HIERARCHICAL MODELING FOR NONRESPONSE AND VICTIMIZATIONS IN THE NATIONAL CRIME SURVEY

Elizabeth A. Stasny, Ohio State University
Department of Statistics, Columbus, Ohio 43210

Key words: Categorical data, Empirical Bayes, Nonrandom nonresponse

1. INTRODUCTION

Data from large-scale sample surveys are often used to estimate the probability that an individual falls into a particular survey classification or has a certain characteristic. For example, data from the National Crime Survey (NCS) is used to estimate the probability of being victimized, and data from the Current Population Survey is used to estimate the probability of being unemployed. It is often of interest to obtain such estimates for sub-groups of the population, such as neighborhoods or age/sex/race groups, as well as for the entire population. If the amount of data available for a population sub-group is small, then it may be difficult to obtain an accurate estimate of the desired probability within that sub-group. A further complication is that not all sampled units respond to a survey and the probability that a sampled unit responds may be related to the survey classification of that unit. The work described here addresses this problem of estimating probabilities in population sub-groups in the presence of possibly nonrandom nonresponse.

This paper presents hierarchical models for the probabilities of the classification of interest and the probabilities of response within sub-groups of the population. Under these hierarchical models, we think of the probabilities that individuals within sub-groups have the characteristic of interest or respond to the survey as belonging to distributions of such probabilities. The advantage to using such models is that information from the entire sample may be used to estimate parameters of the distributions of probabilities and, hence, information from the entire sample is used to estimate the probabilities for a single sub-group (see, for example, Morris, 1983). Such hierarchical models have been proposed in the survey sampling context by Lehoczky and Schervish (1987), who worked with data from the NCS and modeled the distribution of the probabilities of victimization as a beta distribution.

In this paper, we extend the model for victimization probabilities proposed by Lehoczky and Schervish to allow for nonresponse. The models presented here allow the nonresponse probabilities to come from a single distribution, which corresponds to random nonresponse, or from two distributions depending on the presence or absence of the characteristic of interest, which corresponds to informative or nonrandom nonresponse.

The general hierarchical model for survey classification and nonrandom nonresponse is presented in Section 2 of this paper. The special case of the model corresponding to random nonresponse is presented in Section 3. In Section 4 the models are fit to simulated data, which were generated based on probabilities obtained from NCS data, and to actual NCS data. Conclusions and areas for future research are presented in Section 5.

2. THE GENERAL HIERARCHICAL MODEL

This section presents a general form of the hierarchical model for the probabilities of having a particular survey classification and of responding to the survey. The model is an extension of the hierarchical model for victimization probabilities proposed by Lehoczky and Schervish (1987). We shall refer to the sampled units as "individuals" although they may actually be households or some other units rather than single persons. We will refer to the population sub-groups of interest as "strata". We will assume that the sample is chosen using a stratified random sampling plan and that the goal is to estimate the probabilities that individuals within each stratum have the characteristic of interest. In practice, of course, one will typically wish to estimate probabilities in sub-groups of the population that are smaller than strata. In smaller sub-groups, it is often the case that limited data are available for estimating the desired probability within a sub-group. In such cases, the empirical-Bayes procedure, which allows us to borrow information from the entire sample in order to estimate probabilities in small sub-groups, may provide more accurate estimates within sub-groups than do standard procedures.

Since the example of Section 4 uses data from the NCS, in the development of the hierarchical model we will let the characteristic or survey classification of interest be whether or not the individual reported being victimized. The model, of course, is applicable to surveys other than the NCS.

2.1 Model for the Observed Data

Suppose that the population of interest has been divided into K strata. We assume that individuals within a single stratum have a common probability of being victimized while individuals in different strata may have different probabilities. Let p_i be the probability that an individual in the i^{th} stratum is victimized. We will model the distribution of the p_i as a beta distribution with parameters a and b. That is, we will assume that $p_i \overset{iid}{\sim} \text{Beta}(a,b)$ for $i = 1, 2, ..., K$.

Within each stratum, suppose we take a random sample of n_i individuals and observe the survey classification of each individual. Thus we observe, say,

$$X_{ij} = \begin{cases} 1 & \text{if } j^{th} \text{ individual in } i^{th} \text{ stratum is victimized} \\ 0 & \text{otherwise.} \end{cases}$$

We assume that the victimizations within a stratum are conditionally independent of each other given the probability of victimization within the stratum. Thus,

$$X_{i1}, X_{i2}, ..., X_{in_i} | p_i \overset{iid}{\sim} \text{Bernoulli}(p_i) \quad \text{for } i = 1, 2, ..., K.$$

Naturally, we only observe the X_{ij} for individuals who respond to the survey. Let us denote the response status for a sampled individual by Y_{ij}, where

$$Y_{ij} = \begin{cases} 1 & \text{if } j^{th} \text{ individual in } i^{th} \text{ stratum responds} \\ 0 & \text{otherwise.} \end{cases}$$

In the following, we denote summary counts for the observed data in stratum i as follows:

$$\begin{aligned} Y_{i+} &= \text{\# of respondents,} \\ n_i - Y_{i+} &= \text{\# of nonrespondents,} \\ Z_{i+} &= \textstyle\sum_{j=1}^{n_i} X_{ij} Y_{ij} = \text{\# of responding victims, and} \\ Y_{i+} - Z_{i+} &= \text{\# of crime-free respondents.} \end{aligned}$$

We now add the hierarchical model for the response probabilities to the hierarchical model for the victimization probabilities. We allow the probability of nonresponse to differ by stratum and by victimization status. For $i = 1, 2, ..., K$ let

π_{i1} = probability a victimized individual in stratum i is a respondent (ie. $X_{ij} = 1$) and
π_{i0} = probability a crime-free individual in stratum i is a respondent (ie. $X_{ij} = 0$).

We also model the distributions of the π_{i1} and π_{i0} probabilities as beta distributions. In particular, we assume that, given X_{ij}, the π_{i1} and π_{i0} are random samples from beta distributions with parameters α_1 and β_1, and α_0 and β_0 respectively. That is, $\pi_{iv}|X_{ij}=v \overset{iid}{\sim} \text{Beta}(\alpha_v,\beta_v)$ for v = 1, 0. We assume that the response statuses of individuals within a stratum are conditionally independent of each other given the probability of responding within the stratum and the victimization status of the individual. Thus, for i = 1, 2, ..., K

$$Y_{i1}, Y_{i2}, ..., Y_{in_i}|X_{ij} = v, \pi_{iv} \overset{iid}{\sim} \text{Bernoulli}(\pi_{iv}).$$

We take an empirical Bayes approach to estimating the parameters of the model described above. (Note that a full Bayes approach could also be taken by placing prior distributions on the a, b, α_1, β_1, α_0, and β_0 parameters. We will not, however, consider the full Bayes approach here.) From the likelihood for the observed data, we integrate over the unobservable p_i, π_{i1}, and π_{i0} parameters to obtain the marginal distribution of the data given the a, b, α_1, β_1, α_0, and β_0 parameters. Then we obtain maximum likelihood estimators (MLE's) of the a, b, α_1, β_1, α_0, and β_0 parameters from this marginal distribution. The p_i, π_{i1}, and π_{i0} parameters are then treated as a random sample from a distribution with parameters equal to the MLE's, $\hat{a}$, $\hat{b}$, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_0$, and $\hat{\beta}_0$. This distribution is used as a prior distribution for the p_i, π_{i1}, and π_{i0} parameters, and a posterior distribution given the data, $\{X_{ij}, Y_{ij}\}$, is computed. The means of the posterior distribution are used as estimates of the p_i, π_{i1}, and π_{i0} parameters.

2.2 Estimates of the Parameters of the Beta Distributions

Under the model described above, the probabilities for each possible type of observation in the survey data for individual j within stratum i are as follows:

$$\begin{aligned}P(X_{ij} = 1, Y_{ij} = 1) &= P(\text{responding and victimized})\\ &= P(X_{ij} = 1)P(Y_{ij} = 1| X_{ij} = 1)\\ &= p_i\pi_{i1}\end{aligned}$$

$$\begin{aligned}P(X_{ij} = 0, Y_{ij} = 1) &= P(\text{responding and crime free})\\ &= P(X_{ij} = 0)P(Y_{ij} = 1| X_{ij} = 0)\\ &= (1 - p_i)\pi_{i0}\end{aligned}$$

$$\begin{aligned}P(Y_{ij} = 0) &= P(\text{nonresponding})\\ &= P(X_{ij} = 1)P(Y_{ij} = 0| X_{ij} = 1)\\ &\quad + P(X_{ij} = 0)P(Y_{ij} = 0| X_{ij} = 0)\\ &= p_i(1 - \pi_{i1}) + (1 - p_i)(1 - \pi_{i0})\end{aligned}$$

Thus, the likelihood function for the observed data using the hierarchical model described above is

$$\prod_{i=1}^{K} \{ \binom{n_i}{Y_{i+}}\binom{Y_{i+}}{Z_{i+}}[p_i\pi_{i1}]^{Z_{i+}} [(1-p_i)\pi_{i0}]^{Y_{i+}-Z_{i+}} \times [p_i(1-\pi_{i1}) + (1-p_i)(1-\pi_{i0})]^{n_i-Y_i}.$$

To find the marginal distribution of the data given the a, b, α_1, β_1, α_0, and β_0 parameters we must complete the following integration:

$$\begin{aligned}&f(\{X_{ij}, Y_{ij}\}| a, b, \alpha_1, \beta_1, \alpha_0, \beta_0)\\ &= \int_0^1\int_0^1\int_0^1 \prod_{i=1}^{K} \{ \binom{n_i}{Y_{i+}}\binom{Y_{i+}}{Z_{i+}}[p_i\pi_{i1}]^{Z_{i+}} [(1-p_i)\pi_{i0}]^{Y_{i+}-Z_{i+}}\\ &\quad\times [p_i(1-\pi_{i1}) + (1-p_i)(1-\pi_{i0})]^{n_i-Y_i}\\ &\quad\times [\Gamma(a+b)/\Gamma(a)\Gamma(b)]\, p_i^{a-1} (1-p_i)^{b-1} \qquad (1)\\ &\quad\times [\Gamma(\alpha_1+\beta_1)/\Gamma(\alpha_1)\Gamma(\beta_1)]\pi_{i1}^{\alpha_1-1} (1-\pi_{i1})^{\beta_1-1}\\ &\quad\times [\Gamma(\alpha_0+\beta_0)/\Gamma(\alpha_0)\Gamma(\beta_0)]\pi_{i0}^{\alpha_0-1} (1-\pi_{i0})^{\beta_0-1}\}\, dp_i\, d\pi_{i1}\, d\pi_{i0}\end{aligned}$$

Using only the terms involving the p_i in the inner-most integral of equation (1), we can use a binomial expansion to complete that integration as follows:

$$\begin{aligned}&\int_0^1 p_i^{Z_{i+}+a-1} (1-p_i)^{Y_{i+}-Z_{i+}+b-1}[p_i(1-\pi_{i1})+(1-p_i)(1-\pi_{i0})]^{n_i-Y_i}\, dp_i\\ &= \int_0^1 p_i^{Z_{i+}+a-1} (1-p_i)^{Y_{i+}-Z_{i+}+b-1} \sum_{r=0}^{n_i-Y_{i+}} \binom{n_i-Y_{i+}}{r}\\ &\quad\times [p_i(1-\pi_{i1})]^r[(1-p_i)(1-\pi_{i0})]^{n_i-Y_{i+}-r}\, dp_i\\ &= \sum_{r=0}^{n_i-Y_{i+}} \binom{n_i-Y_{i+}}{r} (1-\pi_{i1})^r(1-\pi_{i0})^{n_i-Y_{i+}-r}\,\Gamma(Z_{i+}+a+r)\\ &\quad\times \Gamma(n_i -Z_{i+}+b-r)/\Gamma(n_i+a+b)\end{aligned}$$

where the final step is obtained by rewriting the integrand as a beta probability density function. The remaining two integrals can be solved simply by rewriting the integrands as beta probability density functions. The result of the integration, therefore, is

$$\begin{aligned}&\{\prod_{i=1}^{K} \sum_{r=0}^{n_i-Y_{i+}} \binom{n_i}{Y_{i+}}\binom{Y_{i+}}{Z_{i+}}\binom{n_i-Y_{i+}}{r}B(Z_{i+}+a+r,\ n_i -Z_{i+}+b-r)\\ &\quad\times B(Z_{i+}+\alpha_1, \beta_1+r) \times B(Y_{i+}-Z_{i+}+\alpha_0, n_i -Y_{i+}+\beta_0-r)\}\\ &\quad\times \{B(a,b)B(\alpha_1,\beta_1)B(\alpha_0,\beta_0)\}^{-1} \qquad (2)\end{aligned}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the complete beta function. Note that the summation in equation (2) is over all possible combinations of victimized and crime free for the nonrespondents in each stratum.

The expression in equation (2) must be maximized using numerical methods in order to obtain the MLE's of the a, b, α_1, β_1, α_0, and β_0 parameters. The methods used to obtain the MLE's for the examples in Section 4 will be discussed in that section.

2.3 Estimating the Probabilities of Victimization and Response

The MLE's described in the previous subsection are now used to obtain the joint posterior distribution for the p_i, π_{i1}, and π_{i0} parameters. In the i^{th} stratum, the desired posterior distribution is

$$\begin{aligned}&f(p_i, \pi_{i1}, \pi_{i0}|\{X_{ij}, Y_{ij}\})\\ &\quad= \{f(\{X_{ij}, Y_{ij}\}| p_i, \pi_{i1}, \pi_{i0})\, \xi(p_i, \pi_{i1}, \pi_{i0})\}\\ &\quad\times \{ \int_0^1\int_0^1\int_0^1 f(\{X_{ij}, Y_{ij}\}| p_i, \pi_{i1}, \pi_{i0})\\ &\quad\times \xi(p_i, \pi_{i1}, \pi_{i0})dp_i\, d\pi_{i1}\, d\pi_{i0}\}^{-1} \qquad (3)\end{aligned}$$

where $\xi(p_i, \pi_{i1}, \pi_{i0}) = \xi_V(p_i|\hat{a},\hat{b})\, \xi_R(\pi_{i1}|X_{ij} = 1, \hat{\alpha}_1, \hat{\beta}_1)\, \xi_R(\pi_{i0}|X_{ij} = 0, \hat{\alpha}_0, \hat{\beta}_0)$. Using the MLE's obtained from maximizing equation (2) and the same binomial expansion as was used in the integration of equation (1) we have that the numerator of equation (3) is

$$\begin{aligned}&f(\{X_{ij}, Y_{ij}\}| p_i, \pi_{i1}, \pi_{i0})\, \xi(p_i, \pi_{i1}, \pi_{i0})\\ &\quad= \binom{n_i}{Y_{i+}}\binom{Y_{i+}}{Z_{i+}}\{B(\hat{a},\hat{b})B(\hat{\alpha}_1,\hat{\beta}_1)B(\hat{\alpha},\hat{\beta}_0)\}^{-1}\\ &\quad\times \{ \sum_{r=0}^{n_i-Y_{i+}}\binom{n_i-Y_{i+}}{r} p_i^{Z_{i+}+\hat{a}+r-1} (1-p_i)^{n_i-Z_{i+}+\hat{b}-r-1}\\ &\quad\times \pi_{i1}^{Z_{i+}+\hat{\alpha}_1-1}(1-\pi_{i1})^{\hat{\beta}_1+r-1}\, \pi_{i0}^{Y_{i+}-Z_{i+}+\hat{\alpha}_0-1}\\ &\quad\times (1-\pi_{i0})^{n_i-Y_{i+}+\hat{\beta}_0-r-1}\}.\end{aligned}$$

The integral of the above function with respect to p_i, π_{i1}, and π_{i0}, needed in the denominator of equation (3), is found using the same methods that were used to complete the integration of equation (1). Canceling the common terms in the numerator and denominator, we find that the desired joint posterior distribution is

$$f(p_i, \pi_{i1}, \pi_{i0} | \{X_{ij}, Y_{ij}\}) = \{ \sum_{r=0}^{n_i-Y_{i+}} \binom{n_i-Y_{i+}}{r} p_i^{Z_{i+}+\hat{a}+r-1} (1-p_i)^{n_i-Z_{i+}+\hat{b}-r-1} \pi_{i1}^{Z_{i+}+\hat{\alpha}_1-1} \times (1-\pi_{i1})^{\hat{\beta}_1+r-1} \pi_{i0}^{Y_{i+}-Z_{i+}+\hat{\alpha}_0-1} (1-\pi_{i0})^{n_i-Y_{i+}+\hat{\beta}_0-r-1} \}$$
$$\times \{ \sum_{r=0}^{n_i-Y_{i+}} \binom{n_i-Y_{i+}}{r} B(Z_{i+}+\hat{a}+r, n_i-Z_{i+}+\hat{b}-r) \times B(Z_{i+}+\hat{\alpha}_1, \hat{\beta}_1+r) \times B(Y_{i+}-Z_{i+}+\hat{\alpha}_0, n_i-Y_{i+}+\hat{\beta}_0-r) \}^{-1}.$$

The expected values of the p_i, π_{i1}, and π_{i0} parameters under the posterior distribution may be used as the corresponding parameter estimates within each stratum. For any stratum i, the integration is easily completed to show that the expected value of p_i is

$$\{ \sum_{r=0}^{n_i-Y_{i+}} K_i[Z_{i+}+\hat{a}+r/n_i+\hat{a}+\hat{b}] \} / \{ \sum_{r=0}^{n_i-Y_{i+}} K_i \}$$

where $K_i = \binom{n_i-Y_{i+}}{r} B(Z_{i+}+\hat{a}+r, n_i-Z_{i+}+\hat{b}-r) \times B(Z_{i+}+\hat{\alpha}_1, \hat{\beta}_1+r) \times B(Y_{i+}-Z_{i+}+\hat{\alpha}_0, n_i-Y_{i+}+\hat{\beta}_0-r)$. Similarly, the expected value of π_{i1} may be shown to be

$$\{ \sum_{r=0}^{n_i-Y_{i+}} K_i[Z_{i+}+\hat{\alpha}_1/Z_{i+}+\hat{\alpha}_1+\hat{\beta}_1+r] \} / \{ \sum_{r=0}^{n_i-Y_{i+}} K_i \}$$

and the expected value of π_{i0} may be shown to be

$$\{ \sum_{r=0}^{n_i-Y_{i+}} K_i[Y_{i+}-Z_{i+}+\hat{\alpha}_0/n_i-Z_{i+}+\hat{\alpha}_0+\hat{\beta}_0-r] \} / \{ \sum_{r=0}^{n_i-Y_{i+}} K_i \}.$$

3. SPECIAL CASE OF RANDOM NONRESPONSE

In this section we consider the special case of the model described above in which individuals within different stratum may have different probabilities of responding but those probabilities do not depend directly on victimization status. The distributional assumptions concerning the p_i are as in Section 2. But we now assume that $\pi_{i1} = \pi_{i0} = \pi_i$ with $\pi_i \overset{iid}{\sim}$ Beta(α, β) for i = 1, 2, ..., K. The likelihood function for the observed data using this model is

$$\prod_{i=1}^{K} \{ \binom{n_i}{Y_{i+}} \binom{Y_{i+}}{Z_{i+}} [p_i\pi_i]^{Z_{i+}} [(1-p_i)\pi_i]^{Y_{i+}-Z_{i+}} [(1-\pi_i)]^{n_i-Y_i} \}.$$

To find the marginal distribution of the data given the a, b, α, and β parameters we must complete the integration:

$$f(\{X_{ij}, Y_{ij}\} | a, b, \alpha, \beta) = \int_0^1 \int_0^1 \prod_{i=1}^{K} \{ \binom{n_i}{Y_{i+}} \binom{Y_{i+}}{Z_{i+}} [p_i\pi_i]^{Z_{i+}} [(1-p_i)\pi_i]^{Y_{i+}-Z_{i+}} [1-\pi_i]^{n_i-Y_i} \times [\Gamma(a+b)/\Gamma(a)\Gamma(b)] p_i^{a-1} (1-p_i)^{b-1} \times [\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)] \pi_i^{\alpha-1} (1-\pi_i)^{\beta-1} \} dp_i \, d\pi_i.$$

This integration can be solved simply by rewriting the integrands as beta probability density functions. The result of the integration is

$$\{ \prod_{i=1}^{K} \binom{n_i}{Y_{i+}} \binom{Y_{i+}}{Z_{i+}} B(Z_{i+}+a, Y_{i+}-Z_{i+}+b) \, B(Y_{i+}+\alpha, n_i-Y_{i+}+\beta) \} \times \{ B(a,b)B(\alpha,\beta) \}^{-1} \quad (4)$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ as before. The expression in equation (4) must be maximized using numerical methods in order to obtain the MLE's of the a, b, α, and β parameters. The maximization is made easier, in this case, by the fact that equation (4) may be factored into two parts - one a function of the a and b parameters alone and the other a function of the α and β parameters alone - which may be maximized separately. The part of the likelihood function involving only the a and b parameters for the distribution of the probabilities of victimization is the same as that given by Lehoczky and Schervish (1987).

The MLE's of the a, b, α, and β parameters are used to obtain the posterior distribution for the p_i and π_i parameters. Then the expected values of these parameters under the posterior distribution are used as the parameter estimates within each stratum.

For any stratum, i, the expected value of p_i is easily shown to be $(Z_{i+} + \hat{a})/(Y_{i+} + \hat{a} + \hat{b})$. Note that this is the usual mean of a posterior distribution for the binomial parameter, p, with a Beta(a,b) prior. For any stratum, i, the expected value of π_i is easily shown to be $(Y_{i+} + \hat{\alpha})/(n_i + \hat{\alpha} + \hat{\beta})$. This is also the usual mean of a posterior distribution for the binomial parameter with a beta prior.

4. FITS OF THE MODELS TO NCS DATA

In this section, we first present a brief description of the NCS and the data from that survey. Then we discuss the algorithm used to fit the two hierarchical models described in Sections 2 and 3. Finally, we present the results of fitting those models to NCS data and some randomly generated data.

4.1 The National Crime Survey

The NCS is a large-scale, household survey conducted by the U.S. Bureau of the Census for the Bureau of Justice Statistics. Data from the NCS is used to produce quarterly estimates of victimization rates and yearly estimates of the prevalence of crime. The survey uses a rotating panel of housing units (HU's) under which members of households (HH's) living in sampled HU's are interviewed up to seven times at six-month intervals. Individuals interviewed for the NCS are asked about crimes committed against them or against their property in the previous six months. The survey covers the following crimes and attempted crimes: assault, auto or motor vehicle theft, burglary, larceny, rape, and robbery. Crimes not covered by the NCS include kidnapping, murder, shoplifting, and crimes that occur at places of business. In this analysis, we let the probability of interest, p_i, be the probability that anyone within a HH in stratum i reports at least one victimization of any type for the previous six-month period.

Previous work with NCS data suggests that nonresponse in the data does not occur at random with respect to victimization status (see, for example, Saphire (1984) and Stasny (1989)). Thus, the hierarchical model of Section 2, under which the probability that a HH responds may depend on the victimization status of the HH, seems to be the more reasonable model to consider. The random nonresponse model of Section 3 will be fit for the purpose of comparison.

The data used in this work are from a large, longitudinal data set which includes all the regular NCS interview information collected from January 1975 to June 1979, except for the HU's that rotated into the sample in 1979. To make it easier to handle the data, this research uses only a subset of this large data set. The subset was created by taking a random start at the record for the eighth HU in the full data set and then every fifteenth record after that. Because the HU's on the original longitudinal file are ordered in such a way that units from the same cluster appear together, the 1-in-15 systematic sample should not include two or more HU's from a single cluster. Thus, this research does not consider the problem of correlations among HU's within clusters.

Additional information on the design and history of the NCS is provided, for example, by the U.S. Department of Justice and Bureau of Justice Statistics (1981).

4.2 The Data

In the following, we describe the fits of the hierarchical models of Sections 2 and 3 to two sets of data. The first set of data was randomly generated from distributions based on summaries of the NCS data. The raw NCS data suggests that the probability that a HH is touched by crime in a six-month period is about 0.2 and the overall probability that a HH responds to the survey is about 0.9. The data set, therefore, was generated to agree with these probabilities. The data was generated for $K = 10$ strata within the population with a sample of $n_i = 100$ HH's sampled from each stratum. The probability of victimization within the i^{th} stratum, p_i, was randomly chosen from a Beta(15,60) distribution so that $E[p_i] = 0.2$. For those HH's that were victimized, the probability of responding to the survey for HH's within stratum i, π_{i1}, was randomly chosen from a Beta(7,3) distribution so that $E[\pi_{i1}] = 0.7$. For HH's that were not victimized, the probability of responding to the survey, π_{i0}, was randomly chosen from a Beta(19,1) distribution so that $E[\pi_{i0}] = 0.95$. IMSL subroutines were used to generate the values of the p_i, π_{i1}, and π_{i0} parameters and the resulting data. The randomly generated data set is shown in Table 1.

The second set of data is NCS data which was collected in the first half of 1975. The data are post-stratified according to three neighborhood characteristics: 1) urban and rural, 2) central city, other incorporated place, and unincorporated or not a place, and 3) low poverty level (9% or fewer of families below poverty level) and high poverty level (10% or more of families below poverty level). Since it is practically impossible for a rural area to be a central city, this post-stratification results in $K = 10$ strata. The NCS data summarized according to these ten strata are shown in Table 2. Note that the sample sizes within eight of these ten strata are rather large and that it may not be necessary to borrow information from other strata in order to estimate the probabilities within a single stratum. In practice, the post-stratification used would most likely define much smaller sub-groupings, the corresponding sample sizes within each stratum would be much smaller, and the hierarchical, empirical-Bayes models may provide more accurate estimates within sub-groups than do standard procedures. We use the larger strata here for illustrative purposes.

4.3 Algorithm for Fitting the Model

Numerical algorithms for obtaining the MLE's of the a, b, α, and β parameters under both the nonrandom and random nonresponse models must be carefully written in order to avoid overflow, underflow, and rounding error problems on the computer. The computer programs for the analyses described here were written in double precision FORTRAN using IMSL subroutines to perform the required maximizations, to evaluate the complete beta functions, and to compute combinations. Since the complete beta function, $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, is large even for moderate values of a and b, the calculations were carried out using logarithms wherever possible.

In the case of the random nonresponse model, the logarithm of equation (4) was maximized using IMSL functions to evaluate the logarithm of the complete beta function. For the nonrandom nonresponse model, since the combinations may be expressed in terms of gamma functions, equation (2) may be rewritten as follows to facilitate maximization:

$$\prod_{i=1}^{K} \sum_{r=0}^{n_i-Y_{i+}} \exp\Big\{ \ln[\Gamma(n_i+1)] - \ln[\Gamma(Z_{i+}+1)] - \ln[\Gamma(Y_{i+}-Z_{i+}+1)] - \ln[\Gamma(r+1)] - \ln[\Gamma(n_i-Y_{i+}-r+1)] + \ln[B(Z_{i+}+a+r,\, n_i-Z_{i+}+b-r)] - \ln[B(a,b)] + B(Z_{i+}+\alpha_1,\, \beta_1+r) - \ln[B(\alpha_1,\beta_1)] + \ln[B(Y_{i+}-Z_{i+}+\alpha_0,\, n_i-Y_{i+}+\beta_0-r)] - \ln[B(\alpha_0,\beta_0)] \Big\}.$$

Again, the logarithm of the above equation was used in the maximization.

An additional problem encountered in the data analyses described here was that, because the probability of responding was rather large, the maximization routine tended to converge towards an estimate of 1.0 for the probability of responding, skipping possible estimates which gave larger values of the likelihood function. To avoid this problem, each Beta(a,b) distribution in both the random and nonrandom nonresponse cases was reparameterized in terms of $a/(a+b)$ and $(a+b)$. In addition, in the nonrandom nonresponse case a partial grid search was used to locate a reasonable starting point for π_{i0} in the iterative procedure.

4.4 Results

The random and nonrandom nonresponse models were used to obtain parameter estimates from both the simulated data and the actual NCS data. These parameter estimates are given in Table 1 for the randomly generated data and in Table 2 for the

Table 1. Data randomly generated with $p_i \sim$ Beta(15,60), $\pi_{i1} \sim$ Beta(7,3), and $\pi_{i0} \sim$ Beta(19,1) so that $E[p_i] = 0.2$, $E[\pi_{i1}] = 0.7$, and $E[\pi_{i0}] = 0.95$. Sample sizes of $n_i = 100$ for $i = 1, 2, \ldots 10$.

	Randomly Generated Data			Randomly Generated Probabilities			Naive Estimator		Random Nonresponse		Nonrandom Nonresponse		
i	$Y_{i+}-Z_{i+}$	Z_{i+}	n_i-Y_{i+}	p_i	π_{i1}	π_{i0}	$\hat{p}_i$	$\hat{\pi}_i$	$\hat{p}_i$	$\hat{\pi}_i$	$\hat{p}_i$	$\hat{\pi}_{i1}$	$\hat{\pi}_{i0}$
1	72	19	9	.267	.765	.920	.209	.91	.168	.906	.176	.861	.914
2	73	21	6	.222	.813	.966	.223	.94	.170	.921	.179	.861	.931
3	76	17	7	.243	.768	.977	.183	.93	.164	.916	.172	.861	.927
4	72	17	11	.226	.541	.977	.191	.89	.165	.897	.173	.861	.903
5	68	15	17	.210	.796	.860	.181	.83	.163	.867	.171	.861	.866
6	80	10	10	.211	.629	.986	.111	.90	.153	.901	.161	.861	.913
7	83	12	5	.189	.738	.976	.126	.95	.155	.926	.163	.861	.940
8	72	14	14	.215	.803	.879	.163	.86	.161	.882	.169	.861	.886
9	76	10	14	.175	.527	.954	.116	.86	.154	.882	.161	.861	.889
10	86	10	4	.171	.565	.986	.104	.96	.152	.931	.159	.861	.946

where Y_{i+} = number responding in stratum i, $Y_{i+}-Z_{i+}$ = number reporting crime free in stratum i, Z_{i+} = number reporting victimizations in stratum i, and n_i-Y_{i+} = number of nonrespondents in stratum i.

NCS data. In addition to the parameter estimates obtained under the two hierarchical models, "naive" parameter estimates are provided in both tables. These estimates are obtained using only the information in an individual stratum to obtain the estimates for that stratum. Thus, the naive estimators in stratum i are $\hat{p}_i = Z_{i+}/Y_{i+}$ and $\hat{\pi}_i = Y_{i+}/n_i$.

Consider the effects of the hierarchical estimation schemes on the parameter estimates in the case of the randomly generated data in Table 1. The naive estimates of p_i are simply the observed proportions of victimized HH's in the ten strata ignoring nonrespondents. The estimates of p_i in Table 1 under the random nonresponse model are pulled from the naive estimates towards the overall proportion of victimized HH's, $Z_{++}/Y_{++} = 145/903 = .161$. Similarly, the estimates of π_i under the random nonresponse model are pulled from the naive estimates, the observed proportions of respondents in each stratum, towards the overall proportion of respondents, $Y_{++}/n_+ = 903/1000 = .903$. In this way the information from all strata is used to estimate the probabilities of victimization and nonresponse in each individual stratum.

Under the nonrandom nonresponse model, the estimates of p_i shown in Table 1 are again pulled towards an overall probability of victimization but in this case that overall probability is somewhat larger than the naive overall estimate because it has been adjusted for the fact that victimized HH's are less likely to respond than are crime-free HH's. Thus the estimates of the probabilities of victimization are all larger under the nonrandom nonresponse model than under the random nonresponse model. The estimates of the probabilities of responding for crime-free HH's are generally larger under the nonrandom nonresponse model than are the single response probabilities under the random nonresponse model (the only exception occurs in the fifth stratum). The estimates of the probabilities of responding for victimized HH's are all smaller under the nonrandom nonresponse model than are the single response probabilities under the random nonresponse model. The values of $\hat{\pi}_{i0}$ shown in Table 1 are all identical to three decimal places because, in this example, the estimated value of $\alpha_0 + \beta_0$ is very large. Since this term appears in the denominator of the variance of the prior beta distribution for the π_{i0}, the prior variance is quite small. Thus, the information from the sample does not greatly effect the estimates of π_{i0}'s.

Using the results presented in Table 1 for the simulated data, we may compare the naive, random nonresponse, and nonrandom nonresponse estimates to the actual parameter values to determine how effective the hierarchical models are. The mean absolute errors and root mean squared errors for the $\hat{p}_i$, $\hat{\pi}_{i1}$, and $\hat{\pi}_{i0}$ are given in Table 3. Note that for the naive and random nonresponse estimators, the single estimator of the probability of responding, $\hat{\pi}_i$, is compared to both π_{i1} and π_{i0} since in those cases the probability of responding is taken to be the same for both victimized and crime-free HH's. The errors shown in Table 3 indicate that the naive and random nonresponse estimators are approximately the same in terms of mean absolute errors and root mean squared errors while the errors associated with the nonrandom nonresponse model are somewhat smaller.

Now consider the results for the actual NCS data presented in Table 2. Again, the estimates of the p_i under the random nonresponse model are pulled from the naive estimates towards the overall proportion of victimized HH's, $Z_{++}/Y_{++} = 727/3630 = .2003$. Similarly, the estimates of π_i under the random nonresponse model are pulled from the naive estimates towards the overall proportion of respondents, $Y_{++}/n_+ = 3630/4155 = .8736$. In this way the information from all strata is used to estimate the probabilities of victimization and nonresponse in each individual stratum. Note, of course, that in the stratum where the sample sizes are particularly large the estimate is not pulled towards the overall proportion as much as it is in cases where the sample size is smaller.

In the case of the nonrandom nonresponse model, the estimates of the p_i are again pulled towards a larger overall probability of victimization which has been adjusted for the fact that victimized HH's appear to be less likely to respond than are crime-free HH's. Thus the estimates of the probabilities of victimization are all larger under the nonrandom nonresponse model than under the random nonresponse model. The estimates of the probabilities of responding for crime-free HH's are all larger under the nonrandom nonresponse model than are the single response probabilities under the random nonresponse model. The estimates of the probabilities of responding for victimized HH's are all smaller under the nonrandom nonresponse model than is the single probability of responding under the random nonresponse model. The values of $\hat{\pi}_{i1}$ are all identical to three decimal places because the estimated value of $\alpha_1 + \beta_1$ is very large and, hence, the variance of the prior distribution of π_{i1} is quite small. Thus, the information from the sample does not greatly effect the estimates of π_{i1}.

5. CONCLUSIONS AND FUTURE WORK

We have developed hierarchical models for the probabilities of victimizations and nonresponse and fit those models to randomly generated data and actual data from the NCS. The hierarchical models allow for either random or nonrandom nonresponse. The nonrandom nonresponse model fit to the

Table 2. National Crime Survey data from 1/75 - 6/75 stratified by neighborhood characteristics urban or rural (U or R), central city, other incorporated place, or unincorporated or not a place (C, I, or N), and low or high poverty level (L or H).

	NCS Data			Naive Estimator		Random Nonresponse		Nonrandom Nonresponse		
strata	$Y_{i+}-Z_{i+}$	Z_{i+}	n_i-Y_{i+}	$\hat{p}_i$	$\hat{\pi}_i$	$\hat{p}_i$	$\hat{\pi}_i$	$\hat{p}_i$	$\hat{\pi}_{i1}$	$\hat{\pi}_{i0}$
U/C/L	555	156	104	.219	.872	.217	.873	.272	.689	.937
U/C/H	364	95	73	.207	.863	.205	.869	.265	.684	.937
U/I/L	557	162	101	.225	.877	.222	.876	.276	.692	.937
U/I/H	262	72	36	.216	.903	.212	.885	.254	.694	.937
U/N/L	297	92	79	.237	.831	.230	.855	.305	.679	.937
U/N/H	40	15	9	.273	.859	.228	.872	.287	.687	.937
R/I/L	36	11	7	.234	.870	.210	.873	.265	.687	.937
R/I/H	105	10	20	.087	.852	.130	.870	.185	.682	.937
R/N/L	274	35	32	.113	.906	.129	.886	.166	.687	.937
R/N/H	413	79	64	.161	.885	.165	.879	.213	.686	.937

where Y_{i+} = number responding in stratum i, $Y_{i+}-Z_{i+}$ = number reporting crime free in stratum i, Z_{i+} = number reporting victimizations in stratum i, and n_i-Y_{i+} = number of nonrespondents in stratum i.

Table 3. Errors in estimation for randomly generated data.

	Naive Estimator			Random Nonresponse			Nonrandom Nonresponse		
	$\hat{p}_i$	$\hat{\pi}_{i1}$	$\hat{\pi}_{i0}$	$\hat{p}_i$	$\hat{\pi}_{i1}$	$\hat{\pi}_{i0}$	$\hat{p}_i$	$\hat{\pi}_{i1}$	$\hat{\pi}_{i0}$
mean absolute error	.052	.209	.045	.052	.208	.047	.045	.167	.039
root mean squared error	.058	.240	.054	.057	.237	.055	.050	.200	.047

simulated data succeeded in capturing the difference in response probabilities for victims and non-victims which was present in the distributions from which the data were generated. Since the parameter estimates obtained when the nonrandom nonresponse model was fit to the actual NCS data show similar differences for victims and non-victims, it seems reasonable to conclude that nonresponse in the NCS is informative nonresponse. The values of the parameter estimates suggest that victims of crime are less likely to respond to the survey than are non-victims. Any estimation procedures that do not allow for this difference will result in estimates of probabilities of victimizations that are biased downwards.

The empirical-Bayes approach taken here has the advantage of allowing information from all strata to be used to provide estimates of probabilities within each stratum. The disadvantage is that the computations are more difficult than for the standard, non-hierarchical approach. It may be possible to improve the computation procedure in the future. Obtaining variance estimates under these hierarchical models is an additional problem. One must be wary of variance estimates based on using the MLE's as the parameters in the beta priors for these hierarchical models because such variance estimates would not include the uncertainty in the MLE's themselves. A possible remedy for this problem is suggested by Morris (1983).

Areas for future research include extending these hierarchical models to allow the probabilities of victimizations to be influenced by covariates in the data. Saphire (1989) has developed hierarchical models for estimating the number of victimizations experienced by a HH which make use of covariates but do not address the nonresponse problem. Another extension of the models would be to allow them to handle the longitudinal nature of NCS data. Lehoczky and Schervish (1987) suggest a hierarchical Markov-chain model for victimizations and Stasny (1987) has presented Markov-chain models which are not hierarchical but which do allow for random or nonrandom nonresponse. Perhaps these ideas could be combined to develop hierarchical Markov-chain models for both victimizations and nonresponse.

REFERENCES

Lehoczky, J. P. and Schervish, M. J. (1987), "Hierarchical Modelling and Multi-level Analysis Applied to the National Crime Survey", paper presented at the Workshop on the National Crime Survey, July 6-17, 1987.

Morris, C. N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications", Journal of the American Statistical Association, **78**, 47-65.

Saphire, D. G. (1984), Estimation of Victimization Prevalence Using Data from the National Crime Survey, Lecture Notes in Statistics, Vol. 23, Springer-Verlag, New York.

Saphire, D. G. (1989), "An Empirical Bayes Model with Covariates Applied to Victimization", paper presented at the Joint Statistical Meetings, August 6-10, 1989.

Stasny, E. A. (1987), "Some Markov-Chain Models for Nonresponse in Estimating Gross Labor Force Flows", Journal of Official Statistics, **3**, 359-373.

Stasny, E. A. (1989), "Symmetry in Flows Among Reported Victimization Classifications with Nonrandom Nonresponse", Technical Report No. 412, Department of Statistics, The Ohio State University.

U.S. Department of Justice and Bureau of Justice Statistics (1981), The National Crime Survey: Working Papers, Volume I: Current and Historical Perspectives, Washington DC.

The National Crime Survey data utilized in this paper were made available by the Inter-University Consortium for Political and Social Research. The data were originally collected by the United States' Law Enforcement Assistance Administration. The longitudinal data set used here was created by the Bureau of Justice Statistics using the quarterly public-use data files.