Journal of Official Statistics, Vol. 14, No. 1, 1998, pp. 61–78

# Random-effects Models for Smoothing Poststratification Weights

*Laura C. Lazzeroni*[1] *and Roderick J.A. Little*[2]

Poststratification is a common technique for adjusting survey data using external data from a census or larger survey. When the respondent counts in the poststrata are small, modifications of the method, such as collapsing over adjacent poststrata, are needed to reduce variability in the poststratification weights. We consider here inference about a population mean with ordered poststrata. One approach is to treat poststratum means as random effects, yielding shrinkage towards the unweighted mean, but this method provides unsatisfactory inferences when the means vary systematically across the poststrata. We consider alternative model-based extensions of this method, where the poststratum means are assumed to be distributed about a linear regression line, and where the poststratum means are assumed to have an autoregressive covariance structure. The methods are illustrated on a real data set from the Epidemiologic Catchment Area study, and compared with other procedures in a simulation study. The latter suggests that the autoregressive random effects model may be a useful approach to the problem.

*Key words:* Empirical Bayes; random effects; survey inference; superpopulation model; simulation study.

## 1. Introduction

In the survey setting, stratified sampling is useful when survey outcome variables are related to survey design variables that are observed for all units in the population. Stratification based on the design variables can reduce bias due to fallible sampling procedures, and reduce variance by eliminating the between-stratum component of the variability of the resulting estimates.

Sometimes, the population distribution is known for a secondary variable that is impossible to observe on a given individual prior to sampling. In this case, it is still possible to use the distribution to adjust estimates of the outcome in the analysis stage using the technique known as poststratification. Poststratification can reduce bias caused by problems in the sampling frame or unit nonresponse, and it can also increase the precision of estimates. However, since the secondary variable is not a stratifier in the sample design, the respondent counts in the poststrata are not under the control of the sampler and hence the method can lead to a reduction in precision. This article concerns modifications

[1] Division of Biostatistics, Stanford University, HRP Redwood Bldg., T101D, Stanford, CA 94305-5405, U.S.A.
[2] Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A.
**Acknowledgments:** This research was supported by NIMH grant MH 37188, and by U.S. Bureau of the Census Contract No. 50-YABC-2-66023 with the National Opinion Research Center. We thank the Associate Editor and three referees for valuable comments on an earlier draft.

of the usual poststratified estimator of the mean to limit the effects of excessive variability in the poststratum counts.

We consider the problem of estimating the population mean of an outcome variable when an ordinal poststratifying variable such as an index, or an interval-scaled poststratifier such as age or income, is available. We propose estimates based on two models for the outcome that reflect the ordinal nature of the poststratifier. An example is given using data from the 1979 Los Angeles Epidemiological Catchment Area survey. The methods are compared with previously suggested alternatives in a large simulation study.

Section 2 briefly reviews alternative approaches to this problem. Section 3 presents the model based methods. Section 4 contains an example with comparative results for some of these methods. Section 5 describes a simulation study comparing inferences under a variety of approaches. Section 6 summarizes our conclusions.

## 2. Current Approaches

We consider the estimation of a finite population mean of a continuous outcome variable $Y$ based on a sample survey, in the presence of an ordinal poststratifier $W$ with known population distribution. Suppose $W$ has $H$ levels. For poststratum $h$, let $N_h$ be the population size; $n_h$, the sample size; $Y_{hi}$, the value of $Y$ for the $i$th individual in the population; and $y_{hi}$, the value of $Y$ for the $i$th sample observation. Also, let:

$$N_+ = \sum_h N_h; n_+ = \sum_h n_h; \bar{Y}_+ = \sum_{h,i} Y_{hi}/N; \bar{Y}_h = \sum_i Y_{hi}/N_h; \bar{y}_h = \sum_i y_{hi}/n_h$$

denote respectively the total population size, the total sample size, the population mean, the within poststratum population mean, and the within poststratum sample mean.

It is assumed that given the value of $W$, the probability of inclusion in the sample does not depend upon the value of $Y$. Two standard approaches are to estimate $\bar{Y}_+$ by (a) the unweighted sample mean (unw) $\bar{y}_+ = \Sigma_h p_h \bar{y}_h$, where $p_h = n_h/n_+$; and (b) the poststratified mean (psm) $\bar{y}_{ps} = \Sigma_h P_h \bar{y}_h$, where $P_h = N_h/N_+$. The sample mean is an appropriate estimate of $\bar{Y}_+$ when $Y$ and $W$ are unrelated. Even if the variables are related, $\bar{y}_+$ is design unbiased so long as the probability of inclusion in the sample does not depend upon $W$. However, $\bar{y}_+$ does not use information from the known distribution of $W$ so that it may not be the best estimate. Due to sampling variability or systematic bias in the sampling procedures, the proportion $p_h$ falling within stratum $h$ of a given sample deviates from its respective population proportion $P_h$. Hence $\bar{y}_+$ is biased for $\bar{Y}_+$ conditional upon the sample poststratum counts $\{n_h\}$.

The poststratified mean $\bar{y}_{ps}$ is unbiased given the sample poststratum counts, and can have much smaller mean squared error than $\bar{y}_+$ (Holt and Smith 1979; Little 1993). Although $\bar{y}_{ps}$ incorporates information about the population distribution of $W$, $\bar{y}_h$ is used to estimate the poststratum mean, regardless of how many respondents fall in the poststratum. When a stratum contains few observations, the estimator of the stratum mean might be improved by borrowing strength from information from neighboring strata. Also, the poststratified mean has the same form as for an unordered categorical poststratifier, and thus does not reflect the ordinal nature of the poststratifier. From a model-based perspective, the objective is to improve precision by using estimates of the poststratum means that combine strength across poststrata.

A different perspective is to regard estimates as weighted averages of the sample observations. Both $\bar{y}_+$ and $\bar{y}_{ps}$ can be written as

$$\sum_{h,i} w_h y_{hi} / \sum_{h,i} w_h$$

where $w_h$ is a weight attached to each observation in stratum $h$. Conditional on the sample poststratum counts, the variance of these estimates under simple random sampling is

$$\sum_{h,i} w_h^2 \left(1 - \frac{n_h}{N_h}\right) S_h^2 / \left(\sum_{h,i} w_h\right)^2$$

where $S_h^2$ is the within-stratum variance of $Y_{hi}$. For $\bar{y}_{ps}$, $w_h = P_h/p_h$, and for $\bar{y}_+$, $w_h = 1$ for all $h$. When $p_h$ is much smaller than $P_h$, cases in that poststratum receive a high weight, inflating the variance of $\bar{y}_{ps}$. In fact, $p_h$ can equal zero, in which case adjustments are needed for $\bar{y}_{ps}$ to be defined.

Modifications of $\bar{y}_{ps}$ that reduce its variance can often be written as weighted averages of the observations where the original poststratification weights have been smoothed to reduce variability. One method is to truncate the poststratification weights larger than some maximum allowable value. Simultaneously, smaller weights are adjusted upwards. The truncation point may be fixed in an *ad hoc* way, or based on the data (Potter 1990). A second approach is to pool or collapse strata. Strategies for choosing how and when to collapse strata have been suggested by Kalton and Maligalig (1991), Little (1993), and Tremblay (1986). If sampling or nonresponse depends upon $W$, modeling of these rates has been suggested (e.g., Kalton and Maligalig 1991). Observation weights can be based on the estimated rates, which will usually be smoother than the observed rates.

## 3. Proposed Methods

The ideal compromise between $\bar{y}_+$ and $\bar{y}_{ps}$ would use the ordinal structure of $W$ to aid in the prediction of the stratum means. It would control variance by not weighting any individual observation too highly. When the sample means are well-observed and $Y$ is strongly related to $W$, the estimate should look like $\bar{y}_{ps}$. When $Y$ and $W$ are not strongly related, the estimate should look like $\bar{y}_+$. Also, since surveys contain large numbers of variables, the ideal method would have general applicability without requiring a lot of hands-on modeling for each outcome. However, if arbitrary choices are needed for the sake of generality, the resulting estimates should be insensitive to these choices.

We consider methods based on models for the outcome, which can be viewed within either a superpopulation or Bayesian framework. The general form of these models is

$$Y_{hi} | \mu_h \sim_{ind} N(\mu_h, \sigma^2) \tag{1}$$

and

$$\mu \sim N_H(X\beta, D) \tag{2}$$

where $\mu = (\mu_1, \ldots, \mu_H)^T$, $X$ is a known $H \times Q$ design matrix, $\beta$ is a $Q \times 1$ vector of unknown parameters and $D$ is an $H \times H$ covariance matrix (Harville 1977; Laird and Ware 1982). For applications of this model to cluster sampling, see Scott and Smith

(1969) and Pfefferman and Nathan (1981). For individuals not included in the sample, $Y_{hi}$ can be estimated by $\hat{\mu}_h$, its expected value given the data. The estimated finite population mean is

$$\bar{y}_{\text{mod}} = \sum_h [n_h \bar{y}_h + (N_h - n_h)\hat{\mu}_h]/N_+ \tag{3}$$

The estimates $\hat{\mu}_h$ of the stratum means shrink the sample means $\bar{y}_h$ towards the $h$th element of $X\hat{\beta}$, with a degree of shrinkage that tends to zero as the within-stratum sample size $n_h$ increases. The estimator smoothes the within-stratum means when the sample size is small, but it behaves like the poststratified mean in large samples and hence is design consistent.

Previous work has discussed inference under the exchangeable random effects (XRE) model obtained from Equation (2) by setting $X = (1, \ldots, 1)^T$, $D = I\sigma_\mu^2$, where $I$ is the identity matrix (Holt and Smith 1979; Little 1983, 1991; Ghosh and Meeden 1986). If the between poststratum variance $\sigma_\mu^2$ is set equal to zero then $\mu_h \equiv \mu$, $\hat{\mu}_h = \bar{y}_+$ and $\bar{y}_{\text{mod}} = \bar{y}_+$. If $\sigma_\mu^2$ is set equal to infinity, a fixed effects ANOVA model is obtained, $\hat{\mu}_h = \bar{y}_h$ and $\bar{y}_{\text{mod}} = \bar{y}_{ps}$. If $\sigma_\mu^2$ is estimated from the data, an empirical Bayes approach, the resulting estimate $\bar{y}_{\text{mod}}$ moves the means $\bar{y}_h$ towards $\bar{y}_+$. While these properties are appealing, simulations in Little (1991) indicate that confidence intervals based on the XRE model are sensitive to departures from the assumption of exchangeability in the post-stratum means. This finding is consistent with comments by Morris (1983) in the general context of empirical Bayes estimation. The exchangeability assumption is highly questionable when $W$ is ordinal, since a systematic relationship between $Y$ and $W$ might be expected. We propose and study one-parameter extensions of the XRE model for an ordinal poststratifier.

Two ways are suggested for modeling the ordinal nature of $W$ based on Equation (2). The more standard approach is to model the mean structure by including functions of $W$ in $X$. In particular, the (REG) model for the stratum means sets

$$X = \begin{bmatrix} 1 & W_1 \\ \ldots & \ldots \\ 1 & W_H \end{bmatrix} \quad \text{and} \quad D = I\sigma_\mu^2 \tag{4}$$

In our application the poststrata are equally-spaced and $W_h = h$. Another approach is to model the covariance matrix $D$ to incorporate greater positive correlation between $\mu_h$ and $\mu_{h'}$ when $h$ and $h'$ are close in value. A simple way to do this is to assume an AR1 model of the stratum means, namely

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \ldots \\ 1 \end{bmatrix}, D = \sigma_\mu^2 \begin{bmatrix} 1 & \rho & \ldots & \rho^{H-1} \\ \rho & 1 & \ldots & \rho^{H-2} \\ \rho^2 & \rho & \ldots & \rho^{H-3} \\ \ldots & \ldots & \ldots & \ldots \\ \rho^{H-1} & \rho^{H-2} & \ldots & 1 \end{bmatrix} \tag{5}$$

The AR1 model is proposed as a pragmatic device, since it induces differential shrinkages between the poststrata so that closer poststrata contribute more than more distant poststrata. The AR1 structure of $D$ effectively assumes the levels of the poststratifier are equally-spaced, but this could be modified to reflect unequal spacings if necessary, by raising the correlations to powers of the absolute differences of the values of $W$. The REG and AR1 models each contain four parameters, compared with three for the XRE model. Of course these models could be elaborated, or a combined REG/AR1 model with five parameters could be fitted, with some loss of parsimony; here we examine the properties of the REG and AR1 models treated separately.

Assume that exactly $G$ poststrata contain data. The sufficient statistics are $\bar{y}$ and $S$ where $\bar{y}$ is the $G \times 1$ vector consisting of the ordered observed poststratum means and $S = \Sigma_{h,i}(y_{hi} - \bar{y}_h)^2$. Let $Z$ be the $G \times H$ submatrix of the $H \times H$ identity matrix with row $h$ deleted if $n_h = 0$. Let $R$ be the $G \times G$ diagonal matrix with the $g$th diagonal element equal to $\sigma^2/n_h$ if $h$ is the $g$th stratum for which $n_h > 0$. Define $V = ZDZ^T + R$. Conditional on the observed poststratum counts,

$$\bar{y} \sim N_G(ZX\beta, V)$$

and independently

$$S/\sigma^2 \sim \chi^2_{n-G}$$

The maximum likelihood estimate of $\beta$ for known $V$ is

$$\hat{\beta} = (X^T Z^T V^{-1} ZX)^{-1} X^T Z^T V^{-1} \bar{y}$$

yielding predicted means

$$\hat{\mu} = X\hat{\beta} + DZ^T V^{-1}(\bar{y} - ZX\hat{\beta}) = A\bar{y}$$

say, where $A$ is the matrix obtained by substituting for $\hat{\beta}$. Substitution in Equation (3) yields a model estimate $\bar{y}_{\text{mod}}$ of the finite population mean, which has the form of a weighted average of the observations for some set of smoothed weights, $\{w_h, h = 1, \ldots, H\}$. Estimates of variance based on Model (2), ignoring error in estimating $V$, can be derived by a standard empirical Bayes analysis, yielding

$$Var(\bar{y}_{\text{mod}} - \bar{Y}) = (N - n)^T (T + ARA^T + (I - AZ)D(I - AZ)^T)(N - n)/N_+^2 \tag{6}$$

where $N$ and $n$ are $(G \times 1)$ vectors of population and sample counts in the poststrata with data, and $T$ is the $(G \times G)$ diagonal matrix with diagonal elements $\sigma^2/(N_h - n_h)$. Expressions (3) and (6) are also the posterior mean and variance of $\bar{Y}$ from a Bayesian analysis with a uniform prior for $\beta$, ignoring uncertainty in estimating $V$. When, as is usual, the variance parameters in $V$ are unknown, the approach adopted here is to substitute maximum likelihood estimates, so that $D$ becomes $\hat{D}$ and $V$ becomes $\hat{V}$ in Equations (3) to (6). Approximate $t$-based corrections are applied for interval estimation, as discussed in Section 5.4. Exact Bayesian methods can be developed that allow for uncertainty in estimating $V$, but they are not considered here to keep calculations simple.

## 4. Example

To illustrate our methods, we apply them to data from the Los Angeles Epidemiologic

Catchment Area survey (for example, Eaton and Kessler 1985). This large mental health survey was based on an equal probability sample of households in two areas, East Los Angeles and West Los Angeles. Population distributions within these catchment areas were taken from the 1980 U.S. Census. Data for the age-specific sampling proportions and population distributions are given in Little (1993).

Eight demographic groups defined by Ethnicity (H = Hispanic, N = Not Hispanic), Gender (F = Female, M = Male) and Catchment Area (E = East, W = West) were each analyzed separately. Sample sizes varied from 112 to 738. The outcome variable $Y$ was a score measuring depression based on a set of questions from the survey. Although not continuous, $Y$ takes 71 distinct values from 0 to 51. The sample was poststratified on age, $W$, with each year representing one poststratum. Respondents varied in age from 18 to 96 while individuals in the population were recorded at age 102. All demographic groups had some strata that were not represented in the sample. (Single year poststrata are chosen to yield an extreme method with minimal shrinkage; in practice the number of poststrata would usually be reduced, say by choosing five-year age intervals).

Figure 1 displays data from 3 groups, HMW, NME, and NMW. The left panels plot $Y$ against $W$, and show a downward trend with age but with lots of variability especially at younger ages. Plots of $P_h$ and $p_h$ against age not included here suggest that some age ranges may be systematically undersampled.

Fits of the AR1 and REG models applied to these data are summarized in Table 1. We used a modified version of the scoring algorithm described by Jennrich and Schluchter (1986) that includes information from $S$ as well as the poststratum means. Penalty functions were used to insure that $\hat{\sigma}^2 > 0$, $\hat{\sigma}_\mu^2 > 0$ and $0 < \hat{\rho} < 1$. Under these constraints, $\hat{\sigma}_\mu^2$ was essentially zero in three cases for the AR1 model and six cases for the REG model. For the AR1 model, this means that $\hat{\rho}$ is meaningless and it is not reported. The estimated slope is always negative under the REG model. Because the AR1 model describes the stratum means rather than the observations, $\hat{\rho}$ can be quite large, greater than 0.9 in two cases. A simple intercept model for the outcomes was also fit to the data. This is the simplest superpopulation model that leads to the estimate $\bar{y}_+$. For the AR1 and the REG models, Table 1 contains the difference of the log-likelihood from this simplest model as well as maximum likelihood estimates of the parameters.

For three groups, the right panels of Figure 1 show $\hat{\mu}_h$, the estimated or predicted means, plotted against age, from poststratification, and from the AR1 and REG models. For the sample mean, the predicted means are identically equal to $\bar{y}_+$. Straight lines correspond to $\hat{\sigma}_\mu = 0$ since the predicted mean is then just the estimated fixed effect. The NME group has the least smooth predicted values and the largest ratio of $\hat{\sigma}_\mu^2$ to $\hat{\sigma}^2$. This group also has the smallest meaningful estimates of $\rho$ for the AR1 model. Plots of the weights against age are shown for two groups in Figure 2, with the circles denoting poststratification weights, dashed lines weights from the REG model and continuous lines weights from the AR1 model. The variability of the weights from REG and AR1 in this example is small compared to truncation-based methods that have been proposed. The AR1 model displays weights that vary with the observed sampling rates whereas the weights of the REG model are usually linear.

Table 2 shows five estimates of the finite population mean, namely the unweighted mean, the poststratified mean and the predicted mean from the XRE, REG and AR1 models,

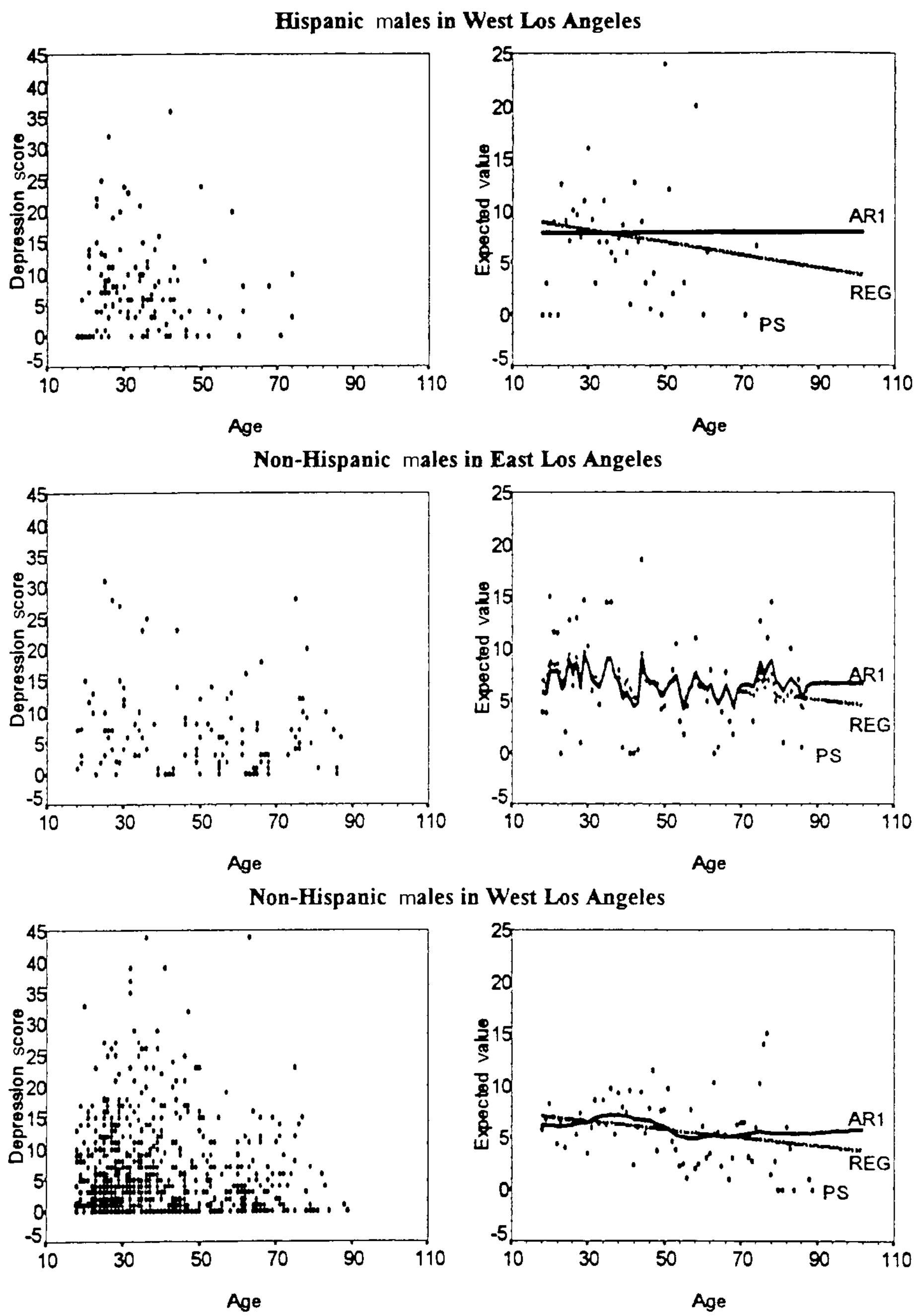


*Fig. 1. Observed depressions scores (on left) and predicted stratum means under PS, REG, and AR1 models (on right)*

*Table 1. Maximum likelihood estimation for AR1 and REG models in eight groups*

| Group | Model | Difference in log-likelihood | MLE of intercept | MLE of $\sigma^2$ | MLE of $\sigma_\mu^2$ | MLE of $\rho$ | MLE of slope |
|---|---|---|---|---|---|---|---|
| HFE | AR1 | 0.00 | 8.88 | 62.13 | 0.00 | NA | |
| $N = 46,301$ | REG | 0.60 | 9.72 | 61.99 | 0.00 | | −0.02 |
| $n = 543$ | | | | | | | |
| HME | AR1 | 0.30 | 6.82 | 42.66 | 0.28 | 0.82 | |
| $N = 42,349$ | REG | 1.14 | 7.93 | 42.76 | 0.00 | | −0.03 |
| $n = 492$ | | | | | | | |
| HFW | AR1 | 0.01 | 8.81 | 97.46 | 0.66 | 0.50 | |
| $N = 11,120$ | REG | 0.27 | 10.41 | 97.74 | 0.00 | | −0.04 |
| $n = 142$ | | | | | | | |
| HMW | AR1 | 0.00 | 7.90 | 51.15 | 0.00 | NA | |
| $N = 11,448$ | REG | 0.63 | 10.05 | 49.59 | 0.00 | | −0.06 |
| $n = 112$ | | | | | | | |
| NFE | AR1 | 0.12 | 6.87 | 65.51 | 0.92 | 0.90 | |
| $N = 12,581$ | REG | 2.08 | 10.06 | 64.87 | 0.00 | | −0.06 |
| $n = 178$ | | | | | | | |
| NME | AR1 | 1.47 | 6.66 | 37.18 | 6.46 | 0.30 | |
| $N = 11,287$ | REG | 1.98 | 8.66 | 37.37 | 5.64 | | −0.04 |
| $n = 124$ | | | | | | | |
| NFW | AR1 | 0.00 | 6.99 | 67.22 | 0.00 | NA | |
| $N = 55,499$ | REG | 0.63 | 7.88 | 67.11 | 0.00 | | −0.02 |
| $n = 738$ | | | | | | | |
| NMW | AR1 | 2.01 | 5.84 | 47.66 | 0.98 | 0.92 | |
| $N = 53,593$ | REG | 2.79 | 7.81 | 48.01 | 0.07 | | −0.04 |
| $n = 711$ | | | | | | | |

for all eight groups. Estimated standard errors based on Equation (6) are also shown for each estimator. For these data sets the XRE model estimates are generally much closer to the unweighted mean than to the poststratified mean, reflecting the large degree of shrinkage of the weights. The AR1 and REG estimates are often similar to those from XRE and have comparable estimated standard errors, but occasionally move towards the poststratified mean. The poststratified means usually have somewhat larger standard errors than the other methods. Confidence intervals based on the poststratified mean would be even wider (by 20–30%) if a *t*-correction is applied, which appears necessary from the results of the simulation study in Section 5.

## 5. Simulation Study

### *5.1. Overview*

A sizable simulation study was carried out to assess the inferential properties of alternative methods under correctly-specified and misspecified models. In all, $9 \times 2 \times 2 \times 2 = 144$ simulation conditions were created, by crossing $9 \times 2 = 18$ population types, 9 normal and 9 lognormal, with $2 \times 2 = 4$ sample types, defined by mechanism (MAR, MCAR) and sample size (50, 250). MAR mechanisms, where selection rates vary across the

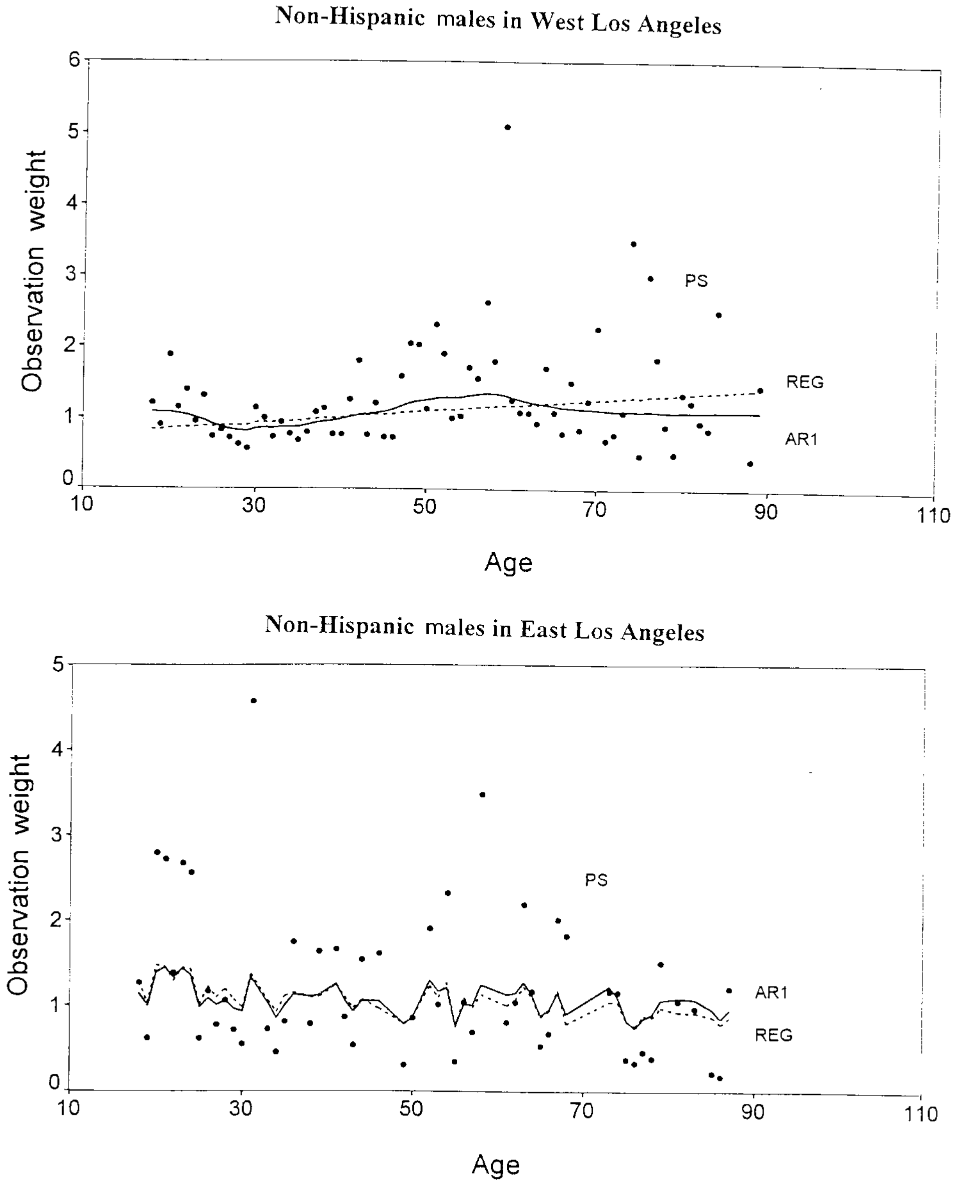


*Fig. 2. Observation weights from PS, REG, and AR1 models for two groups*

poststratifier, are included to model situations where frame limitations or unit nonresponse result in a sample that is not missing completely at random. Non-MAR situations are not included since none of the methods are designed to address non-ignorable selection. For each simulation condition, 50 populations were generated and 10 samples generated within each population, yielding 500 replications. Six estimation methods for the finite population mean were applied to these replicates, and the bias and mean squared error assessed. Also, the width and coverage of 95% intervals for the finite population mean were compared for a variety of procedures.

*Table 2. Estimate (standard error) of mean depression score from five methods in eight subgroups*

| Method | Subgroup HFE | HME | HFW | HMW |
|---|---|---|---|---|
| $\bar{y}_{+}$ | 8.88 (.34) | 6.84 (.30) | 8.82 (.83) | 7.90 (.67) |
| $\bar{y}_{ps}$ | 8.92 (.35[a]) | 7.00 (.35[a]) | 8.69 (.88[a]) | 7.63 (.76[a]) |
| $\bar{y}_{xre}$ | 8.88 (.34) | 6.84 (.29) | 8.82 (.83) | 7.90 (.67) |
| $\bar{y}_{ar1}$ | 8.88 (.34) | 6.89 (.30) | 8.82 (.83) | 7.90 (.67) |
| $\bar{y}_{reg}$ | 8.88 (.34) | 6.85 (.29) | 8.77 (.83) | 7.85 (.66) |

| Method | Subgroup NFE | NME | NFW | NMW |
|---|---|---|---|---|
| $\bar{y}_{+}$ | 6.94 (.61) | 6.61 (.60) | 6.68 (.30) | 6.25 (0.26) |
| $\bar{y}_{ps}$ | 6.73 (.66[a]) | 7.09 (.70[a]) | 6.99 (.30[a]) | 6.01 (0.26[a]) |
| $\bar{y}_{xre}$ | 6.94 (.61) | 6.70 (.59) | 6.99 (.30) | 6.24 (0.26) |
| $\bar{y}_{ar1}$ | 6.91 (.61) | 6.67 (.59) | 6.99 (.30) | 6.14 (0.26) |
| $\bar{y}_{reg}$ | 6.93 (.60) | 6.74 (.59) | 6.95 (.30) | 6.17 (0.26) |

Notes: [a] Variance computed assuming unequal variances across the poststrata, and the asymptotic normal reference distribution (method psvn in Section 5.4).

## 5.2. *Details of the simulations*

### 5.2.1. Simulated populations

The population counts and age structure were based on one of the ECA demographic groups, namely Non-Hispanic Females in East Los Angeles, collapsing population counts from four successive single age groups to yield 20 poststrata. The total population size was 12,575 and poststratum counts ranged from 47 to 982, with a mean of 629 and a standard deviation of 283. Population values were generated as follows:

**(A) Normal** $Y_{hi}$ normal as in Equations (1) and (2), with $H = 20$ poststrata and the following choices of $X, D$ and parameters:

(1) NULL: $X = (1, 1, \ldots, 1)^T, D = I, \beta = 0, \sigma_\mu^2 = 0, \sigma^2 = 1$
(2) XRE, Within Var = low: $X = (1, 1, \ldots, 1)^T, D = I, \beta = 0, \sigma_\mu^2 = 1/3, \sigma^2 = 1/3$;
(3) XRE, Within Var = high: as for (2) but with $\sigma^2 = 1$
(4) AR1, Within Var = low: $X$ and $D$ in Equation (5), $\rho = 0.9, \beta = 0, \sigma_\mu^2 = 1/3$, $\sigma^2 = 1/3$;
(5) AR1, Within Var = high: as for (4) but with $\sigma^2 = 1$
(6) REG, Within Var = low: $X$ and $D$ in Equation (4) with $W_h = h, \beta = (0, 0.2)$, $\sigma_\mu^2 = 1/3, \sigma^2 = 1/3$;
(7) REG, Within Var = high: as for (6) but with $\sigma^2 = 1$
(8) QUAD, Within Var = low : $X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ & \ldots & \ldots \\ 1 & H & H^2 \end{bmatrix}$

$D = I, \beta = (4.41, -0.84, 0.04), \sigma_\mu^2 = 1/3, \sigma^2 = 1/3$

(9) QUAD, Within Var = high: as for (8) but with $\sigma^2 = 1$.

**(B) Lognormal:** (10)–(18) as in (1)–(9) but with $Y_{hi}$ lognormally distributed. Specifically, $Y_{hi} = \mu_h + \sigma\sqrt{64/151}(\exp(Z_{hi}) - 13/8), Z_{hi} \sim N(0, 1)$. When $\sigma = 1$, $Y_{hi}$ has approximate mean $\mu_h$, variance 1.

5.2.2. Sampling mechanism

The total sample size in the simulations was fixed but sample sizes in the poststrata were random and hence vary over simulations. Sample sizes and mechanisms were as follows:

**Sample *n***: (1) Low: $n = 50$; (2) High: $n = 250$

(car) Missing completely at random: constant rates of selection across H poststrata
(mar) Missing at random: variable rates of selection across H poststrata. The logit of the probability of selection in poststratum $h$ is computed as $-9.3 - .07W_h$, yielding probabilities of selection such that the highest is 3.8 times the lowest. The selection probabilities are standardized to yield the desired sample size.

*5.3. Point estimation*

5.3.1. Methods

The following methods for estimating $\bar{Y}_+$ were compared in the simulation: (psm) Poststratified mean, with some collapsing to yield at least 5 cases in each poststratum. (cps) Collapsed poststratified mean. The 20 poststrata were collapsed into 5 by combining adjacent sets of four poststrata together. Any collapsed poststratum with less than 5 observations was combined with its neighbor, as with the original poststratified mean.

(reg) Maximum likelihood estimate for REG Model.
(ar1) Maximum likelihood estimate for AR1 Model.
(xre) Maximum likelihood estimate for XRE Model.
(unw) Unweighted Sample Mean.

The average root mean squared error of each of these methods was computed over the 50 population replicates of each simulation condition.

5.3.2. Results

An overview of the results is given in Table 3A, which presents average root mean squared errors for the six methods listed above by simulation population type and sample size. For ease of interpretation, the average root mean squared errors in each row are presented as per cent deviations from the corresponding average root mean squared error for psm. We note the following from this table:

(1) The method corresponding to the model used to generate the data has the lowest root mean squared error, for example reg is the best method for the REG populations and unw is the best method for the NULL populations. This result is expected given optimality properties of maximum likelihood when the assumed model holds.
(2) The unweighted mean unw does well for the NULL and XRE populations, but very

*Table 3. Average root mean squared errors of alternatives to the poststratified mean, expressed as percentage deviations from average root mean squared error of poststratified mean*

A) By population and sample size, averaged over sampling mechanism, error distribution and within variance

| Population ↓ | Alternative estimator cps | ar1 | reg | xre | unw | | cps | ar1 | reg | xre | unw |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NULL | −3 | −6 | −2 | −7 | −10 | ♦ | −2 | −5 | −2 | −5 | −5 |
| QUAD | 0 | −7 | 4 | −3 | 53 | ♦ | 15 | −2 | 2 | 0 | 186 |
| AR1 | 0 | −3 | −1 | 4 | 27 | ♦ | 1 | 0 | −1 | 2 | 57 |
| REG | −1 | −3 | −7 | 13 | 74 | ♦ | 12 | −1 | −1 | 5 | 230 |
| XRE | −2 | −7 | −4 | −8 | −4 | ♦ | 9 | −2 | −1 | −1 | 15 |
| Sample $n$ → | 50 | 50 | 50 | 50 | 50 | | 250 | 250 | 250 | 250 | 250 |

B) By population, sample size and sampling mechanism, averaged over error distribution and within variance

| Population ↓ | Alternative estimator ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NULL | −3 | −2 | −3 | ♦ | −8 | −2 | −11 | ♦ | −2 | −2 | −2 | ♦ | −7 | −3 | −8 |
| QUAD | −6 | 1 | 0 | ♦ | −8 | 7 | −5 | ♦ | −1 | 1 | 0 | ♦ | −3 | 4 | −1 |
| AR1 | −2 | 0 | 0 | ♦ | −3 | −1 | 8 | ♦ | −1 | 0 | 0 | ♦ | 0 | −1 | 5 |
| REG | −2 | −6 | 4 | ♦ | −4 | −7 | 21 | ♦ | 0 | −1 | 0 | ♦ | −1 | −2 | 10 |
| XRE | −5 | −4 | −6 | ♦ | −8 | −4 | −10 | ♦ | −1 | −1 | −1 | ♦ | −2 | −2 | −2 |
| Sample $n$ → | 50 | 50 | 50 | | 50 | 50 | 50 | | 250 | 250 | 250 | | 250 | 250 | 250 |
| Mechanism → | car | car | car | | mar | mar | mar | | car | car | car | | mar | mar | mar |

C) By population type, sample size and error distribution, averaged over sampling mechanism and within variance

| Population ↓ | Alternative estimator ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NULL | −8 | −5 | −9 | ♦ | −4 | 0 | −6 | ♦ | −5 | −3 | −6 | ♦ | 4 | −2 | −4 |
| QUAD | −14 | 3 | −8 | ♦ | −1 | 5 | 1 | ♦ | −3 | 3 | −2 | ♦ | −1 | 2 | 0 |
| AR1 | −5 | −3 | −5 | ♦ | −1 | 0 | 3 | ♦ | −1 | −1 | 1 | ♦ | 0 | −1 | 2 |
| REG | −4 | −11 | 16 | ♦ | −2 | −3 | 10 | ♦ | −1 | −2 | 8 | ♦ | −1 | −1 | 4 |
| XRE | −10 | −8 | −11 | ♦ | −4 | −1 | −6 | ♦ | −2 | −2 | −2 | ♦ | −1 | −1 | −1 |
| Sample $n$ → | 50 | 50 | 50 | | 50 | 50 | 50 | | 250 | 250 | 250 | | 250 | 250 | 250 |
| Error distribution → | nml | nml | nml | | lno | lno | lno | | nml | nml | nml | | lno | lno | lno |

D) By population type, sample size and within variance, averaged over sampling mechanism and error distribution

| Population ↓ | Alternative estimator ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre | | ar1 | reg | xre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NULL | −6 | −2 | −7 | ♦ | NA | NA | NA | ♦ | −5 | −2 | −5 | ♦ | NA | NA | NA |
| QUAD | −13 | 4 | −7 | ♦ | −3 | 4 | 0 | ♦ | −3 | 4 | −1 | ♦ | −1 | 2 | 0 |
| AR1 | −4 | −1 | 7 | ♦ | −2 | −1 | 3 | ♦ | 0 | 0 | 2 | ♦ | 0 | −1 | 2 |
| REG | −6 | −10 | 15 | ♦ | −1 | −4 | 11 | ♦ | −1 | −2 | 6 | ♦ | −1 | −1 | 5 |
| XRE | −9 | −7 | −10 | ♦ | −5 | −2 | −7 | ♦ | −2 | −2 | −2 | ♦ | −1 | −1 | −1 |
| Sample $n$ → | 50 | 50 | 50 | | 50 | 50 | 50 | | 250 | 250 | 250 | | 250 | 250 | 250 |
| Within variance → | low | low | low | | hi | hi | hi | | low | low | low | | hi | hi | hi |

poorly for populations with more structure, such as those generated by the REG, QUAD and AR1 populations, when the method is seriously biased, particularly under MAR selection and the larger sample size. This method is clearly not useful for routine use. The ar1, reg and xre methods are all consistent and behave like post-stratification when the sample size is large, and they are much better than unw for the structured populations.

(3) The ar1 method is the best overall in terms of root mean squared error. The ar1 and reg methods display advantages over xre for the QUAD, AR1, and REG populations, reflecting gains from including the added parameter in the models that underlie these methods. Comparing ar1 and reg, ar1 has a noticeable advantage for the QUAD population, when the mean structure differs from that assumed by the models that underlie either method. The ar1 method yields small but consistent reductions of root mean squared error over psm for all the populations. In contrast, the collapsing method cps yields smaller reductions at the low sample size and is worse than psm for the large sample sizes, except for the null model. The advantages of ar1 and reg over the psm are considerably greater for $n = 50$ than for $n = 250$, as one would expect since the gains of smoothing dissipate as the sample size increases, and ar1, reg, and xre all converge towards the psm method.

Analyses of variance of these summaries were conducted to discover the main sources of variation over simulation conditions, and these were used to form summary Tables 3B, 3C and 3D, which retain subsets of the simulation factors. In these more detailed tables we restrict attention to root mean squared error comparisons of ar1, reg, and xre with psm. Table 3B indicates that advantages of the modeling methods are greater for mar than for car selection, as might be expected. In Table 3C it can be seen that gains from the model methods are reduced for the lognormal populations, reflecting the lack of normality. However, the root mean squared errors of ar1 and reg are quite robust to this form of misspecification. Table 3D show predictable gains of the model-based methods when the within-group variance is large, since in this circumstance smoothing of the poststratum weights is relatively advantageous.

## *5.4. Interval estimation methods*

### 5.4.1. Methods

Methods of interval estimation involve other choices concerning conditional versus unconditional approaches to estimating standard errors, assumptions about the variance and $t$-type modifications of the reference distribution. We compared twenty methods in our simulations, of which ten are discussed here:

(A) Intervals centered at the poststratified mean:

(pscn): the variance of the poststratified mean (ps) is computed by pooling across strata, assuming a constant variance across poststrata (c) and using the asymptotic normal reference distribution (n).

(psvn): as for pscn, but assuming a variance that varies across the poststrata (v). A distinct variance is estimated for each poststratum, rather than pooling.

(psct): as for pscn, but with a $t$-like correction for estimating the variance (t). The normal reference distribution was retained, but the variance was inflated by $(n-1)/(n-3)$ to reflect the increased uncertainty from estimating the pooled variance (e.g., Scott and Smith 1971). For a Bayesian justification of the adjustment see Little (1993). Although not strictly a $t$ inference, a very similar interval would be obtained by replacing the normal reference distribution by the $t$ with $n-1$ degrees of freedom.

(psvt): as psvn, but with a t correction for estimating the variances within each poststratum. As before, a normal reference distribution was used, but the contribution of the variance from poststratum $h$ was inflated by $(n_h-1)/(n_h-3)$ to reflect increased uncertainty from estimating the within-stratum variance. A Bayesian justification of this adjustment is given in Little (1993).

(B) Intervals centered on the collapsed poststratified mean:
(cpsct): as psct, but applied to the poststratified mean after collapsing.
(cpsvt): as psvt, but applied to the poststratified mean after collapsing.

(C) Intervals centered on model estimates (ar1, reg, xre). The variances for each of these methods were computed via Equation (6). The reference distribution was Student's $t$ with degrees of freedom equal to the number of poststrata with observations minus the number of parameters estimated, including fixed-effects terms and $\rho$. For example, if 16 poststrata had observations then degrees of freedom are $16-2=14$ for REG and AR1 and $16-1=15$ for XRE. An exact t correction is not available for this model, but the proposed choice of degrees of freedom was expected to improve the asymptotic normal intervals, which treat the covariance matrix as effectively known.

(D) Interval centered on the unweighted mean (unw): The standard normal interval was computed for the unweighted mean, with the sample variance computed collapsing over the poststrata.

#### 5.4.2. Results

Table 4A shows the coverage rates of these ten methods by population and sample size, averaged over mechanism, distribution and within variance. The coverages are expressed as deviations from the nominal 95% level, so for example the first entry of 2.3 in Table 4A means an empirical coverage of 97.3%. From this table, the uncorrected normal intervals for the poststratified mean, pscn and psvn, undercover slightly for $n=250$ and markedly for $n=50$. The corrected intervals, psct and psvt, are somewhat conservative, and fairly similar for these simulations. The collapsed methods, cpsct and cpsvt yield slightly conservative intervals. Of the model-based methods ar1, reg, xre, and unw, ar1 has the best coverage over all problems, although its coverage is slightly liberal; reg seriously undercovers for the QUAD population and xre seriously undercovers for the REG and QUAD populations. The intervals based on the unweighted mean have poor coverage in the QUAD, AR1 and REG populations, particularly for the large sample size where bias predominates over variance.

Tables 4B and 4C compares coverage of the psvt, ar1, reg, and xre intervals in a less aggregated fashion, for sample size $n=50$. From Table 4B, it can be seen that the undercoverages of reg and xre under model misspecification are greater for the mar than for the car mechanism. From Table 4C, it is clear that undercoverage prevails in the lognormal populations. Thus modifications of these models for non-normality might be expected to improve their performance in such settings.

Table 5 summarizes the average width of the intervals from the ten methods, expressed as percentage deviation of the width of the psvt, chosen as the poststratified method of choice for comparisons. The methods pscn and psvn have intervals that are 20% narrower on average for $n = 50$ and 10% narrower on average for $n = 250$, but this is achieved at the expense of undercoverage. The intervals based on the collapsed poststratified mean are

*Table 4. Coverage rates for interval estimates, expressed as deviations from 95% nominal coverage*

A) By population type and sample size, averaged over mechanism, distribution, and within variance

| Population ↓ | Interval estimate psct | pscn | psvt | psvn | cpsct | cpsvt | ar1 | reg | xre | unw |
|---|---|---|---|---|---|---|---|---|---|---|
| NULL | 2.3 | −2.7 | 1.0 | −3.2 | 0.6 | 0.1 | −0.5 | −1.1 | −0.7 | −2.0 |
| QUAD | 1.1 | −4.0 | 1.1 | −3.1 | 1.6 | 1.4 | −0.4 | −3.2 | −0.1 | −7.2 |
| AR1 | 1.9 | −3.6 | 2.0 | −2.7 | 0.0 | 0.0 | −1.3 | −1.1 | −2.4 | −6.4 |
| REG | 0.6 | −1.9 | 0.6 | −1.3 | 1.1 | 1.0 | −1.3 | −0.1 | −5.1 | −17.4 |
| XRE | 2.2 | −3.4 | 2.2 | −2.2 | 1.4 | 1.4 | −0.7 | −0.1 | −0.6 | −1.7 |
| Sample $n$ → | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| NULL | 0.9 | −1.5 | 0.5 | −1.4 | −0.6 | −1.0 | −0.2 | −0.1 | −0.3 | −1.5 |
| QUAD | 0.5 | −2.3 | 0.6 | −2.0 | 0.6 | 0.1 | −0.1 | −1.7 | −0.8 | −32.4 |
| AR1 | 0.9 | −1.4 | 0.8 | −1.3 | 0.4 | 0.0 | 0.0 | 0.3 | −0.4 | −12.6 |
| REG | 1.0 | −0.5 | 0.9 | 0.0 | −0.3 | −0.9 | −0.3 | 0.3 | −2.2 | −43.1 |
| XRE | 0.9 | −1.5 | 0.7 | −1.8 | 0.2 | −0.1 | 0.2 | 0.3 | 0.1 | −3.0 |
| Sample $n$ → | 250 | 250 | 250 | 250 | 250 | 250 | 250 | 250 | 250 | 250 |

B) By population and mechanism, averaged over distribution and within variance, for $n = 50$

| Population ↓ | Interval estimate psvt | ar1 | reg | xre | | psvt | ar1 | reg | xre |
|---|---|---|---|---|---|---|---|---|---|
| NULL | 1.3 | −0.4 | −0.5 | −0.5 | ♦ | 0.7 | −0.6 | −1.8 | −0.9 |
| QUAD | 1.9 | −0.1 | −1.7 | −0.6 | ♦ | 0.4 | −0.7 | −4.6 | 0.5 |
| AR1 | 2.0 | −0.7 | −0.4 | −0.9 | ♦ | 1.9 | −1.9 | −1.7 | −3.9 |
| REG | 1.2 | −0.6 | 0.2 | −1.2 | ♦ | −0.1 | −2.0 | −0.5 | −9.1 |
| XRE | 2.4 | 0.0 | 0.2 | 0.1 | ♦ | 1.9 | −1.5 | −0.5 | −1.2 |
| Sample $n$ → | 50 | 50 | 50 | 50 | | 50 | 50 | 50 | 50 |
| Mechanism → | car | car | car | car | | mar | mar | mar | mar |

C) By population and distribution, averaged over mechanism and within variance, for $n = 50$

| Population ↓ | Interval estimate psvt | ar1 | reg | xre | | psvt | ar1 | reg | xre |
|---|---|---|---|---|---|---|---|---|---|
| NULL | 1.9 | 1.0 | 0.9 | 1.0 | ♦ | −0.9 | −1.5 | −2.0 | −1.7 |
| QUAD | 2.6 | 1.4 | −2.0 | 1.4 | ♦ | −0.3 | −2.2 | −4.4 | −1.5 |
| AR1 | 3.5 | 0.6 | 0.6 | −0.4 | ♦ | 0.4 | −3.2 | −2.7 | −4.5 |
| REG | 0.0 | 0.2 | 1.2 | −2.7 | ♦ | 1.1 | −2.9 | −1.5 | −7.5 |
| XRE | 3.2 | 0.6 | 1.2 | 0.8 | ♦ | 1.2 | −2.0 | −1.5 | −1.9 |
| Sample $n$ → | 50 | 50 | 50 | 50 | | 50 | 50 | 50 | 50 |
| Distribution → | nor | nor | nor | nor | | lno | lno | lno | lno |

*Table 5. Average width of 95% interval, expressed as percentage deviations from average width of 95% interval from PSVT method, by population type and sample size, averaged over mechanism, distribution, and within variance*

| Population | Interval estimate | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ↓ | psct | pscn | psvn | cpsct | cpsvt | ar1 | reg | xre | unw |
| NULL | 0 | −20 | −20 | −12 | −12 | −22 | −19 | −22 | −27 |
| QUAD | −4 | −24 | −20 | −7 | −7 | −24 | −18 | −18 | 2 |
| AR1 | −1 | −22 | −20 | −11 | −10 | −20 | −18 | −18 | −15 |
| REG | −3 | −23 | −20 | −10 | −9 | −21 | −23 | −14 | 0 |
| XRE | −2 | −23 | −20 | −10 | −10 | −24 | −21 | −24 | −23 |
| Sample $n$ → | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 | 50 |
| NULL | 1 | −9 | −10 | −9 | −10 | −9 | −6 | −9 | −15 |
| QUAD | 0 | −11 | −11 | 9 | 7 | −5 | −4 | −5 | 41 |
| AR1 | 0 | −10 | −10 | −3 | −4 | −5 | −4 | −5 | 0 |
| REG | 0 | −10 | −10 | 4 | 3 | −4 | −5 | −4 | 30 |
| XRE | 0 | −10 | −10 | 2 | 1 | −5 | −4 | −6 | −2 |
| Sample $n$ → | 250 | 250 | 250 | 250 | 250 | 250 | 250 | 250 | 250 |

about 10% narrower for $n = 50$, suggesting a benefit from collapsing at this sample size that is less evident from the root mean squared errors in Table 3. At $n = 250$ the reductions in interval width from collapsing have disappeared. The ar1, reg, and xre methods achieve a reduction in width of around 20% for $n = 50$ and around 6% for $n = 250$. More detailed classifications by the simulation factors, omitted here, indicate that these reductions are fairly stable over simulation conditions. Thus ar1, which achieves these reductions with good coverage (at least in the normal populations) is seen as a useful generalization over xre and an attractive alternative to collapsing for small samples, when some smoothing of the poststratification weights is desired.

## 6. Discussion

This article has examined methods for an ordinal stratifier that modify the poststratified mean by imposing random-effects models on the poststratum mean differences. Other model based weight smoothing strategies use fixed-effects regression models for more than one continuous or categorical poststratifier, for equal and unequal probability sample designs (Huang and Fuller 1978; Bardsley and Chambers 1984; Bethlehem and Keller 1987). The methods discussed in these articles reduce to the poststratified mean for the case of a single categorical stratifier and equal probability sampling that we consider here. Like many of the methods in those articles, the methods described here yield design-consistent estimates and hence are somewhat protected against model misspecification. The REG and AR1 models for the poststratum means appear reasonable generalizations of the previously-considered XRE model. Smoothing based on a model like REG and AR1 seems less arbitrary than procedures that truncate the weights or collapse the poststrata. A disadvantage over global weight smoothing strategies (e.g., Bardsley and Chambers 1984; Little 1993) is that our models are specific to survey outcomes and hence in effect provide different weights for each survey outcome. Thus they are less convenient and involve more computation than methods that provide a single smoothed weight for all outcomes. But computation is less of an issue in the era of high-speed computers, and ML estimation for random-effects models such as (1) and (2) is becoming more accessible in statistical software. Any procedure directed at reducing variance must tailor the weights, depending on the degree of association of the poststratifier with the outcome.

One approach to allowing the uncertainty in estimating variance parameters is to apply Bayesian methods, although for our model this involves some additional computational complexity. We applied a simple $t$-based correction to account for uncertainty due to estimation of the dispersion parameters, as in Scott and Smith (1971). This did not appear to seriously distort coverages in the simulation study.

Our methods are model based, raising the important issue of performance when the model is misspecified. The effect of model misspecification is limited here since the estimates converge to the poststratified mean as the sample size increases, and hence are design consistent. In small samples the models yield optimal shrinkage methods when correctly specified, but the model estimates may still dominate *ad hoc* alternative methods even when the model is misspecified. Indeed the model-based estimates proved quite robust to misspecification in the simulation study, which included populations with heteroscedasticity and non-normality. All the methods discussed here assumed ignorability of the selection mechanism given the poststratifier. Methods based on nonignorable models can be developed, but are hard to estimate since information on systematic differences between selected and non-selected cases within poststrata is rarely available.

Although both the REG and AR1 models were better than the XRE model in our simulations, AR1 appeared a slightly better model for general use, since it was nearly as good as REG when the REG model was correct, and provided better inferences when the post-stratum means deviated from linearity. The AR1 covariance structure in (5) leads to local smoothing of the poststratum weights, which seems intuitively sensible for an ordinal post-stratifier. The model is applied here to a single equally-spaced ordinal poststratifier, but other categorical or continuous poststratifiers could be modeled by including them as regressors in the mean structure of (5), or through further elaboration of the covariance structure.

## 7. References

Bardsley, P. and Chambers, R.L. (1984). Multipurpose Estimation from Unbalanced Samples. Applied Statistics, 33, 290–299.

Bethlehem, J.G. and Keller, W.J. (1987). Linear Weighting of Sample Survey Data. Journal of Official Statistics, 3, 141–153.

Eaton, W.W. and Kessler, L.G. (1985, eds.). Epidemiologic Field Methods in Psychiatry: The NIMH Epidemiologic Catchment Area Program. New York: Academic Press.

Ghosh, M. and Meeden, G. (1986). Empirical Bayes Estimation of Means from Stratified Samples. Journal of the American Statistical Association, 81, 1058–1062.

Harville, D.A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. Journal of the American Statistical Association, 72, 320–340 (with Discussion).

Holt, D. and Smith, T.M.F. (1979). Post Stratification. Journal of the Royal Statistical Society, Series A, 142, 33–46.

Huang, E.T. and Fuller, W.A. (1978). Nonnegative Regression Estimator for Sample Survey Data. Proceedings of the American Statistical Association, Social Statistics Section, 300–305.

Jennrich, R.I. and Schluchter, M.D. (1986). Incomplete Repeated-Measures Models with Structured Covariance Matrices. Biometrics, 42, 805–820.

Kalton, G. and Maligalig, D.S. (1991). A Comparison of Methods of Weighting Adjustment for Nonresponse. Proceedings of the 1991 Annual Research Conference, 409–428. U.S. Bureau of the Census, Arlington, VA.

Laird, N.M. and Ware, J.H. (1982). Random-Effects Models for Longitudinal Data. Biometrics, 38, 963–974.

Little, R.J.A. (1983). Estimating a Finite Population Mean from Unequal Probability Samples. Journal of the American Statistical Association, 78, 596–604.

Little, R.J.A. (1991). Inference with Survey Weights. Journal of Official Statistics, 7, 405–424.

Little, R.J.A. (1993). Post-stratification: A Modeler's Perspective. Journal of the American Statistical Association, 88, 1001–1012.

Morris, C.N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. Journal of the American Statistical Association, 78, 47–55.

Pfefferman, D. and Nathan, G. (1981). Regression Analysis of Data from a Cluster Sample. Journal of the American Statistical Association, 76, 681–689.

Potter, F. (1990). A Study of Procedures to Identify and Trim Extreme Sampling Weights. Proceedings of the American Statistical Association, Survey Research Methods Section, 225–230.

Scott, A. and Smith, T.M.F. (1969). Estimation in Multistage Sampling. Journal of the American Statistical Association, 64, 830–840.

Scott, A. and Smith, T.M.F. (1971) Interval Estimates for Linear Combinations of Means. Applied Statistics, 20, 276–285.

Tremblay, V. (1986). Practical Criteria for Definition of Weighting Classes. Survey Methodology, 12, 85–97.

Received April 1996
Revised May 1997