Survey Methodology, December 1997
Vol. 23, No. 2, pp. 119-126
Statistics Canada

Empirical Bayes Estimation of Small Area Proportions Based on Ordinal Outcome Variables

PATRICK J. FARRELL[1]

ABSTRACT

Much research has been conducted into the modelling of ordinal responses. Some authors argue that, when the response variable is ordinal, inclusion of ordinality in the model to be estimated should improve model performance. Under the condition of ordinality, Campbell and Donner (1989) compared the asymptotic classification error rate of the multinominal logistic model to that of the ordinal logistic model of Anderson (1984). They showed that the ordinal logistic model had a lower expected asymptotic error rate than the multinominal logistic model. This paper also aims to compare the performance of ordinal and multinomial logistic models for ordinal responses. However, rather than focussing on classification efficiency, the assessment is made in the context of an application where the objective is to estimate small area proportions. More specifically, using multinominal and ordinal logistic models, the empirical Bayes approach proposed by Farrell, MacGibbon and Tomberlin (1997a) for estimating small area proportions based on binomial outcome data is extended to response variables consisting of more than two outcome categories. The properties of estimators based on these two models are compared via a simulation study in which the empirical Bayes methods proposed here are applied to data from the 1950 United States Census with the objective of predicting, for a small area, the proportion of individuals who belong to the various categories of an ordinal response variable representing income level.

KEY WORDS: Bootstrap; Complex survey design; Logistic regression; Random effects models; Small area summary statistics; Taylor series.

1. INTRODUCTION

Much research has been conducted into the modelling of ordinal responses (see Albert and Chib 1993, Anderson 1984, Crouchley 1995, and McCullagh 1980). Some authors argue that, when the response variable is ordinal, inclusion of ordinality in the model to be estimated should improve model performance. Under the condition of ordinality, Campbell and Donner (1989) theoretically compared the asymptotic classification error rate of the multinomial logistic model to that of the ordinal logistic model of Anderson (1984), demonstrating that the ordinal model had a lower expected asymptotic error rate. However, in a subsequent simulation study, Campbell, Donner, and Webster (1991) illustrated that ordinal models classify less accurately than multinomial models under a variety of circumstances, and concluded that ordinal models confer no advantage when the main purpose of an analysis is classification.

This paper also aims to compare the performance of ordinal and multinomial logistic models for ordinal responses. However, rather than focussing on classification efficiency, the assessment is made in the context of an application where the objective is to estimate small area proportions.

The estimation of small area parameters is a finite population sampling problem which has received considerable attention. An excellent review of such research appears in Ghosh and Rao (1994). These authors demonstrate that as a compromise between synthetic and direct survey estimators, estimators based on empirical or hierarchical Bayes procedures are not subject to the large bias that is sometimes associated with a synthetic estimator (see Gonzales 1973), nor are they as variable as a direct survey estimator. A similar conclusion was drawn by Farrell, MacGibbon, and Tomberlin (1997a) in a study of the properties of an empirical Bayes estimator for small area proportions based on a binomial outcome variable.

Despite the numerous studies aimed at predicting small area proportions based on binomial response variables (see Dempster and Tomberlin 1980, MacGibbon and Tomberlin 1989, Farrell 1991, Farrell *et al.* 1997a, Malec, Sedransk, and Tompkins 1993, Stroud 1991, and Wong and Mason 1985), little attention has been given to estimating proportions based on response variables with more than two outcome categories. This paper extends the empirical Bayes approach of Farrell *et al.*, (1997a), to such response variables by basing the estimates on multinomial and ordinal logistic models. To compare the estimates of small area proportions based on an ordinal outcome variable using multinomial and ordinal models, the proposed empirical Bayes methods are applied to data from the 1950 United States Census in order to predict, for a given small area, the proportion of individuals who belong to the various categories of an ordinal response variable representing income level.

For such an estimation problem, there are many issues which require attention. They include the selection of predictor variables for the model, model diagnostics, the sample design, and the properties of the estimators

[1] Patrick J. Farrell, Assistant Professor, Department of Mathematics and Statistics, Acadia University, Wolfville, Nova Scotia, B0P 1X0.

employed. For example, among the model diagnostics for the multinomial and ordinal models was an assessment of model fit which was based on residuals. For a description of this diagnostic and others, see Farrell (1991). The findings did not appear to indicate a lack of fit for either model. In this study, the focus is on investigating the properties of empirical Bayes estimators over repeated realizations of the sample design using a simulation. For many survey practitioners, such properties are of prime importance.

One concern associated with using an empirical Bayes estimation approach is that interval estimates do not attain the desired level of coverage, since the uncertainty that arises from having to estimate the parameters of the prior distribution is not accounted for. This study incorporates the suggestion of Laird and Louis (1987) to use bootstrap techniques for adjusting naive estimates of accuracy. Alternatively, Prasad and Rao (1990) have developed a procedure which attempts to account for the uncertainty not captured by the naive estimates. Although their approach was designed for three specific linear models containing random effects, Cressie (1992) has made certain conjectures as to when the procedure is appropriate. Of importance is the constraint that the outcome variable must follow a normal distribution.

The proposed empirical Bayes procedures based on multinomial and ordinal logistic models are presented in Section 2. The simulation study to compare multinomial and ordinal logistic models for ordinal responses is described in Section 3, while the conclusions and discussion are presented in Section 4.

2. ESTIMATION PROCEDURES

Consider a discrete small area characteristic of interest with M possible outcomes. The subscript m will reference these categories, where $m = 1, ..., M-1$ and $m^+ = 1, ..., M$. In addition, underlined lower case and capital letters will designate vectors, while bold capital letters will represent matrices.

The estimation procedures are illustrated under a two stage sample design, where individuals are sampled from selected local areas. Thus, local areas are the primary sampling units here. Let p_{im+} be the proportion of individuals in the i-th local area that belong to category m^+ of the response variable. Then

$$p_{im^+} = \sum_j y_{ijm^+} / N_i, \tag{2.1}$$

where y_{ijm+} is either zero or one, depending upon whether the j-th individual in local area i belongs to category m^+ of the characteristic of interest, and N_i is the population size of the i-th local area.

The approach employed by Farrell *et al.*, (1997a), to estimate small area proportions based on binomial outcome variables is extended here to allow for the estimation of p_{im+}. The procedure follows the explicitly model-based approach proposed by Dempster and Tomberlin (1980). Let π_{ijm+} represent the probability that the j-th individual within the i-th local area belongs to category m^+ of the response variable. Then, according to Royall (1970), p_{im+} in (2.1) is estimated by

$$\hat{p}_{im^+} = \left(\sum_{j \in S} y_{ijm^+} + \sum_{j \in S'} \hat{\pi}_{ijm^+} \right) \Big/ N_i, \tag{2.2}$$

where S is the set of n_i sampled individuals from local area i, and S' is the set of individuals in local area i not included in the sample. Values for the $\hat{\pi}_{ijm+}$ are required. To obtain these estimates, logistic regression models are used to describe the probabilities associated with individuals in the population.

Under a multinomial logistic model, the π_{ijm+} are described as follows:

$$\begin{aligned} \log(\pi_{ijm} / \pi_{ijM}) &= \underline{X}_{ij}^T \underline{\beta}_m + \delta_{im}, \\ \underline{\delta}_i &\sim \text{i.i.d. Normal}\ (\underline{0}, \boldsymbol{D}), \end{aligned} \tag{2.3}$$

where $\underline{\delta}_i^T = (\delta_{i1}, ..., \delta_{i(M-1)})$, $i = 1, ..., I$, and $\boldsymbol{D}$ is an unknown covariance matrix. In this model, $\underline{X}_{ij}$ is a vector of fixed effects predictor variables, the vector $\underline{\beta}_m$ contains the fixed effects parameters associated with the m-th category of the outcome variable of interest, and δ_{im} is a normally distributed random effect associated with the m-th category of the characteristic of interest in the i-th local area. The vector $\underline{X}_{ij}$ may include covariates at both the individual and aggregate levels. For sample designs of more than two stages, an analogous model would contain random effects for the sampling units at each stage, excluding the final one.

Note that the model in (2.3), unlike a similar model proposed by Malec *et al.*, (1993), does not contain interaction terms between the local area effects and the fixed effects predictor variables. However, terms to acknowledge such interaction could be included if they were deemed necessary.

To obtain Bayes estimates of the model parameters, values are assumed for the unknown parameters of the random effects distribution. Let $\underline{y}_{ij}^T = (y_{ij1}, ..., y_{ijM})$ be a vector for the ij-th sampled individual where the component associated with the category of the outcome variable to which the individual belongs has a value of one. The remaining entries are zero. If $\boldsymbol{Y}$ is a matrix with rows $\underline{y}_{ij}^T$, then the data are distributed as:

$$f(\boldsymbol{Y} \mid \underline{\beta}, \underline{\delta}_c) \propto \prod_{ij} \pi_{ij1}^{y_{ij1}} \pi_{ij2}^{y_{ij2}} \dots \pi_{ijM}^{y_{ijM}},$$

where $\underline{\beta}^T = (\underline{\beta}_1^T, ..., \underline{\beta}_{M-1}^T)$, and $\underline{\delta}_c^T = (\underline{\delta}_1^T, ..., \underline{\delta}_I^T)$. If a flat distribution is specified for the fixed effects, the distribution of the parameters is $f(\underline{\beta}, \underline{\delta}_c | \boldsymbol{D}_c) \propto \exp(-\tfrac{1}{2}\underline{\delta}_c^T \boldsymbol{D}_c \underline{\delta}_c)$, where $\boldsymbol{D}_c = \text{diag}(\boldsymbol{D}, \boldsymbol{D}, ..., \boldsymbol{D})$. The joint distribution of the data and the parameters is determined using $f(\boldsymbol{Y} \mid \underline{\beta}, \underline{\delta}_c)$ and $f(\underline{\beta}, \underline{\delta}_c \mid \boldsymbol{D}_c)$, and subsequently employed to obtain the posterior distribution of the parameters. Unfortunately, a closed form for this posterior distribution cannot be derived due to the intractable integration required to obtain the marginal distribution of $\boldsymbol{Y}$. A possible approach could be a stochastic integration method such as Gibbs sampling (see Zeger and Karim 1991). Ripley and Kirkland (1990) indicate that the drawbacks of such an approach include the intensive computations and questions about when the sampling process has achieved equilibrium. Since computing time is of particular concern for the simulation discussed in Section 3, this approach will not be pursued here. Alternatively, Breslow and Clayton (1993) state that there is still room for simple, approximate methods. Many authors have found that a multivariate normal approximation of the posterior works very well in practice (see Farrell *et al.* 1997a, Laird 1978, Tomberlin 1988, and Wong and Mason 1985). Breslow and Lin (1995) warn, however, that such an approach might yield inconsistent estimates for the fixed effects parameters. Thus, if $\hat{p}_{im+}$ is to be based on fixed effects estimates obtained in this manner, the same might apply to the consistency of $\hat{p}_{im+}$ as an estimator for p_{im+}.

Following Farrell *et al.* (1997a), the posterior distribution of the parameters is approximated as a multivariate normal distribution having its mean at the mode and covariance matrix equal to the inverse of the information matrix evaluated at the mode. The information matrix here is simply the second derivative of the posterior distribution taken with respect to $\underline{\beta}$ and $\underline{\delta}_c$. When values are specified for the unknown parameters of the random effects distribution, the resulting mode and covariance matrix constitute an initial set of estimates of the model parameters. Empirical Bayes estimates are then obtained by using the EM algorithm described by Dempster, Laird, and Rubin (1977) to determine estimates for the parameters of the random effects distribution. The algorithm converges quickly, taking only a few minutes in real time. For details on how the empirical Bayes estimates are obtained for a model based on a two stage sample design and a binomial response variable, see MacGibbon and Tomberlin (1989).

The empirical Bayes estimates of the model parameters are used in (2.2) to determine $\hat{p}_{im+}$. In developing an expression for the uncertainty of $\hat{p}_{im+}$, N_i is assumed to be known. Since the approach being used is model-based and predictive in nature, the uncertainty in $\hat{p}_{im+}$ arises solely from the $\sum \hat{\pi}_{ijm+}$ term; the $\sum y_{ijm+}$ term has zero variance. Thus, the mean square error of $\hat{p}_{im+}$ as a predictor for p_{im+} can be estimated as

$$\widehat{\mathrm{MSE}}(\hat{p}_{im^+}) = \widehat{\mathrm{Var}}\left(\frac{\sum\limits_{j\in S'} \hat{\pi}_{ijm^+}}{N_i}\right) + \frac{\sum\limits_{j\in S'} \hat{\pi}_{ijm^+}(1-\hat{\pi}_{ijm^+})}{N_i^2}. \qquad (2.4)$$

For sampled local areas, where n_i is greater than zero, the first term of (2.4) 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 $\hat{p}_{im+}$ is based on the first term only, which yields a useful approximation provided that N_i is large compared to n_i. For nonsampled local areas, the first term in (2.4) is of order 1; therefore it always dominates the second term.

To estimate the uncertainty of $\hat{p}_{im+}$, which is expressed as a non-linear function of the estimators of the fixed and random effects, the expression for $\hat{p}_{im+}$ is linearized by taking a first order multivariate Taylor series expansion about the realized values of the fixed and random effects. The variance of the resulting expression, call it $\widehat{\mathrm{Var}}(\hat{p}_{im+})$, is taken as an estimate of the uncertainty of $\hat{p}_{im+}$. Details of the Taylor series expansion are given in Farrell *et al.*, (1997a), for a binomial outcome variable.

When population micro-data for auxiliary variables are not available, $\hat{p}_{im+}$ in (2.2) cannot be determined. For non-linear models such as (2.3), prediction is not straightforward in this situation. However, an alternative estimator to $\hat{p}_{im+}$, say $\tilde{p}_{im+}$, which requires only local area summary statistics (a mean vector and finite population covariance matrix) for both continuous and categorical variables can be obtained by extending the approach proposed by Farrell, MacGibbon, and Tomberlin (1997b) for achieving this objective when estimating binomial small area parameters. The same Taylor series expansion that was used to estimate the accuracy of $\hat{p}_{im+}$ can be employed to obtain a measure of the uncertainty for $\tilde{p}_{im+}$, $\widehat{\mathrm{Var}}(\tilde{p}_{im+})$.

The approach described in this section can also be used to develop point and interval estimates for small area proportions based on $\hat{p}_{im+}$ and $\tilde{p}_{im+}$ when an ordinal model is used. In this study, a fixed and random effects model is proposed for the π_{ijm+} which is based on the ordinal model proposed by McCullagh (1980)

$$\log\left(\frac{\pi_{ij1} + \dots + \pi_{ijm}}{\pi_{ij(m+1)} + \dots + \pi_{ijm}}\right) = \beta_{0m} - \underline{X}_{ij}^T \underline{\beta} + \delta_{im}, \qquad (2.5)$$

$$\underline{\delta}_i \sim \text{i.i.d. Normal } (\underline{0}, \boldsymbol{D}).$$

The vector $\underline{X}_{ij}$ contains the values of the fixed effects predictor variables for the ij-th individual, while $\underline{\beta}$ represents a vector of fixed effects parameters. Associated with the m-th category of the response variable is a constant term, β_{0m}. The random effects are again assumed to be normally distributed. Note that an important feature of the model in (2.5) is that the restriction $\beta_{0(m+1)} - \beta_{0m} \geq \delta_{im} - \delta_{i(m+1)}$ must hold in order for $\pi_{ij(m+1)} \geq 0$. A discussion concerning this constraint is given in Section 3.

The approach used to approximate the uncertainty in $\hat{p}_{im+}$ and $\tilde{p}_{im+}$ when π_{ijm+} is based on either (2.3) or (2.5) can be described as naive, since $\widehat{\mathrm{Var}}(\hat{p}_{im+})$ and $\widehat{\mathrm{Var}}(\tilde{p}_{im+})$ do not account for the uncertainty which results from estimating the parameters of the random effects distribution. Thus, interval estimates for p_{im+} that are based on $\widehat{\mathrm{Var}}(\hat{p}_{im+})$ and $\widehat{\mathrm{Var}}(\tilde{p}_{im+})$ are typically too short. Many approaches have been proposed for addressing this issue (see Carlin and Gelfand 1990, and Laird and Louis 1987). In this study, the Type III bootstrap proposed by Laird and Louis (1987) is used to adjust naively-estimated measures of uncertainty. The procedure is described in Farrell *et al.*, (1997a), for a binomial outcome variable. It can be extended to (2.3) and (2.5), and is applicable regardless of whether estimation is based on $\hat{p}_{im+}$ or $\tilde{p}_{im+}$.

The procedure requires that a number of bootstrap samples, N_B, be generated from a given set of data. Suppose that small area estimation is to be based on $\hat{p}_{im+}$. For the b-th bootstrap sample, an estimate $\hat{p}_{bim+}$ for p_{im+} based on (2.3) or (2.5), along with a naive estimate of the variability of $\hat{p}_{bim+}$, $\widehat{\text{Var}}(\hat{p}_{bim+})$ are obtained. The quantities $\hat{p}_{bim+}$ and $\widehat{\text{Var}}(\hat{p}_{bim+})$ are determined for each of N_B bootstrap samples, and used to calculate a bootstrap-adjusted estimate of the variability associated with $\hat{p}_{im+}$:

$$\widehat{\text{Var}}^{\{B\}}(\hat{p}_{im+}) = \frac{\sum_b \widehat{\text{Var}}(\hat{p}_{bim+})}{N_B} + \frac{\sum_b (\hat{p}_{bim+} - \hat{p}^{\{B\}}_{im+})^2}{N_B - 1},$$

$$\text{where} \quad \hat{p}^{\{B\}}_{im+} = \frac{\sum_b \hat{p}_{bim+}}{N_B}.$$

Note that even though individuals are not selected by simple random sampling without replacement in this study, survey weights have not been attached to the records. However, in practice, the weights attached to a record will vary due to features of the survey design, such as differential nonresponse and clustering. In this study, the models account for the effects of these features. Further research is necessary to determine what impact the incorporation of survey weights into the models would have on the bootstrapping procedure.

3. A DATA EXAMPLE

A comparison of the estimates for small area proportions based on multinomial and ordinal logistic models was carried out using a simulation study where the response variable was ordinal. The data set is based on a 1% sample of the 1950 United States Census (United States Bureau of the Census 1984). Data based on the 1950 Census is used since it constitutes a public use microdata sample, and none of the more recent census data is available in this form. Thus, the results below for the multinomial and ordinal models are obtained by using predictor variable data for each individual within a local area. For a discussion of the difficulties encountered in obtaining microdata, see Bethlehem, Keller, and Pannekoek (1990).

The application considered is the estimation of the proportion of individuals in a given local area associated with each of the three categories of an ordinal outcome variable representing total personal income, where a local area is typically specified to be a state. This variable encompasses all sources of income, including wages and salaries, business income, and net income from other sources. An individual is regarded as having a low (less than \$2,500), medium (\$2,500 to under \$10,000) or high (\$10,000 and over) level of total personal income during 1949. Thus, $m = 1$ for low income (Category 1), $m = 2$ for medium income (Category 2), and $m = 3$ for high income (Category 3). The multinomial and ordinal models were each used to obtain point and interval estimates for 42 local areas. Twenty of these areas were sampled, the others were not. Note that individuals with no income were included in Category 1. An alternative approach would have been a two stage model; a first stage logistic model for the probability of non-zero income, and a second stage multinomial or ordinal model for income category conditional on non-zero income.

In practice, historical data are often available for survey planning purposes. For example, variable selection for purposes of model predictions could be based on previous census data. To emulate this situation, a random sample of size 2,000 was selected from the 1% sample. Variables for model prediction were determined by applying a stepwise logistic regression procedure. The variables selected were age, gender, and race. With regards to race, individuals were categorized as white, negro, or other.

Thus, the multinomial and ordinal models used in this study included four individual level predictor variables for age, gender, and race (two indicator variables were required to code the various races). However, they also contained four local area variables representing average age, the proportion of males, the proportion of whites, and the proportion of negroes. Regardless of which model is considered, these local area variables are necessary since, when they are excluded, a relationship is noted between the expected value of $\hat{p}_{im+}$ and its bias, where as the expected value increases, the bias increases from large negative to large positive values. The inclusion of domain level covariates removes this correlation. Therefore, since local area variables are also included in the models, the multinomial model contains eighteen fixed effects parameters (two for each of the individual level and local area predictor variables, and two constant terms) and forty random effects (two for each of the twenty sampled local areas), while the ordinal model contains ten fixed effects parameters (one for each of the individual level and local area predictor variables, and two constant terms) and forty random effects (two for each of the twenty sampled local areas). For a detailed study comparing logistic regression models for estimating small area proportions with and without domain level covariates which uses binomial outcome data, see Farrell *et al.*, (1997a).

The data for estimating the proportions of individuals in each local area belonging to the various income level categories were obtained from the 1% sample using a self-weighting two stage sample design. In the first stage, 20 out of 42 local areas were selected, without replacement, using probabilities proportional to size (PPS). More specifically, the approach used to select these local areas was randomized systematic selection of primary sampling units with PPS (see Kish 1965, p. 230). Then, at the second stage, 50 individuals were randomly selected from each chosen local area. A total of 500 samples were drawn using this two stage design; however, resampling was not performed at the local area selection stage. Thus, the same 20 local areas were sampled in each of the 500 replicates. For these 20 sampled local areas, the average local area proportions for Categories 1, 2, and 3 of income level are 0.7142, 0.2260, and 0.0598.

Note that for the ordinal model, the constraint $\beta_{02} - \beta_{01} \geq \delta_{i1} - \delta_{i2}$ must hold in order for $\pi_{ij2} \geq 0$. A check of this constraint for each of the 500 samples using the estimates for the constant terms and the random effects indicated that it held at all times. In fact, it was discovered that in each of the 500 samples taken, the difference in the estimates for the constant terms was always positive, at least two orders of magnitude larger than the majority of the absolute differences of the random effects estimates, and always one order of magnitude bigger. Thus, the constant terms in the model dominate over the random effects.

To compare the properties of estimators for small area proportions over repeated realizations of the sample design, for each of the 500 samples selected the quantities $\hat{p}_{im+}$, $\widehat{\mathrm{Var}}(\hat{p}_{im+})$, and $\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})$ associated with each income level category were obtained for each local area, sampled or not, using both the multinomial and ordinal models. For each model, the estimates for $\widehat{\mathrm{Var}}(\hat{p}_{im+})$ and $\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})$ were used to construct naive and bootstrap-adjusted empirical Bayes symmetric 95% confidence intervals, respectively. Estimates for $\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})$ were obtained by using the bootstrap procedure to generate 100 bootstrap samples from each of the 500 simulation samples.

Note that for the ordinal model, the constraint $\beta_{02} - \beta_{01} \geq \delta_{i1} - \delta_{i2}$ must also hold in the bootstrap procedure for random effects generated from an estimated distribution; otherwise negative estimates for some of the probabilities π_{ijm+} will result when creating bootstrap samples. Over the course of the simulation for the application considered here, no negative probabilities were encountered when bootstrapping. One approach for assessing the likelihood of negative probabilities during the bootstrap procedure is to consider the ratio of the difference $\hat{\beta}_{02} - \hat{\beta}_{01}$ to the estimated prior standard deviation of the difference $\hat{\delta}_{i1} - \hat{\delta}_{i2}$. This ratio was determined for each sampled local area in each of the 500 simulation samples taken. The average of this entire set of ratios was 6.8, and none were found to be less than 5.8. Thus, the difference $\hat{\beta}_{02} - \hat{\beta}_{01}$ was determined to always be at least 5.8 times the estimated standard deviation of the difference $\hat{\delta}_{i1} - \hat{\delta}_{i2}$. Based on the empirical rule, a rule of thumb would be to conclude that when the ratio described above is at least three, it is highly unlikely that negative probabilities will arise when bootstrapping.

Table 1 presents average summary statistics over the 500 simulation samples obtained for the multinomial and ordinal models across all sampled local areas for each of three income level categories. A study of the stability of these statistics was conducted by investigating how they changed as additional samples were taken. Only slight changes were observed once 150 samples had been reached. Table 1 includes the summary statistics obtained for the first 200 samples in brackets for comparative purposes.

For each income category, two summary statistics shown in Table 1 were evaluated to compare the design bias of $\hat{p}_{im+}$ for the multinomial and ordinal models; the average bias of $\hat{p}_{im+}$, and the average absolute bias of $\hat{p}_{im+}$. The average bias is simply the mean over all sampled local areas of the differences obtained when the actual proportion, p_{im+}, for the i-th local area is subtracted from the average point estimate for the area over the 500 simulation samples. The average absolute bias is defined similarly, except that the absolute value of each difference is used. Generally speaking, the results obtained for these two summary statistics were slightly better for the ordinal model, regardless of the income category considered. However, the multinomial model did result in a somewhat smaller average bias for $\hat{p}_{im+}$ for the low income category.

For each sampled local area, empirical root mean square errors (RMSE's) were computed over the 500 simulation samples under each model for the three income categories. For each model and income level combination, the appropriate empirical RMSE's were averaged over all sampled local areas, resulting in the average empirical RMSE's presented in Table 1. Once again, the performance of the ordinal model is slightly better for all three income level categories.

To study the reduction in empirical RMSE when a model-based approach to estimation is used instead of a classical design unbiased method, average empirical RMSE's analogous to those in Table 1 based on the 500 samples were computed using the observed local area sample proportions in place of $\hat{p}_{im+}$. The average empirical RMSE's obtained were substantially larger (0.0617, 0.0564, and 0.0311 for the low, medium, and high income level categories) than those based on $\hat{p}_{im+}$ under either model.

Table 1 also includes summary statistics over all sampled local areas which relate naive and bootstrap measures of variability in $\hat{p}_{im+}$ to average empirical RMSE. For each income level category, the average relative bias and the average absolute relative bias of the square root of $\widehat{\mathrm{Var}}(\hat{p}_{im+})$ as an estimate of empirical RMSE are shown in Table 1 for the multinomial and ordinal models. The average relative bias is simply the mean over all sampled local areas of the values obtained when the difference resulting from the subtraction of the empirical RMSE for the i-th local area from the average of the square root of $\widehat{\mathrm{Var}}(\hat{p}_{im+})$ for the area over the 500 simulation samples is divided by the empirical RMSE. The average absolute bias is defined similarly, expect that the absolute value of each difference is used. The table also presents similar averages for the bootstrap-adjusted measures of variability, $\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})$. For both the multinomial and ordinal logistic models, the average relative bias and average absolute relative bias of the bootstrap-adjusted estimates of variability are substantially smaller in magnitude than their naive counterparts for all three income level categories. In

Table 1
Average Summary Statistics based on 500 Simulation Samples for the Multinomial and Ordinal Logistic Models across all Sampled Local Areas for each Income Level Category.
The average summary statistics obtained over the first 200 simulation samples are included in brackets for comparative purposes

Average	Low Income Level		Medium Income Level		High Income Level	
	Multinomial	Ordinal	Multinomial	Ordinal	Multinomial	Ordinal
Bias of $\hat{p}_{im+}$	-0.0004 (-0.0004)	-0.0005 (-0.0006)	-0.0007 (-0.0006)	-0.0004 (-0.0003)	0.0011 (0.0010)	0.0009 (0.0009)
Absolute Bias of $\hat{p}_{im+}$	0.0076 (0.0078)	0.0051 (0.0055)	0.0089 (0.0085)	0.0048 (0.0046)	0.0108 (0.0106)	0.0074 (0.0073)
Empirical RMSE	0.0479 (0.0483)	0.0467 (0.0469)	0.0417 (0.0414)	0.0401 (0.0402)	0.0236 (0.0233)	0.0231 (0.0229)
Relative Bias of $\sqrt{\widehat{\mathrm{Var}}(\hat{p}_{im+})}$	-0.1192 (-0.1197)	-0.1125 (-0.1128)	-0.1273 (-0.1276)	-0.1180 (-0.1186)	-0.1524 (-0.1521)	-0.1376 (-0.1372)
Absolute Relative Bias of $\sqrt{\widehat{\mathrm{Var}}(\hat{p}_{im+})}$	0.1192 (0.1197)	0.1125 (0.1128)	0.1273 (0.1276)	0.1180 (0.1186)	0.1524 (0.1521)	0.1376 (0.1372)
Relative Bias of $\sqrt{\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})}$	-0.0275 (-0.0272)	-0.0173 (-0.0175)	-0.0309 (-0.0314)	-0.0204 (-0.0207)	-0.0391 (-0.0393)	-0.0273 (-0.0269)
Absolute Relative Bias of $\sqrt{\widehat{\mathrm{Var}}^{\{B\}}(\hat{p}_{im+})}$	0.0294 (0.0290)	0.0227 (0.0228)	0.0349 (0.0343)	0.0263 (0.0265)	0.0450 (0.0446)	0.0353 (0.0347)
Naive Coverage Rate	91.35 (91.325)	91.91 (91.875)	91.19 (91.225)	91.78 (91.750)	90.67 (90.650)	91.26 (91.300)
Absolute Deviation of Naive Coverage from the 95% Nominal Rate	3.65 (3.675)	3.09 (3.125)	3.81 (3.775)	3.22 (3.250)	4.33 (4.350)	3.74 (3.700)
Adjusted Coverage Rate	94.44 (94.400)	94.75 (94.775)	94.37 (94.350)	94.68 (94.650)	93.91 (93.925)	94.40 (94.375)
Absolute Deviation of Adjusted Coverage from the 95% Nominal Rate	1.58 (1.600)	1.43 (1.425)	1.71 (1.725)	1.50 (1.525)	1.91 (1.900)	1.62 (1.650)

addition, these bootstrap-adjusted average summary statistics are all very small, which indicates that the bootstrap-adjusted estimates of variability are capable of incorporating most of the uncertainty that arises from having to estimate the distribution of the random effects.

For each sampled local area, naive and bootstrap-adjusted coverage rates based on 95% interval estimates were computed over the 500 samples under each model for the three income level categories. Over all income level and model combinations, the bootstrap-adjusted coverage rates for individual local areas ranged from 92.2% to 97.6%. Since an approximate bound for the Monte Carlo error is 3 $\sqrt{(0.95)(0.05)/500}$, or 0.029, all bootstrap-adjusted coverage rates are within 3 standard errors of 95%.

For each model and income level combination, the appropriate coverage rates were averaged over all sampled local areas, resulting in the average naive and bootstrap-adjusted coverage rates in Table 1. A number of observations can be made which hold for each income level category. For both multinomial and ordinal models, the average coverage rates for the bootstrap-adjusted intervals are much closer to the 95% nominal rate than those associated with the naive intervals. However, both the average naive and bootstrap-adjusted coverage rates for the ordinal model are slightly better than counterparts for the multinomial model. This is also the case for the average absolute deviation of both the naive and bootstrap-adjusted coverage rates from the 95% nominal rate. The average absolute deviation of the naive coverage rates from the 95% nominal rate is simply the mean over all sampled local areas of the absolute values of the differences obtained when the 95% nominal rate is subtracted from the naive coverage rates for the sampled local areas over the 500 simulation samples. The average absolute deviation of the bootstrap-adjusted coverage rates from the 95% nominal rate is defined analogously.

Twenty-two local areas were not sampled. Estimates for the proportion of individuals associated with each income level category were also obtained for these areas using the multinomial and ordinal models. The findings were similar to those for sampled local areas. However, the performance of the models deteriorated somewhat, since nonsampled local areas constitute a holdout sample. For a detailed evaluation of results associated with nonsampled local areas, see Farrell *et al.* (1997a).

A comparison of the estimates for the three income level categories based on micro-data, $\hat{p}_{im+}$, with those based on local area summary statistics, $\tilde{p}_{im+}$, was also made for each model. For both models, the results obtained for $\tilde{p}_{im+}$ were gratifyingly close to those obtained using $\hat{p}_{im+}$, although those obtained for $\hat{p}_{im+}$ were slightly better. Similar findings were obtained by Farrell *et al.*, (1997b) in a detailed comparison of $\hat{p}_{im+}$ and $\tilde{p}_{im+}$ for a binomial outcome variable.

4. CONCLUSION

Using multinomial and ordinal logistic models, the empirical Bayes approach proposed by Farrell *et al.*, (1997a), for estimating small area proportions based on binomial outcome data has been extended to accommodate outcome variables with more than two categories. It was found that the performance of the approach is preserved for multicategorical outcome data.

To compare the estimates of small area proportions based on an ordinal outcome variable using multinomial and ordinal logistic models, the proposed empirical Bayes methods based on these two models were applied to data from the 1950 United States Census with the objective of predicting, for a small area, the proportion of individuals who belong to the various categories of an ordinal response variable representing income level. The estimates based on the ordinal model were only slightly better in terms of design bias, empirical RMSE, and coverage rates. In addition, an important feature of the ordinal logistic model is that the constraint $\beta_{0(m+1)} - \beta_{0m} \geq \delta_{im} - \delta_{i(m+1)}$ must hold in order for $\pi_{ij(m+1)} \geq 0$. Since the results for the multinomial and ordinal models in the simulation were very similar, a multinomial model could be used for estimating small area proportions based on ordinal outcome variables when there is concern that fitting an ordinal model may result in negative estimates for some of these probabilities.

ACKNOWLEDGEMENTS

This research was supported by NSERC of Canada. The author is grateful to the associate editor and the referees for their valuable comments and suggestions.

REFERENCES

ALBERT, J.H., and CHIB, S. (1993). Bayesian analysis of binary and polytomous response data. *Journal of the American Statistical Association*, 88, 669-679.

ANDERSON, J.A. (1984). Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B*, 46, 1-30.

BETHLEHEM, J.G., KELLER, W.J., and PANNEKOEK, J. (1990). Disclosure control of microdata. *Journal of the American Statistical Association*, 85, 38-45.

BRESLOW, N.E., and CLAYTON, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

BRESLOW, N.E., and LIN, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82, 81-91.

CAMPBELL, M.K., and DONNER, A. (1989). Classification efficiency of multinomial logistic regression relative to ordinal logistic regression. *Journal of the American Statistical Association*, 84, 587-591.

CAMPBELL, M.K., DONNER, A., and WEBSTER, K.M. (1991). Are ordinal models useful for classification? *Statistics in Medicine*, 10, 383-394.

CARLIN, B.P., and GELFAND, A.E. (1990). Approaches for empirical Bayes confidence intervals. *Journal of the American Statistical Association*, 85, 105-114.

CRESSIE, N. (1992). REML Estimation in empirical Bayes smoothing of census undercount. *Survey Methodology*, 18, 75-94.

CROUCHLEY, R. (1995). A random-effects model for ordered categorical data. *Journal of the American Statistical Association*, 90, 489-498.

DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

DEMPSTER, A.P., and TOMBERLIN, T.J. (1980). The analysis of census undercount from a postenumeration survey. *Proceedings of the Conference on Census Undercount*, Arlington, VA, 88-94.

FARRELL, P.J. (1991). Empirical Bayes Estimation of Small Area Proportions. PhD. dissertation, Department of Management Science, McGill University, Montreal, Quebec, Canada.

FARRELL, P.J., MacGIBBON, B., and TOMBERLIN, T.J. (1997a). Empirical Bayes estimators of small area proportions in multistage designs. *Statistica Sinica*, 7, 1065-1083.

FARRELL, P.J., MacGIBBON, B., and TOMBERLIN, T.J. (1997b). Empirical Bayes small area estimation using logistic regression models and summary statistics. *Journal of Business and Economic Statistics*, 15, 101-108.

GHOSH, M., and RAO, J.N.K. (1994). Small area estimation: an appraisal. *Statistical Science*, 9, 55-93.

GONZALES, M.E. (1973). Use and evaluation of synthetic estimation. *Proceedings of the Social Statistics Section, American Statistical Association*, 33-36.

KISH, L. (1965). *Survey Sampling*. New York: John Wiley & Sons Inc.

LAIRD, N.M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, 65, 581-590.

LAIRD, N.M., and LOUIS, T.A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association*, 82, 739-750.

MacGIBBON, B., and TOMBERLIN, T.J. (1989). Small area estimates of proportions via empirical Bayes techniques. *Survey Methodology*, 15, 237-252.

MALEC, D., SEDRANSK, J., and TOMPKINS, L. (1993). Bayesian predictive inference for small areas for binary variables in the National Health Interview Survey. In *Case Studies in Bayesian Statistics*, (Eds. C. Gatsonis, J.S. Hodges, R. Kasf, and N.D. Singpurwalla). New York: Springer Verlag.

McCULLAGH, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42, 109-142.

PRASAD, N.G.N., and RAO, J.N.K. (1990). On the estimation of mean square error of small area predictors. *Journal of the American Statistical Association*, 85, 163-171.

RIPLEY, B.D., and KIRKLAND, M.D. (1990). Iterative simulation methods. *Journal of Computational and Applied Mathematics*, 31, 165-172.

ROYALL, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 74, 1-12.

STROUD, T.W.F. (1991). Hierarchical Bayes predictive means and variances with application to sample survey inference. *Communications in Statistics, Theory and Methods*, 20, 13-36.

TOMBERLIN, T.J. (1988). Predicting accident frequencies for drivers classified by two factors. *Journal of the American Statistical Association*, 83, 309-321.

UNITED STATES BUREAU OF THE CENSUS (1984). Census of the Population, 1950: Public Use Microdata Sample Technical Documentation, edited by J.G. Keane, Washington, D.C.

WONG, G.Y., and MASON, W.M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.

ZEGER, S.L., and KARIM, M.R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79-86.