school? (*Sch?*, 2);
High school diploma? (*H.S.*, 2);

Military service? (*Mil.S.*, 2);
Registered to vote? (*RgVot*, 2);
How often use math on the job? (*MatU*, 5);
Reading skills good enough for your job? (*ReadJ*, 3);
Writing skills good enough for your job? (*WritJ*, 3);
Math skills good enough for your job? (*MathJ*, 3);
Better job if more English training? (*MorEn*, 2);
Better job if more math training? (*MorEn*, 2);
How often read English newspaper? (*ENws*, 5);
How many people in household? (*#Hhld*);
How often read/use reports on the job? (*Reps*);
How often write memos on the job? (*Mems*).

A number of other variables, related to the use of English language, employment, income, education, and housing, were not considered because they were either defined only for a small fraction of the population (e.g., not applicable for many subjects), or they were not collected in the statewide surveys.

Listwise deletion using these variables led to a reduction of the dataset by about 900 subjects in both datasets; from 3277 to 2367 for ES/UI, and from 2501 to 1604 for JTPA. Inclusion in the regression model of the variables listed above resulted in small changes of the state-level variance for both datasets; the within-state variance was reduced by about 13 per cent for the ES/UI dataset, and by about 8 per cent for the JTPA dataset. Most of the regression parameters are nominally statistically significant (at the 5 percent level). For brevity, details of the regression fits are omitted.

The standard errors of the prediction based on these variables are slightly higher than for the reduced model. Improvement in the model fit is negated by the reduced effective sample size. The estimates for Mississippi are 0.278 (standard error 0.067) based on the ES/UI survey, and 0.135 (0.105) based on JTPA. They are even more distant from the weighted means based on the respective statewide surveys than the estimates given in Table 5. The predictions for Oregon are changed only slightly (0.665 for ES/UI and 0.513 for JTPA).

The 900 or so subjects discarded by listwise deletion for each dataset are informative subsamples of the respective datasets. For instance, the weighted mean for the 1604 included subjects for Mississippi is –0.021, while the weighted mean for the entire sample of 2501 subjects is –0.124. The corresponding means for Oregon are 0.587 and 0.520.

Clearly, too high a price is paid for inclusion of many variables in the prediction model. In the next round of refinement we drop the quadratic age term (it has a low t-ratio for all model fits), collapse the categories 4 – 6 of *Ethnicity* (Native Americans, Asian Americans, and 'Other', each with small counts), collapse the categories 4 and

5 of *Education* (very few subjects in category 5), discard variables *H.S.*, *Mil.S.*, *MatU*, *ReadJ*, *WritJ*, *MathJ*, *ENws*, *#Hhld*, *Reps*, and *Mems* because their values are missing for large numbers of subjects, and/or they are not important in the regression model.

Table 6

Regression model fits for ES/UI and JTPA; maximum likelihood estimates

	ES/UI		JTPA	
Parameter	Estimate	St. error	Estimate	St. error
Intercept	−0.236	(0.245)	−0.345	(0.264)
Sex 2–1	−0.197	(0.050)	−0.125	(0.069)
Ethnicity 2–1	−0.703	(0.079)	−0.689	(0.078)
3–1	−0.510	(0.069)	−0.555	(0.080)
(>3)–1	−0.443	(0.116)	0.367	(0.145)
Education 2–1	0.751	(0.159)	0.690	(0.171)
3–1	1.091	(0.151)	1.051	(0.151)
(>3)–1	1.528	(0.160)	1.489	(0.168)
Age category 2–1	−0.028	(0.106)	−0.032	(0.104)
3–1	0.060	(0.120)	0.089	(0.120)
4–1	0.058	(0.163)	0.126	(0.171)
5–1	0.198	(0.262)	0.244	(0.269)
In School?	−0.259	(0.072)	−0.212	(0.077)
Reg. Voter?	−0.234	(0.055)	−0.181	(0.068)
More English	0.461	(0.066)	0.437	(0.065)
More Maths	−0.018	(0.067)	−0.041	(0.071)
Engl. pps 2–1	0.062	(0.057)	0.067	(0.057)
3–1	−0.140	(0.083)	−0.127	(0.082)
4–1	−0.231	(0.110)	−0.165	(0.122)
5–1	−0.329	(0.175)	−0.293	(0.178)
Mother's educ.	0.0163	(0.0093)	0.0176	(0.0086)
Father's educ.	−0.0026	(0.0077)	−0.0030	(0.0075)
Personal income	0.026	(0.014)	0.015	(0.013)
Household income	0.027	(0.011)	0.021	(0.011)
Age (years)	−0.0128	(0.0067)	−0.0109	(0.0069)
σ_1^2	0.7332		0.5825	
τ	0.0068	(0.0078)	0.0039	(0.0078)

Table 7

Predictions for Mississippi and Oregon based on the original model (see Table 5) and the refined model

State	ES/UI		JTPA		State	
Model	Mean	St. error	Mean	St. error	Mean	St. error
Mississippi						
Original	0.146	(0.067)	0.071	(0.095)	–0.120	(0.027)
Refined	0.140	(0.060)	0.091	(0.084)	–0.107	(0.027)
Oregon						
Original	0.624	(0.042)	0.501	(0.067)	0.571	(0.020)
Refined	0.648	(0.039)	0.499	(0.067)	0.570	(0.021)

The new regression model contains 25 regression parameters (degrees of freedom). The respective raw and effective sample sizes after listwise deletion are 3089 and 1323 for ES/UI, and 2364 and 991 for JTPA (compare with Table 1). Thus, the selected model is a compromise of adequacy (more covariates), and small loss of observations due to listwise deletion.

Table 6 contains the maximum likelihood fits to the two national survey datasets (averaged over the five sets of imputed values). Although the covariates included at the last (refinement) stage are nominally significant, their impact on the model fit is only marginal, especially for the JTPA dataset. The subject-level variance estimates are reduced insubstantially, and the variance ratio is reduced only for ES/UI (by about 30 per cent).

The predictions for Mississippi and Oregon using the original and the refined models are summarized in Table 7. The predictions of the state means, based on the refined model differ from their counterparts based on the original model (Table 5) only slightly, and the differences in predictions are trivial compared to the standard errors. The standard errors are reduced somewhat, but the conclusions drawn using the original model are not affected.

We see that even though the model fit can be improved without sacrificing a large number of records the improvement in prediction is insubstantial and a large apparent bias (for Mississippi) remains unexplained. This 'robustness' feature of the prediction model is very desirable because model selection has to rely to a large extent on expediency (availability of data) rather than familiarity with the subject matter, underlying theory, or formal statistical procedures.

7. CONCLUSION

A prediction method for estimating within-state means of literacy proficiency scores from national surveys, based on a random-effects regression model, is presented and illustrated on two states, Mississippi and Oregon. The method combines the advantages of the design-based and model-based approaches by incorporating sampling weights, by imposing a model which captures the clustered structure of the data, and by using linear regression to reduce the residual variation at both cluster and elementary levels. For the two states statewide survey data are available, and so the predictions from national surveys can be compared with more reliable estimates from the state-wide surveys. The estimated standard errors of the prediction can be compared with the standard errors of the (weighted) sample means to assess the usefulness of the prediction.

For instance, the predictions for Mississippi, based on the ES/UI and JTPA data, have about 2.5 and 3.5 times greater standard errors than the sample mean. This can be interpreted that information about Mississippi in the national surveys is equivalent to that in a sample about ($2.5^2 =$) six times (ES/UI) and twelve times (JTPA) smaller than the sample collected in Mississippi. The corresponding factors for Oregon are smaller, about four for ES/UI and eleven for JTPA. These factors can be adjusted for unequal (effective) sample sizes in the obvious manner. Although these conclusions are contingent on the selected regression model, the prediction appears to be fairly robust with respect to model specification.

The only iterative component of the prediction procedure is the fitting of the random coefficient model. Using the 'weighted' version of the Fisher scoring algorithm (Jennrich and Schluchter, 1986; Longford, 1987) no problems with convergence or multiple local maxima arise; usually less than twelve iterations are required to achieve convergence using any reasonable criterion for convergence.

The method can be extended to multiple layers of clustering (see Longford, 1987) and to non-normally distributed data by application of generalized linear models with random coefficients (Longford 1994). In the latter case it is assumed that the regression estimator $\hat{\beta}$ and the estimator $\bar{\mathbf{x}}$ of the state's mean are (approximately) normally distributed. Owing to the asymptotic theory, this is a realistic assumption.

The predictions based on the two national surveys can be combined. The coefficients of the convex combination of the predictions can be determined so as to minimize the standard error of the combined estimator. The coefficients of this convex combination are inversely proportional to the variances of the components. Thus, the combined estimator for Mississippi has coefficients 0.67 (ES/UI) and 0.33 (JTPA), and the resulting prediction for the state is 0.121 (standard error 0.054). Of course, prediction from one or several national surveys can also be combined with an estimate based on the data from the state. In the case of Mississippi this would lead to only a marginal improvement because the estimate based on the Mississippi survey is far superior to the prediction from either national survey, or their combination. The combined prediction for Oregon is equal to 0.590 (standard error 0.035), very close to the weighted sample mean of 0.571.

ACKNOWLEDGEMENTS

Norma Norris and Kate Pashley assisted me with orientation in the databases. Discussions with Irwin Kirsch, Peter Pashley, and Kentaro Yamamoto are acknowledged. The last revision of the manuscript benefited from the referees' comments and suggestions.

APPENDIX: SAMPLING VARIANCE OF THE PREDICTOR $\hat{y}^{(j)}$

Suppose $\hat{\mathbf{x}}^{(j)} \sim \mathcal{N}(\mathbf{x}^{(j)}, \Sigma_S)$, $\hat{\beta} \sim \mathcal{N}(\beta, \Sigma_b)$ and $\mathrm{cov}(\hat{\mathbf{x}}^{(j)}, \hat{\beta}) = \Sigma_{bS}$. By conditioning on $\hat{\beta}$ we obtain

$$\begin{aligned} \mathrm{var}\left(\hat{y}^{(j)}\right) &= \mathbf{E}_\beta\left\{\mathrm{var}\left(\hat{\mathbf{x}}^{(j)}\hat{\beta} \mid \hat{\beta}\right)\right\} + \mathrm{var}_\beta\left\{\mathbf{E}\left(\hat{\mathbf{x}}^{(j)}\hat{\beta} \mid \hat{\beta}\right)\right\} \\ &= \mathbf{E}_\beta\left\{\hat{\beta}^\top\left(\Sigma_S - \Sigma_{bS}^\top\Sigma_b^{-1}\Sigma_{bS}\right)\hat{\beta}\right\} \\ &\quad + \mathrm{var}_\beta\left[\left\{\mathbf{x}^{(j)} + \left(\hat{\beta} - \beta\right)^\top \Sigma_b^{-1}\Sigma_{bS}\right\}\hat{\beta}\right]. \end{aligned}$$

For an arbitrary $p \times p$ matrix of constants $\mathbf{A}$ we have the following identities:

$$\begin{aligned} \mathbf{E}\left(\hat{\beta}^\top\mathbf{A}\hat{\beta}\right) &= \beta^\top\mathbf{A}\beta + \mathrm{tr}\left(\mathbf{A}\Sigma_b\right) \\ \mathrm{cov}\left(\hat{\beta}, \hat{\beta}^\top\mathbf{A}\hat{\beta}\right) &= \Sigma_b(\mathbf{A} + \mathbf{A}^\top)\beta \\ \mathrm{var}\left(\hat{\beta}^\top\mathbf{A}\hat{\beta}\right) &= \mathrm{tr}\left(\mathbf{A}\Sigma_b\mathbf{A}\Sigma_b\right) + \mathrm{tr}\left(\mathbf{A}\Sigma_b\mathbf{A}^\top\Sigma_b\right) \\ &\quad + 2\beta^\top\mathbf{A}\Sigma_b\mathbf{A}\beta + 2\beta^\top\mathbf{A}\Sigma_b\mathbf{A}^\top\beta. \end{aligned}$$

The versions of these identities for a symmetric matrix $\mathbf{A}$ are well-known, see, e.g., Seber (1977, Ch. 2). Their general versions are derived by application of the 'symmetric' versions to $\mathbf{A} + \mathbf{A}^\top$, noting that $\hat{\beta}^\top\mathbf{A}\hat{\beta} = \hat{\beta}^\top\mathbf{A}^\top\hat{\beta}$. Substitution of these identities for appropriate matrices **A** yields

$$\begin{aligned}
\operatorname{var}\left(\hat{y}^{(j)}\right) = {} & \beta^\top \Sigma_S \beta + \operatorname{tr}(\Sigma_h \Sigma_S) - \beta^\top \Sigma_{hS}^\top \Sigma_h^{-1} \Sigma_{hS} \beta - \operatorname{tr}\left(\Sigma_{hS}^\top \Sigma_h^{-1} \Sigma_{hS} \Sigma_h\right) \\
& + \left(x^{(j)} - \beta^\top \Sigma_h^{-1} \Sigma_{hS}\right) \Sigma_h \left(x^{(j)} - \beta^\top \Sigma_h^{-1} \Sigma_{hS}\right)^\top \\
& + 2\left(x^{(j)} - \beta^\top \Sigma_h^{-1} \Sigma_{hS}\right)\left(\Sigma_{hS} + \Sigma_h \Sigma_{hS}^\top \Sigma_h^{-1}\right)\beta \\
& + \operatorname{tr}\left(\Sigma_{hS}^2\right) + \operatorname{tr}\left(\Sigma_h^{-1} \Sigma_{hS} \Sigma_h \Sigma_{hS}^\top\right) + 2\beta^\top \Sigma_h^{-1} \Sigma_{hS}^2 \beta \\
& + 2\beta^\top \Sigma_h^{-1} \Sigma_{hS} \Sigma_h \Sigma_{hS}^\top \Sigma_h^{-1} \beta ,
\end{aligned}$$

from which equation (6) follows directly.

REFERENCES

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

[2] **Beran, R.,** and **Hall, P.** (1992). «Estimating coefficient distributions in random coefficient regressions». *Annals of Statistics* **20**, 1979–1984.

[3] **Bock, R. D.** and **Aitkin, M.** (1981). «Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm». *Psychometrika* **46**, 443–459.

[4] **Harville, D. A.** (1974). «Bayesian inference for variance components using only error contrasts». *Biometrika* **61**, 383–385.

[5] **Jennrich, R. I.** and **Schluchter, M. D.** (1986). «Unbalanced repeated-measures models with structured covariance matrices». *Biometrics* **42**, 805–820.

[6] **Kackar, R. N.** and **Harville, D. A.** (1984). «Approximations for standard errors of estimators of fixed and random effects in mixed linear models». *Journal of the American Statistical Association* **79**, 853–862.

[7] **Kirsch, I. S.** and **Jungeblut, A.** (1992). *Profiling the literacy proficiencies of JTPA and ES/UI populations.* Final Report to the Department of Labor. Educational Testing Service, Princeton, NJ.

[8] **Longford, N. T.** (1987). «A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects». *Biometrika* **74**, 817–827.

[9] **Longford, N. T.** (1994). «Logistic regression with random coefficients». *Computational Statistics and Data Analysis* **17**, 1–15.

[10] **Longford, N. T.** (1995). «Random coefficient models and missing data». Submitted.

[11] **Mislevy, R. J.** and **Bock, R. D.** (1983). *BILOG: Item analysis and test scoring with binary logistic models.* Scientific Software, Inc., Mooresville, IN.

[12] **Potthoff, R. F., Woodbury, M. A.** and **Manton, K. G.** (1992). «Equivalent sample size" and "equivalent degrees of freedom" refinements for inference using survey weights under superpopulation models». *Journal of the American Statistical Association* **87**, 383–396.

[13] **Seber, G. A. F.** (1977). *Linear Regression Analysis.* Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York.