SUMMARY. Direct survey estimators for small areas are often unstable due to the small (or nonexistent) samples taken from these areas. Estimators with less variability can be derived by "borrowing strength" from related areas. In this study, a hierarchical Bayes methodology for estimating small area proportions is proposed. The idea consists of incorporating random effects which reflect the structure of the sample design into a logistic regression model. A data example involving the estimation of local labour force participation rates is presented. Comparisons are drawn with an empirical Bayes approach used by Farrell, MacGibbon, and Tomberlin (1997a), and an analogous procedure which incorporates a modification to the Laird and Louis (1987) Type III bootstrap suggested by Carlin and Gelfand (1991).

1. Introduction

The terms “small area” and “local area” are used to denote a small geographic area, such as a county or a census division. Unfortunately, direct survey estimators for small area parameters such as means or proportions are unstable due to the small or nonexistent samples taken from these areas. Alternatively, model-based estimators with less variability can be derived by “borrowing strength” from related areas. The first of these was the synthetic estimator (See Gonzales 1973); however it has a tendency to be biased. Additional model-based procedures have been developed which address the above deficiencies, including those based on empirical and hierarchical Bayes approaches. Ghosh and Rao (1994) showed that these Bayesian techniques, for most purposes, seem to have a distinct advantage over other methods. Farrell, MacGibbon and Tomberlin (1997a) reached a similar conclusion in a study comparing an empirical Bayes estimator with an unbiased direct survey estimator and a synthetic estimator. However, they noted that when a naive empirical Bayes approach is used, interval estimates do not attain the desired level of coverage since the uncertainty associated with estimating the parameters of the prior distribution is not accounted for. They used the bootstrap techniques proposed by Laird and Louis (1987) to adjust these naive estimates.
The study of Fay and Herriot (1979) was one of the first to apply linear empirical Bayes models to small area estimation. Datta and Ghosh (1991) considered hierarchical Bayes estimators instead. Recently, a number of extensions to the Fay-Herriot model have been made; see Fay (1987) and Ghosh, Datta and Fay (1991). Lahiri and Rao (1995) robustified the Fay-Herriot model by relaxing the assumption of a normal prior distribution. Datta and Lahiri (1995) extended this work. They developed a robust hierarchical Bayes approach for handling outliers, and studied a number of alternative prior distributions to the normal.

Others have studied the estimation of small area rates and binomial parameters using empirical and hierarchical Bayes approaches. Dempster and Tomberlin (1980) proposed an empirical Bayes method for estimating census undercount for local areas which was based on a logistic regression model containing fixed and random effects. This proposal was further developed by MacGibbon and Tomberlin (1989) and Farrell, MacGibbon and Tomberlin (1994, 1997a). Farrell, MacGibbon and Tomberlin (1994) obtained empirical Bayes point estimates of small area proportions, and also provided guidelines for the choice of the prior distribution which depended upon whether outlying local areas were present in the data. Farrell, MacGibbon and Tomberlin (1997a) extended this approach to interval estimation using a normal prior. Stroud (1994) presented a comprehensive treatment of the hierarchical-conjugate Bayesian predictive approach to binary survey data. Malec, Sedransk and Tompkins (1993) and Malec, Sedransk, Moriarity and Le Clere (1997) used fully Bayes approaches to estimate proportions using data from the National Health Interview Survey.

In this study, a hierarchical Bayes methodology for estimating small area proportions is proposed. The idea consists of incorporating into a logistic regression model containing predictor variables, random effects which reflect the structure of the sample design. The model is similar to that proposed by Dempster and Tomberlin (1980) and used by Farrell, MacGibbon and Tomberlin (1997a) to derive empirical Bayes local area estimates. However, the griddy-Gibbs sampler developed by Ritter and Tanner (1992) will be used here to estimate model parameters. The approach differs from Stroud (1994) in that the Bernoulli parameters themselves, rather than their prior means, are subject to the logistic regression model. The proposed approach is applied to data from a United States Census to determine point and interval estimates for local labour force participation rates. Comparisons are made with the empirical Bayes approach used in Farrell, MacGibbon and Tomberlin (1997a), and an analogous procedure which incorporates a modification to the Laird and Louis (1987) parametric Type III bootstrap suggested by Carlin and Gelfand (1991). The proposed estimation procedures are described in Section 2, a data example is presented in Section 3, while the conclusions and discussion are given in Section 4.

2. Estimation Procedures

Let \( p_i \) represent the proportion of individuals in the \( i \)-th local area which possess
some characteristic of interest. The objective of this study is the development of point and interval estimates for these $p_i$. The data to be employed in determining these estimates will be obtained using a two stage sample design, where individuals are to be sampled from selected local areas. The local areas will therefore reflect primary sampling units here.

In the framework of the sample design, $p_i$ can be written as

$$ p_i = \sum_j y_{ij} / N_i, \quad (2.1) $$

where $N_i$ is the population size of local area $i$, and $y_{ij}$ takes on a value of zero or one, depending upon whether or not the $j$-th individual within the $i$-th local area possesses the characteristic of interest. A predictive model-based approach proposed by Royall (1970) will be used to specify an estimator for $p_i$. Under this approach, $p_i$ in (2.1) is estimated using:

$$ \hat{p}_i = \left( \sum_{j \in S} y_{ij} + \sum_{j \in S'} \hat{y}_{ij} \right) / N_i, \quad (2.2) $$

where the sum over $j \in S$ of $y_{ij}$ is the sum of the values of the outcome variable for sampled individuals from the $i$-th local area, and the sum over $j \in S'$ of $\hat{y}_{ij}$ is the sum of the estimated outcome variables for nonsampled individuals in the $i$-th local area. Values for $\hat{y}_{ij}$ are obtained by initially specifying a model to describe the probability, $\pi_{ij}$, that the $j$-th individual within the $i$-th local area possesses the characteristic of interest. Specifically, the model is given by:

$$ y_{ij} | \pi_{ij} \sim i.i.d. \text{ Bernoulli (} \pi_{ij} \text{), logit (} \pi_{ij} \text{) = } X_{ij}^T \beta + \delta_i, \quad (2.3) $$

so that

$$ \pi_{ij} = \left[ 1 + \exp \left\{ -(X_{ij}^T \beta + \delta_i) \right\} \right]^{-1}. \quad (2.4) $$

Here $X_{ij}$ represents a vector of fixed effects predictor variables which is augmented by the constant one, and $\beta$ refers to a vector of fixed effect logistic regression parameters which contains the constant term $\beta_0$. The vector of predictor variables may include covariates at both the individual and aggregate levels. The quantity $\delta_i$ is a random effect associated with the $i$-th local area. The $\delta_i$ are assumed to follow some prior probability distribution.

Therefore, once estimates $\beta$ and $\delta_i$ have been determined, $\pi_{ij}$ is estimated by

$$ \hat{\pi}_{ij} = \left[ 1 + \exp \left\{ -(X_{ij}^T \hat{\beta} + \hat{\delta}_i) \right\} \right]^{-1}. \quad (2.5) $$

The estimates $\hat{\pi}_{ij}$ used in conjunction with (2.2) ultimately allow for the development of point and interval estimates for $p_i$. The approaches employed in achieving this objective are described in Sections 2.1 and 2.2.

Note that in this study, a joint multivariate normal prior distribution is assumed for the random effects in (2.3). Specifically, these effects are assumed to follow normal distributions, each with a mean of zero, and the same unknown
bayesian inference for small areas

405

variance $\tau^2$. For the application considered in Section 3, a check of this assumption using the approach suggested by Farrell, MacGibbon and Tomberlin (1994) indicated no evidence against normality. Thus, under the two stage sample design described above, the model in (2.3) becomes:

$$y_{ij} | \pi_{ij} \sim i.i.d. \text{Bernoulli } (\pi_{ij}), \quad \text{logit } (\pi_{ij}) = X_{ij}^T \beta + \delta_i, \quad (2.6)$$

$$\delta_i \sim i.i.d. \text{Normal } (0, \tau^2).$$

2.1 Hierarchical Bayes Local Area Estimates. If a hierarchical Bayes approach is employed to estimate small area proportions, it is necessary to specify a distribution for the parameter $\tau^2$ of the prior distribution in (2.6):

$$y_{ij} | \pi_{ij} \sim i.i.d. \text{Bernoulli } (\pi_{ij}), \quad \text{logit } (\pi_{ij}) = X_{ij}^T \beta + \delta_i, \quad \delta_i \sim i.i.d. \text{Normal } (0, \tau^2), \quad \tau^2 \sim \text{Inverse Gamma } (a, b). \quad (2.7)$$

Here, a diffuse version of an inverse gamma distribution will be assumed for $\tau^2$, where the parameters of the distribution, $a$ and $b$, are both set to zero.

To develop hierarchical Bayes estimates for the model in (2.7), we begin by considering the distribution of the data. If $\delta$ and $y$ are vectors containing $\delta_i$ and the data $y_{ij}$, then the data are distributed as the following product binomial:

$$f(y | \beta, \delta) \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}}. \quad (2.8)$$

If a flat prior is placed on the fixed effects parameters, then the distribution of the parameters is:

$$f(\beta, \delta | \tau^2) \propto \tau^{-n} \exp \left( - \sum_i \delta_i^2 / 2\tau^2 \right), \quad (2.9)$$

where $n$ is the number of sampled local areas. The distribution associated with $\tau^2$ is given by:

$$f(\tau^2) = \frac{b^a \exp(-b/\tau^2)}{\tau^{a-1} \Gamma(a)}. \quad (2.10)$$

Thus, the above three distributions can be used to specify that

$$f(y, \beta, \delta, \tau^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}} \tau^{-n} \exp \left( - \sum_i \delta_i^2 / 2\tau^2 \right) \frac{b^a \exp(-b/\tau^2)}{\tau^{2(a+1)} \Gamma(a)}. \quad (2.11)$$

The joint distribution in (2.11) can be employed to determine conditional posterior distributions for the components of $\beta$ and $\delta$, as well as for $\tau^2$:

$$f(\beta_0 | y, \beta_1, \ldots, \beta_m, \delta, \tau^2) = f(y, \beta, \delta, \tau^2) / \int f(y, \beta, \delta, \tau^2) d\beta_0, \quad (2.12)$$
\[f(\beta_u|y, \beta_0, \beta_1, \ldots, \beta_{u-1}, \beta_{u+1}, \ldots, \beta_m, \delta, \tau^2) = f(y, \beta, \delta, \tau^2)/\int f(y, \beta, \delta, \tau^2)d\beta_u,\]  
(2.13)

\[f(\delta_i|y, \beta, \delta_1, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_n, \tau^2) = f(y, \beta, \delta, \tau^2)/\int f(y, \beta, \delta, \tau^2)d\delta_i,\]  
(2.14)

\[f(\tau^2|y, \beta, \delta) = f(y, \beta, \delta, \tau^2)/\int f(y, \beta, \delta, \tau^2)d\tau^2,\]  
(2.15)

where \(u\) refers to the \(u\)-th predictor variable, and \(m\) is the number of predictor variables.

It is not feasible to obtain closed form expressions for the posteriors given in (2.12) through (2.15) due to the intractable integrations required to evaluate the denominators. However, these equations can be used to define the following series of expressions that the posteriors will be proportional to

\[f(\beta_0|y, \beta_1, \ldots, \beta_m, \delta, \tau^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}},\]  
(2.16)

\[f(\beta_u|y, \beta_0, \beta_1, \ldots, \beta_{u-1}, \beta_{u+1}, \ldots, \beta_m, \delta, \tau^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}},\]  
(2.17)

\[f(\delta_i|y, \beta, \delta_1, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_n, \tau^2) \propto \prod_{ij} \pi_{ij}^{y_{ij}}(1 - \pi_{ij})^{1-y_{ij}} \exp(-\delta_i^2/2\tau^2),\]  
(2.18)

\[f(\tau^2|y, \beta, \delta) \propto \frac{1}{\tau_n^a} \exp\left(-\sum_i \delta_i^2/2\tau^2\right) \frac{\exp(-b/\tau^2)}{\tau^{2(a+1)}} = \frac{1}{\tau_n^{a+2}} \exp\left(-\sum_i \delta_i^2/2\tau^2\right),\]  
(2.19)

provided that a diffuse version of the inverse gamma distribution is placed upon \(\tau^2\), where \(a\) and \(b\) are both set to zero.

To determine estimates for \(\beta, \delta\) and \(\tau^2\), the griddy-Gibbs sampler proposed by Ritter and Tanner (1992) is applied to (2.16) through (2.19). Initially, a set of values are assumed as the estimates for \(\beta, \delta\) and \(\tau^2\) say \(\hat{\beta}_0(0), \hat{\delta}(0)\) and \(\hat{\tau}^2(0)\). Equation (2.16) is then used to develop a discrete empirical cumulative distribution function for \(\hat{\beta}(0)\) by evaluating the expression proportional to \(f(\hat{\beta}_0|y, \hat{\beta}_1(0), \ldots, \hat{\beta}_m(0), \hat{\delta}(0), \hat{\tau}^2(0))\) at a number of different values for \(\hat{\beta}_0\). An updated value for the estimate of \(\hat{\beta}_0, \hat{\beta}_0(1)\), is obtained by sampling from this distribution. The algorithm proceeds by using (2.17) to develop a discrete empirical cumulative distribution function for \(\hat{\beta}_1\) by evaluating the expression proportional to \(f(\hat{\beta}_1|y, \hat{\beta}_0(1), \ldots, \hat{\beta}_m(0), \hat{\delta}(0), \hat{\tau}^2(0))\) at a number of different values for \(\hat{\beta}_1\). The resulting distribution is sampled to provide an updated value of the estimate of \(\hat{\beta}_1, \hat{\beta}_1(1)\). This procedure of creating discrete cumulative distribution functions using the most up-to-date revised estimates and subsequently sampling them continues until the revised estimates \(\hat{\beta}(1)\) and \(\hat{\delta}(1)\) have been obtained. These quantities are used to create a discrete empirical cumulative distribution function for \(\hat{\tau}^2\) by using (2.19) to evaluate the expression proportional to \(f(\hat{\tau}^2|y, \hat{\beta}(1), \hat{\delta}(1))\) at different values for \(\hat{\tau}^2\). This distribution is
subsequently sampled to produce the revised estimate $\hat{\tau}_2^2$. The determination of the values $\hat{\beta}(1), \hat{\delta}(1)$, and $\hat{\tau}_2^2$ constitutes the first iteration in the griddy-Gibbs sampler. The second iteration proceeds precisely as the first, except that the starting values $\hat{\beta}(0), \hat{\delta}(0)$ and $\hat{\tau}_2^2(0)$ are replaced by the revised estimates $\hat{\beta}(1), \hat{\delta}(1)$ and $\hat{\tau}_2^2(1)$. At the end of the second iteration, the values $\hat{\beta}(2), \hat{\delta}(2)$ and $\hat{\tau}_2^2(2)$ would be available as the starting values for the third iteration. This procedure is continued for a large number of iterations, say $T$, where the estimates arising from this final iteration would be denoted by $\hat{\beta}(T), \hat{\delta}(T)$ and $\hat{\tau}_2^2(T)$.

This series of $T$ iterations is considered as one path of the griddy-Gibbs sampler. The algorithm also requires the specification of convergence criteria. Cowles and Carlin (1996) summarize and compare a number of the available Markov chain Monte Carlo convergence diagnostics. In this paper, a graphical technique described by Ritter and Tanner (1992) and a numerical approach proposed by Gelman and Rubin (1992) are both used to assess convergence of the griddy-Gibbs sampler. For both methods, it is necessary to consider estimates for small area proportions at the same iteration number across many paths. Therefore, the series of $T$ iterations described above must be repeated a large number of times, say $R$, to produce $R$ different estimates of the same parameter at each of the $T$ iterations. Different values $\beta(0), \delta(0)$ and $\tau^2(0)$ are used as the initial estimates over the $R$ different paths to allow for an investigation of multimodality in the joint posterior distribution of the model parameters.

To check for convergence of the griddy-Gibbs sampler using the approaches suggested by Ritter and Tanner (1992) and Gelman and Rubin (1992), a value for $\hat{p}_i$ is required at each of the $T$ iterations for each of the $R$ different paths. For any given path, the values $\hat{\beta}(t)$ and $\hat{\delta}(t)$ are available at the $t$-th iteration, where $t = 1, ..., T$. These values can be substituted for $\hat{\beta}$ and $\hat{\delta}$ in (2.5) to determine values for $\hat{\pi}_{ij(t)}$ for all $j \in S'$ in the $t$-th local area. A value for $\hat{\gamma}_{ij(t)}$, where $j \in S'$, is then generated from a Bernoulli distribution with parameter $\hat{\pi}_{ij(t)}$. The resulting values for $\hat{\gamma}_{ij(t)}$ are then substituted for $\hat{\gamma}_{ij}$ in (2.2) to obtain a value for $\hat{p}_{ij(t)}$. Since $R$ different paths of the griddy-Gibbs sampler have been performed, there will be $R$ different values for $\hat{p}_{ij(t)}$ at each of the $T$ iterations comprising the paths.

Under the approach suggested by Ritter and Tanner (1992), the $R$ values for $\hat{p}_{ij(t)}$ at the $t$-th iteration are considered as a distribution for $\hat{p}_i$. The quartiles of the distribution for the estimator are then determined at each iteration, and plotted as a function of the iteration number. Convergence of the estimates for $p_i$ is achieved at the iteration where all of these quartiles have stabilized. Gelman and Rubin (1992) propose using the $R$ values for $\hat{p}_{ij(t)}$ at each of the final $T/2$ iterations to check for convergence. The method requires that the mean and the variance of the $R$ values for $\hat{p}_{ij(t)}$ are determined for each of these iterations. Gelman and Rubin (1992) recommend a comparison of the variance of the averages of the $\hat{p}_{ij(t)}$ values over these iterations to the average within-path variance of the $\hat{p}_{ij(t)}$ values over the $R$ different paths. They formalize this comparison in terms of a statistic referred to as a “Shrink Factor”, which should be close to one for convergence to be achieved.

Once the iteration number for which the griddy-Gibbs sampler is deemed to have
converged has been deduced, $R$ values of the estimator $\hat{p}_i$ in (2.2) are available. If a point estimate of the proportion of the $i$-th local area is desired, the mean or the median of these $R$ values could be used. In addition, if a $100(1 - \alpha)\%$ interval estimate is required, the $R$ values for $\hat{p}_i$ can be treated as an empirical distribution. One possible interval estimate is obtained by taking the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of this distribution as the lower and upper limits, respectively.

2.2 Empirical Bayes Local Area Estimates. Empirical Bayes estimates can be derived for the parameters in (2.6) by using the EM algorithm described by Dempster, Laird, and Rubin (1977) to determine a maximum likelihood estimate for $\tau^2$ (See Farrell, MacGibbon and Tomberlin 1997a). Once the empirical Bayes estimates for the model parameters have been obtained, (2.5) is used to determine a value for $\hat{\pi}_{ij}$ for all $j \in S'$ in the $i$-th local area, then (2.2) is used to obtain empirical Bayes point estimates of small area proportions by setting $\sum \hat{y}_{ij} = \sum \hat{\pi}_{ij}$. To develop empirical Bayes interval estimates, we consider the mean square error of $\hat{p}_i$ as a predictor for $p_i$. When $\sum \hat{y}_{ij}$ in (2.2) is replaced by $\sum \hat{\pi}_{ij}$, this mean square error can be estimated as

$$MSE(\hat{p}_i) = Var\left(\frac{\sum_{j \in S'} \hat{\pi}_{ij}}{N_i}\right) + \frac{\sum_{j \in S'} \hat{\pi}_{ij}(1 - \hat{\pi}_{ij})}{N_i^2}$$

(2.20)

For sampled local areas, where $n_i$ is greater than zero, the first term of (2.20) is of order $1/n_i$, while the second term is of order $1/N_i$. In this study, the approximation of the mean square error of $\hat{p}_i$ is based on the first term only, yielding a useful approximation so long as $N_i$ is large compared to $n_i$. For nonsampled local areas, the first term in (2.20) is of order 1; therefore it always dominates the second term.

To develop an expression for the variance of $\hat{p}_i$, we let $Z_{ij}$ represent a vector of fixed effects predictor variables for the $ij$-th item augmented by a series of binary variables, each indicating whether or not the $ij$-th item belongs to a particular local area. We also let $\hat{\Gamma}$ be the vector containing the estimates of the fixed and random effects parameters. Then $Z_{ij}^T \hat{\Gamma} = X_{ij}^T \hat{\beta} + \hat{\delta}_i$ where $\hat{\beta}$ and $\hat{\delta}_i$ are the empirical Bayes estimates of $\beta$ and $\delta_i$, respectively. To obtain an expression for the variance of $\hat{p}_i$, a first order multivariate Taylor series expansion of (2.2) with $\sum \hat{y}_{ij}$ replaced by $\sum \hat{\pi}_{ij}$ is taken with respect to the realized values of the fixed and random effects estimates, yielding an approximate expression which describes $\hat{p}_i$ as a linear function of these estimates. If the variance of this expression is taken, the result is

$$Var(\hat{p}_i) = \left[ \sum_{j \in S'} Z_{ij}^T \hat{\pi}_{ij} (1 - \hat{\pi}_{ij}) \right] \left( \frac{\hat{\Sigma}}{N_i^2} \right) \left[ \sum_{j \in S'} Z_{ij} \hat{\pi}_{ij} (1 - \hat{\pi}_{ij}) \right],$$

(2.21)

where $\hat{\Sigma}$ represents the covariance matrix of the estimated vector $\hat{\Gamma}$.

Estimates of posterior variances given by (2.21) do not include the uncertainty due to estimating the prior parameters; hence empirical Bayes confidence intervals based on these variances are often too short to achieve the desired level of coverage. A number of methods for addressing this shortcoming are available (See Carlin and Gelfand 1990, 1991; and Laird and Louis 1987). Farrell, MacGibbon, and Tomberlin
(1997a) used the parametric Type III bootstrap proposed by Laird and Louis (1987) and showed empirically that it is capable of incorporating the uncertainty that arises from having to estimate the parameters of the prior distribution of the random effects.

The methodology of Laird and Louis (1987) requires the generation of a number of bootstrap samples, $N_B$, from each sample selected from the population under consideration. The procedure for generating a single bootstrap sample, say the $b$-th, is described as follows:

1. For a given set of sample data, obtain empirical Bayes estimates of the fixed and random effects, $\hat{\beta}$ and $\hat{\delta}_i$, for the model in (2.6), along with an estimate of the prior distribution of the random effects.

2. For each sampled local area, generate a random effect $\hat{\delta}_bi$ using the estimated prior distribution of the random effects obtained in step (1).

3. For each sample observation, a probability $\hat{\pi}_{bij}$ is computed by replacing $\beta$ and $\delta_i$ in (2.4) with $\hat{\beta}$ obtained in step (1) and $\hat{\delta}_bi$ generated in step (2).

4. For each sample observation, generate $y_{bij}$ from a Bernoulli distribution with parameter $\hat{\pi}_{bij}$.

5. The values obtained for $y_{bij}$, along with the vectors $Z_{ij}$ constitute the data for a bootstrap sample.

For the $b$-th bootstrap sample, the empirical Bayes estimation approach in Farrell, MacGibbon and Tomberlin (1997a) is applied to the model in (2.6) to obtain an estimate for $\tau^2$. Applying the procedures again to this estimate, the original data $y$, and the vectors $Z_{ij}$ for sampled individuals yields estimates of the fixed and random effects, $\hat{\beta}_i^*$ and $\hat{\delta}_bi^*$, along with an associated covariance matrix, $\hat{\Sigma}_b^*$. These estimates are used to arrive at an estimate, $\hat{p}_i^*$, for the proportion of local area $i$ by replacing $\hat{\pi}_{ij}$ in (2.2) with $\hat{\pi}_{bij}^*$, where $\hat{\pi}_{bij}^*$ is determined by substituting $\hat{\beta}_i^*$ and $\hat{\delta}_bi^*$ into (2.5) for $\beta$ and $\delta_i$, respectively. Replacing $\hat{\pi}_{ij}$ and $\Sigma$ in (2.21) by $\hat{\pi}_{bij}$ and $\hat{\Sigma}_b^*$, respectively, provides an estimate of the variability of $\hat{p}_i^*$.

The quantities $\hat{p}_i^*$ and $\text{Var}(\hat{p}_i^*)$ are determined for each of $N_B$ bootstrap samples, and used to calculate

$$\hat{p}_i^* = \frac{\sum_b \hat{p}_bi^*}{N_B},$$

and the bootstrap-adjusted estimate of variability associated with the empirical Bayes estimator, $\hat{p}_i$:

$$\text{Var}^*(\hat{p}_i) = \frac{\sum_b \text{Var}(\hat{p}_bi^*)}{N_B} + \frac{\sum_b (\hat{p}_bi^* - \hat{p}_i)^2}{N_B - 1}.$$  

A bootstrap-adjusted 100(1-$\alpha$)% confidence interval for $p_i$ can be determined using $\hat{p}_i \pm z_{(1-\alpha/2)} \sqrt{\text{Var}^*(\hat{p}_i)}$, where $z_{(1-\alpha/2)}$ is the 100(1-$\alpha$/2) percentile of a standard normal distribution.

Carlin and Gelfand (1991) have proposed a modification to the Type III bootstrap which corrects for bias by conditioning on the data. For the Type III bootstrap proposed by Laird and Louis (1987), a single set of bootstrap samples is obtained by
randomly generating values for the outcome variable for each sampled individual. Under the modification suggested by Carlin and Gelfand (1991), a set of bootstrap samples is generated for each local area in the sample. A single bootstrap sample for a local area would contain the actual outcome variable values for the individuals sampled in that particular local area, and randomly generated outcome variable values for sampled individuals from the other areas. The procedure for generating a single bootstrap sample for the $i$-th local area based on the modification of Carlin and Gelfand (1991), say the $b$-th sample, is as follows:

1. For a given set of sample data, obtain empirical Bayes estimates of the fixed and random effects, $\hat{\beta}$ and $\hat{\delta}_i$, for the model in (2.6), along with an estimate of the prior distribution of the random effects.

2. For the $k$-th sampled local area, where $k \neq i$, generate a random effect $\delta^*_b$ using the estimated prior distribution of the random effects obtained in step (1).

3. For each observation from the $k$-th sampled local area, where $k \neq i$, a probability $\pi^*_{bkj}$ is computed by replacing $\beta$ and $\delta_i$ in (2.4) with $\hat{\beta}$ obtained in step (1) and $\delta^*_b$ generated in step (2).

4. For each observation from the $k$-th sampled local area, where $k \neq i$, generate $y^*_{bkj}$ from a Bernoulli distribution with parameter $\pi^*_{bkj}$.

5. The $y_{ij}$ associated with the $i$-th local area, the generated $y^*_{bkj}$ values, where $k \neq i$, as well as the $Z_{qj}$ for each sampled local area constitute the data for a bootstrap sample.

Suppose that $N_B$ bootstrap samples are drawn in this fashion, with the $b$-th sample producing the conditional estimates $\hat{p}^*_C_{bi}$ and $\hat{Var}(\hat{p}^*_C_{bi})$. These estimates are determined by applying an approach similar to the one employed with bootstrap samples generated using the Laird and Louis (1987) bootstrap. Once obtained these estimates can then be used to determine

$$\hat{p}^*_C_i = \frac{\sum_b \hat{p}^*_C_{bi}}{N_B}, \quad (2.24)$$

and a modified bootstrap-adjusted estimate of the variability in $\hat{p}_i$:

$$\hat{Var}^*_C(\hat{p}_i) = \frac{\sum_b \hat{Var}(\hat{p}^*_C_{bi})}{N_B} + \frac{\sum_b (\hat{p}^*_C_{bi} - \hat{p}^*_C_i)^2}{N_B - 1}. \quad (2.25)$$

A modified bootstrap-adjusted 100$(1 - \alpha)$% confidence interval for $p_i$ can be determined using the formula $\hat{p}_i \pm z_{(1 - \alpha/2)} \sqrt{\hat{Var}^*_C(\hat{p}_i)}$, where $z_{(1 - \alpha/2)}$ is the 100$(1 - \alpha/2)$ percentile of a standard normal distribution.

3. Data Example

To compare the hierarchical Bayes technique proposed here with the empirical Bayes approach proposed by Farrell, MacGibbon and Tomberlin (1997a) based on the Laird and Louis (1987) Type III bootstrap and a modification to this procedure
which uses the Carlin and Gelfand (1991) bootstrap, a simulation study was carried out using data from a 1% sample of the 1950 United States Census (United States Bureau of the Census 1984). The local area parameter of interest, $p_i$, was the female labour force participation rate, where local areas are more or less confined to states. The choice of the above data set was motivated by the fact that it comprises a public use microdata sample. Unfortunately, data from a more recent census are not yet available in this form. Bethlehem, Keller and Pannekoek (1990) elaborate on the difficulties encountered in obtaining microdata.

Note that when microdata concerning predictor variables for all individuals in a local area are not available, $\hat{p}_i$ in (2.2) cannot be determined. In order to determine values for the $\hat{\pi}_{ij}$ for nonsampled individuals in the expression for $\hat{p}_i$, the $X_{ij}$ predictor variable vectors are required for these individuals. However, Farrell, MacGibbon and Tomberlin (1997b) proposed a model-based procedure for estimating binomial small area parameters which required only local area summary statistics for both continuous and categorical auxiliary variables. Although the procedure was illustrated using estimates for model parameters obtained using an empirical Bayes approach, it is also applicable when estimates are determined using the hierarchical Bayes technique proposed here.

Farrell, MacGibbon and Tomberlin (1997b) have also described conditions under which an estimator based on summary statistics, $\tilde{p}_i$, will serve as a useful approximation to the estimator based on microdata for auxiliary variables. Using the empirical Bayes approach referred to in Section 2.2 to estimate model parameters, they showed that for the application considered in this study, the approximation is useful. Nevertheless, since microdata is available here, the results presented below are obtained by using predictor variable data for each individual included within a local area.

The model used for empirical Bayes estimation in this study is based on (2.6), while the model given in (2.7) is employed for hierarchical Bayes estimation. The fixed effects predictor variables for these models, selected by a stepwise logistic regression procedure, were age, marital status and whether the individual had children. In addition to these individual level predictor variables, the models also contained local area variables representing average age, the proportions of individuals in various marital status categories, and the proportion of individuals having children. These local area level covariates are necessary since their exclusion results in an observed relationship between the expected value of $\hat{p}_i$ and its bias, where as the expected value increases, the bias increases from large negative to large positive values. This correlation is removed through the inclusion of domain level covariates.

Farrell, MacGibbon and Tomberlin (1997a) present a detailed study for estimating small area proportions which compares logistic regression models with and without domain level covariates.

Once the predictor variables were selected, the data for estimating local area female labour force participation rates were obtained from the 1% sample using a two stage sample design, emulating an actual sample survey. First, twenty out of fifty-two local areas were selected without replacement using probabilities proportional to size (PPS). Specifically, randomized systematic selection of primary sampling
units with PPS (See Kish 1965, p.230) was used to select these local areas. Fifty individuals were then randomly selected from each chosen local area, bringing the total sample size to 1,000. Five hundred samples were drawn in this fashion. In order to investigate the properties of the estimator \( \hat{p}_i \) conditionally on a particular set of local areas being sampled, resampling was not performed at the local area selection stage.

For each of the 500 replicates, three hundred paths of the griddy-Gibbs sampler were performed. Each path consisted of a series of five hundred iterations of the sampler. These paths produced three hundred sets of parameter estimates at each iteration number. For all iteration numbers, the discrete empirical cumulative distribution functions for the estimators of the parameters were constructed using one hundred equi-spaced values for the appropriate estimator. These values were selected in such a fashion so that the probability of choosing the value with the greatest chance of being selected was approximately one hundred times greater than the probability of selecting the value with the smallest chance of being chosen.

For a particular replicate, three hundred sets of values for \( \hat{p}_i \) in (2.2) are available at the iteration at which the griddy-Gibbs sampler was deemed to have converged, one for each of the paths performed. These estimates for \( \hat{p}_i \) in (2.2) are then treated as an empirical distribution. The 50th percentile of this distribution is taken as a hierarchical Bayes point estimate of the proportion for the \( i \)-th local area. A 95% hierarchical Bayes interval estimate for this proportion is created by using the 2.5th and 97.5th percentiles of the distribution for \( \hat{p}_i \) as the lower and upper limits, respectively. Note that it is possible to produce 500 hierarchical Bayes point and interval estimates for the proportion of the \( i \)-th local area, one for each replicate.

An empirical Bayes point estimate was also obtained for \( p_i \) for each of the 500 replicates using the approach described in Section 2.2. In addition, two bootstrap-adjusted 95% confidence intervals for \( p_i \) were also determined for each replicate. One was based on the Laird and Louis (1987) bootstrap, the other on the modification proposed by Carlin and Gelfand (1991). Thus, 500 empirical Bayes point estimates, bootstrap-adjusted interval estimates, and modified bootstrap-adjusted interval estimates were obtained for the proportion of the \( i \)-th local area.

Table 1 presents some of the results obtained for sampled local areas using the empirical and hierarchical Bayes estimation approaches. The table presents the true small area proportions, and the averages of the empirical and hierarchical Bayes point estimates over the 500 simulation samples. Regardless of whether an empirical or hierarchical Bayes estimation approach is employed, the design bias of \( \hat{p}_i \) is quite small for most local areas. In order to compare the empirical and hierarchical Bayes estimation procedures, the mean absolute difference between the small area proportions and the average estimated rates was determined for each approach. The hierarchical Bayes procedure produced a mean absolute difference of 0.0031, as compared to 0.0056 for the empirical Bayes approach. In addition, the absolute differences between the small area proportions and the average estimated rates obtained using the hierarchical Bayes approach were smaller than the empirical Bayes counterparts for sixteen of the twenty local areas.
Table 1: Average estimates, average interval lengths and coverage rates for sampled local areas

<table>
<thead>
<tr>
<th>Local Area</th>
<th>Ave. $\hat{p}_i$</th>
<th>Ave $\hat{p}_i$:</th>
<th>Average Interval Length</th>
<th>Coverage Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EB</td>
<td>HB</td>
<td>EB:LL</td>
<td>EB:CG</td>
</tr>
<tr>
<td>01</td>
<td>0.2191</td>
<td>0.2143</td>
<td>0.1543</td>
<td>0.1528</td>
</tr>
<tr>
<td>02</td>
<td>0.2343</td>
<td>0.2467</td>
<td>0.1691</td>
<td>0.1685</td>
</tr>
<tr>
<td>03</td>
<td>0.2393</td>
<td>0.2315</td>
<td>0.1630</td>
<td>0.1653</td>
</tr>
<tr>
<td>04</td>
<td>0.2446</td>
<td>0.2523</td>
<td>0.1695</td>
<td>0.1677</td>
</tr>
<tr>
<td>05</td>
<td>0.2455</td>
<td>0.2429</td>
<td>0.1673</td>
<td>0.1689</td>
</tr>
<tr>
<td>06</td>
<td>0.2456</td>
<td>0.2441</td>
<td>0.1666</td>
<td>0.1662</td>
</tr>
<tr>
<td>07</td>
<td>0.2568</td>
<td>0.2663</td>
<td>0.1784</td>
<td>0.1761</td>
</tr>
<tr>
<td>08</td>
<td>0.2587</td>
<td>0.2656</td>
<td>0.1743</td>
<td>0.1757</td>
</tr>
<tr>
<td>09</td>
<td>0.2690</td>
<td>0.2783</td>
<td>0.1794</td>
<td>0.1808</td>
</tr>
<tr>
<td>10</td>
<td>0.2697</td>
<td>0.2723</td>
<td>0.2685</td>
<td>0.1771</td>
</tr>
<tr>
<td>11</td>
<td>0.2741</td>
<td>0.2729</td>
<td>0.2802</td>
<td>0.1794</td>
</tr>
<tr>
<td>12</td>
<td>0.2858</td>
<td>0.2878</td>
<td>0.2854</td>
<td>0.1845</td>
</tr>
<tr>
<td>13</td>
<td>0.2944</td>
<td>0.2887</td>
<td>0.2998</td>
<td>0.1837</td>
</tr>
<tr>
<td>14</td>
<td>0.2990</td>
<td>0.2961</td>
<td>0.2973</td>
<td>0.1884</td>
</tr>
<tr>
<td>15</td>
<td>0.3087</td>
<td>0.2987</td>
<td>0.3132</td>
<td>0.1896</td>
</tr>
<tr>
<td>16</td>
<td>0.3128</td>
<td>0.3150</td>
<td>0.3076</td>
<td>0.1946</td>
</tr>
<tr>
<td>17</td>
<td>0.3197</td>
<td>0.3162</td>
<td>0.3172</td>
<td>0.1931</td>
</tr>
<tr>
<td>18</td>
<td>0.3461</td>
<td>0.3572</td>
<td>0.3526</td>
<td>0.2113</td>
</tr>
<tr>
<td>19</td>
<td>0.3467</td>
<td>0.3462</td>
<td>0.3481</td>
<td>0.2085</td>
</tr>
<tr>
<td>20</td>
<td>0.3607</td>
<td>0.3520</td>
<td>0.3585</td>
<td>0.2090</td>
</tr>
</tbody>
</table>

The table also presents average interval lengths and coverage rates over the 500 simulation samples which are based on 95% confidence intervals determined using the empirical Bayes approach described in Farrell, MacGibbon and Tomberlin (1997a) which employs the Type III bootstrap of Laird and Louis (1987), the modification to this procedure which uses the Carlin and Gelfand (1991) bootstrap, and the hierarchical Bayes approach proposed here. The coverage rates over all three methods range from 92.2% to 97.8%. Note that an approximate bound for the Monte Carlo error of the simulation study is $3\sqrt{(0.95)(0.05)/500}$, or 0.029. Thus, all coverage rates are within 3 standard errors of 95%.

There is little difference in the average interval lengths obtained using the various procedures. In addition, the average coverage rates over the twenty sampled local areas are similar for the three approaches, and very close to the 95% nominal rate. The average coverage rates for the Laird and Louis (1987) and the Carlin and Gelfand (1991) bootstrap intervals are 94.88% and 94.91%, respectively, while the average coverage rate for the hierarchical Bayes intervals is 95.14%. However, the individual local area coverage rates for the Carlin and Gelfand (1991) bootstrap intervals are consistently closer to the nominal rate than those of the Laird and Louis (1987) bootstrap intervals. The standard deviation of the coverage rates for the Laird and Louis (1987) bootstrap intervals is 1.66%, compared to 1.14% for the Carlin and Gelfand (1991) bootstrap intervals. Similar findings were observed in Carlin and Gelfand (1990). The individual local area coverage rates for the hierarchical Bayes intervals are even closer to the nominal rate than those of the Carlin and Gelfand (1991) bootstrap intervals. The standard deviation of the coverage rates for the hierarchical Bayes intervals is only 0.54%.
Thirty-two of the fifty-two local areas were not sampled. Although the findings were similar in most aspects to those when sampled local areas were considered, the performance of the empirical and hierarchical Bayes procedures deteriorated slightly since nonsampled local areas constitute, in effect, a holdout sample. A key difference in the results between sampled and nonsampled local areas is that for the latter group, the Carlin and Gelfand (1991) modified bootstrap reduces to the Laird and Louis (1987) bootstrap. Therefore, the individual nonsampled local area coverage rates are identical for the two bootstraps.

4. Conclusion and Discussion

In the context of the simulation study conducted here, both the empirical and hierarchical Bayes procedures yielded point estimates with a small design bias, with the hierarchical Bayes approach being slightly better. The average coverage rates for Laird and Louis (1987) bootstrap intervals, Carlin and Gelfand (1991) bootstrap intervals, and hierarchical Bayes intervals are similar and very close to the nominal rate. However, the variability in the individual local area coverage rates based on the hierarchical Bayes intervals is much smaller than that in counterparts based on the two bootstraps. This is clearly an advantage of the hierarchical Bayes approach. In addition, the variability in the individual local area coverage rates for intervals based on the Carlin and Gelfand (1991) bootstrap is smaller than that for analogous rates based on intervals obtained using the Laird and Louis (1987) bootstrap, but only for sampled local areas. A disadvantage of the Carlin and Gelfand (1991) bootstrap is its computer intensity, relative to that of Laird and Louis (1987). The two methods differ in computing time by a factor equal to the number of sampled areas.

The hierarchical Bayes approach proposed here has also been extended to point and interval estimation of small area proportions based on multinomial and ordinal outcome variables, and compared with the empirical Bayes method used by Farrell (1997). The comparison yielded conclusions analogous to those obtained in this study. The proposed approach also allows for the specification of different forms for the prior distribution of the random effects, implying that it can be used according to the guidelines in Farrell, MacGibbon and Tomberlin (1994) when outlying local areas are present in the data. Using the same simulated data sets as in Farrell, MacGibbon and Tomberlin (1994), a comparison of the approach with the method used by Farrell, MacGibbon, and Tomberlin (1994) yields results similar to those of this study.

Acknowledgements. The author would like to acknowledge the financial support of NSERC of Canada. The author is grateful to the editor, co-editor and referee for useful comments and suggestions.
References


Patrick J. Farrell  
School of Mathematics and Statistics  
Carleton University  
1125 Colonel By Drive  
Ottawa, Ontario, CANADA K1S 5B6  
E-mail: pfarrell@math.carleton.ca