MODEL-ASSISTED WEIGHTING FOR SURVEYS WITH MULTIPLE RESPONSE MODE

Eric V. Slud, Census Bureau & Univ. of Maryland
Mathematics Department, University of Maryland, College Park MD 20742

Key words: American Community Survey, demographic weights, logistic regression, mean-squared error, misspecified model, nonresponse adjustment.

Abstract. Several large surveys administered by the Census allow responses in several different modes (mail, telephone, personal interview), in some cases — as in the American Community Survey (ACS) — after sub-sampling from those sample units which have not responded at any earlier stage. Analysis of the survey results usually proceeds by weighting responses roughly by inverse unconditional probability of response to the survey. But often, nonresponse in the different modes can be modelled in terms of demographic and geographic variables (such as dwelling type and aggregated characteristics related to ethnicity and socioeconomic status), and the results of such models derived from previous recent surveys could be used to improve the estimation of population totals and domain subtotals which is ordinarily the goal of large surveys. This paper applies concepts of sample-survey theory to investigate the theoretical improvements possible with both correctly and incorrectly specified models, and illustrates the issues and improvements using data from the 1990 decennial census. In the setting of models of nonresponse no worse than those fitted to one state's decennial-census data and applied to another similar state, the mean-squared error is definitely improved by incorporating the model into sample-weighted estimators, through an adjustment factor which is constant across PSU (within state).

1. INTRODUCTION & PROBLEM

Suppose that individuals in a sampling frame U are selected for inclusion in a large survey S with single and joint inclusion probabilities respectively π_i, π_{ij}, and that each individual can respond to the survey in a succession of K possible modes, with mode 1 being the most direct, such as mail-response in the decennial census or American Community Survey. (In the ACS, the modes are: Mail-, Telephone-, and Personal-Interview response. In the decennial census, the modes are Mail- and Interviewer- response, but the census data can be analyzed further, as in Slud (1998, 1999, 2001) by treating as different modes the responses to interviewers within successive quantile-intervals of interviewer followup time within ARA.) Of those individuals selected for initial inclusion and not responding in any of modes $1, \ldots, k$, where $k = 1, \ldots, K-1$, a fraction a_k are randomly sub-sampled for attempted enumeration under mode $k+1$. For each individual $i \in U$, denote by J_{ik} the indicator that individual i would, if selected and followed up, respond to the survey in mode k and no earlier mode. Then $\sum_{k=1}^{K} J_{ik} = J_i$ is the indicator that individual i responds in any mode to the survey.

Suppose that each individual in the sampling frame comes equipped with a vector $\mathbf{X}_i$ of predictor variables for response, which are observable or known in advance of actual enumeration. Such predictors would include geographic area, along with variables such as housing type which would be known from a master address list and aggregates from recent previous surveys of demographic characteristics for the neighborhood (such as census block-group or tract) containing the individual. A list of such predictors based on block-group aggregated census long-form data is provided in Slud (1998). Assume that the conditional probabilities h_{ik} of survey response by individual i in mode k, *given nonresponse in each of modes* $1, \ldots, k-1$, have been modelled and estimated as parametric functions $h_k(\mathbf{X}_i) = h_k(\mathbf{X}_i, \beta^{(k)})$ of the predictor variables, with the estimated parameters (e.g., regression coefficients in generalized linear models) denoted $\beta^{(k)}$. The model implemented below in analysis of 1990 decennial census data, following Slud (1998, 1999, 2001), is the logistic regression model

$$h_k(\mathbf{x}, \beta) = e^{\mathbf{x}\cdot\beta} / (1 + e^{\mathbf{x}\cdot\beta})$$

The predictors $\mathbf{X}_i$ are mostly the same across different response modes, except that, as in Slud (1999, 2001), block-group aggregated rates of response by earlier modes than k are used in modelling the rate of response by mode k. Thus, based on predictors $\mathbf{X}_i$, the modelled probabilities of response in the k'th mode, with (estimated) model-coefficients $\beta^{(k)}$, are

$$h_{ik} \;\equiv\; P(J_{ik}=1|\mathbf{X}_i,\, J_{ij}=0,\, j<k) \;=\; \frac{e^{\mathbf{X}_i\cdot\beta^{(k)}}}{1+e^{\mathbf{X}_i\cdot\beta^{(k)}}}$$

Assume that the parametric model is a fixed-effect model only. Then prospectively, before sampling,

$$p^0_{ik} \;=\; \prod_{1\le j<k} (1-h_{ij}) \;\cdot\; h_{ik}$$

gives the probability with which individual i would respond to the survey in mode k and not earlier, if selected for inclusion and followed up that far. Taking into account the probabilities with which nonrespondents at each stage are included in later stages of followup, we obtain the probabilities $P(J_{ik}=1) \;=$

$$p_{ik} \;=\; \prod_{j=1}^{k-1} \{(1-h_{ij})\, a_j\}\; h_{ik} \;=\; p^0_{ik} \prod_{1\le j<k} a_j$$

with which individuals assumed to be included in the initial sample are sampled up to and respond within the k'th response mode. Note that as a practical matter, models and estimates for the final-stage response probabilities h_{iK} are largely speculative, because they reflect rates of interview-refusal and omission (e.g., because of failure of interviewers to make personal contact or find proxies) concerning which there is no direct data. Only a followup or post-enumeration study could give more than a hypothetical cast to estimates of these probabilities. However, if these rates are understood reflect response by mode for units on a master address-list (like the MAF in the decennial census), then even the final rates h_{iK} could be meaningfully estimated from survey data and used in weighting.

The data recorded from the survey will be, for each individual $i \in S$, the response indicator vector $(J_{ik},\; k=1,\ldots,K)$ together with a label A_i for the latest mode under which individual i is selected for followup. In case $J_i=1$, the label A_i is equal to that mode k for which $J_{ik}=1$, and an attribute value y_i (such as number in household, or total household income) is also recorded.

The survey data are to be used to estimate the frame population total $t=\sum_U y_i$ for the attribute in question, and the estimators to be considered all take the form of a weighted linear combination $\hat{t}_w \;=\; \sum_S \sum_{k=1}^K w_{ik}\, J_{ik}\, y_i$. The objective of the present research is to investigate what the optimal weights w_{ik} are, from the vantage point of minimum mean-squared error; how much difference it makes to use them by comparison with the weights one would use if the probabilities p_{ik} were completely unknown, and how sensitive the weights are to correct specification of the model $h_k(\mathbf{X}_i,\, \beta^{(k)})$.

Remark 1 *The related problem of optimally designing sub-sampling rates a_k, without allowing them to depend on location and neighborhood demographic characteristics, has been studied in Elliott, Little & Lewitzky (2000).* □

A primary motivation for the work described here was to investigate the desirability of demographic-model-based weighting adjustments to surveys like the ACS. In particular, our interest centers on nonresponse weighting adjustments derived from predictive models. However, it is in the nature of nonresponse that (without additional followup or calibration sampling studies) **design-unbiased estimation is unattainable**. For a description of the existing approach to ACS weighting — which is done essentially through constant (within-state) multiplicative factors for subsampling, seasonal response-rates by mode, and *noninterview-rate* adjustments — see Dahl (1998) and Adeshiyan (1998).

The problem of nonresponse and weighting adjustments is generic for sample surveys (see Groves et al. 1999). We focus here on aspects peculiar to surveys with multiple response modes and a subsampling design like the ACS, under which models for nonresponse would naturally be built for each response mode.

2. MSE FORMULAS, OPTIMAL WEIGHTS

We begin by obtaining formulas for the bias and variance of the statistic $\hat{t}_w$, and by reducing the set of weights $\{w_{ik}\}$ to consider. First, recalling that $E\,J_{ik} \;=\; p_{ik}$ by definition, we find

$$E(\hat{t}_w) \;=\; E\Big(\sum_{i\in U}\sum_k I_{[i\in S]}\, y_i\, w_{ik}\, J_{ik}\Big) \;=\; \sum_{i\in U}\sum_k \pi_i\, w_{ik}\, p_{ik}\, y_i$$

Thus the bias in estimating $\sum_U y_i$ by $\hat{t}_w$ is

$$\sum_{i\in U} y_i \left\{\pi_i \sum_k w_{ik}\, p_{ik} \;-\; 1\right\}$$

If the quantities p_{ik} were known or accurately estimated, then for the estimator $\hat{t}_w$ to be approximately unbiased for all possible attribute-values y_i, the weights w_{ik} would have to satisfy

$$\pi_i \, \Sigma_{k=1}^K \, w_{ik} \, p_{ik} \;=\; 1 \tag{1}$$

Denote the actual probability of *non*response to the survey, for an included (sampled) unit i, to be

$$q_i \;\equiv\; 1 \,-\, \Sigma_{k=1}^K \, p_{ik} \tag{2}$$

Remark 2 *The unbiasedness condition (1) cannot generally be satisfied, even approximately, in model-free fashion. It would force weights w_{ik} to vary with different i whenever the quantities $\pi_i(1-q_i)$ do. Since the q_i are not known, approximating them involves models and estimates.* □

Straightforward algebraic manipulation of the expression (bias squared plus variance) for the mean-squared error (MSE) of $\hat{t}_w$ yields a form from which it is easy to conclude that **the MSE is made as small as possible by equalizing the allowed weights w_{ik} over all response-modes k for each individual** i, specifically by putting

$$w_{ik} \;=\; \tilde{w}_i \;=\; (\pi_i \, \Sigma_{k=1}^K \, p_{ik})^{-1} \;=\; (\pi_i \,(1-\tilde{q}_i))^{-1} \tag{3}$$

for some set of values $\tilde{q}_i$. (Details of this verification, as well as of the following MSE formula (4), can be obtained in an unpublished report Slud 2002.) Assume from now on that $w_{ik} = \tilde{w}_i$ are defined as in (3), not varying over response-mode, and define a modified 'attribute':

$$\tilde{y}_i \;=\; y_i \,(1-q_i)\,/\,(1-\tilde{q}_i)$$

Then the formulas for bias and variance of $\widehat{t}_w$ lead to the general formula: $\mathrm{MSE}(\hat{t}_{\tilde{w}}) \;=$

$$\Big(\sum_U \,(y_i - \tilde{y}_i)\Big)^2 \,+\, \mathrm{Var}(\hat{t}_{\pi,\tilde{y}}) \,+\sum_U \frac{\tilde{y}_i^2 \, q_i}{\pi_i\,(1-q_i)} \tag{4}$$

where following Särndal et al. (1997), we denote by $\widehat{t}_{\pi,\tilde{y}}$ the ordinary Horvitz-Thompson estimator (for the population total of the attributes y based upon a single response-mode) with the same inclusion-probabilities π_i, π_{ij} as above, for which the theoretical variance is

$$V(\widehat{t}_{\pi,\tilde{y}}) \;=\; \sum_{i,\,j\in U} \tilde{y}_i \, \tilde{y}_j \, \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j} \tag{5}$$

Note that the comparison, across different choices $\tilde{q}_i$, of variances and MSE's depends not only upon the initial sampling fraction and the population coefficient of variation of the attribute y, but also on population characteristics relating y_i to q_i.

In the setting of *balanced random sampling without replacement*, the variance formula (5) specializes in a well-known way. Denote by $\pi_i \equiv \pi$ the probability with which each individual i is included in the initial sample, and by $s^2_{\tilde{y}U} \;=\; \frac{1}{|U|-1} \sum_U (\tilde{y}_i - \bar{\tilde{y}}_U)^2$ the frame-population modified-attribute variance. Then the term (5) becomes

$$|U| \; (1/\pi - 1) \; s^2_{\tilde{y}U} \tag{6}$$

Several possible choices of $\tilde{q}_i$ are of particular interest in defining weights $\tilde{w}_i$ as in (3), and we label these choices for later comparison. First, denote as choice (a) the *correct* demographic weights using $\tilde{q}_i = q_i$ of (2). Choice (b) is the population-wide nonresponse rate

$$q^* \;=\; 1 \,-\, |U|^{-1} \, \Sigma_U \Sigma_{k\le K} \, p_{ik} \tag{7}$$

This may be the most obvious choice for $\tilde{q}_i$ to be constant over i, but it is not the best. In fact, we will find below that bias-squared, the first of the three terms on the right-hand side of (4), is usually the dominant term in MSE. Thus the best constant to use should be approximately the one which zeroes out the bias, that is, which satisfies the equation

$$\Sigma_{i\in U} \, y_i \,(q_i - \tilde{q}) \;\;=\;\; 0$$

This value, which we define as choice (c) for $\tilde{q}_i$, is

$$\tilde{q}_{opt} \;=\; \Sigma_{i\in U} \, y_i \, q_i \,/\, \Sigma_{i\in U} \, y_i \tag{8}$$

So far, these choices (a)–(c) provide useful MSE benchmarks but *are not feasible*, in the sense that they describe estimators which depend on the unknown values (2). The three corresponding choices based upon empirically modelled and estimated nonresponse probabilities $\hat{p}_{ik}$ are (d), (e), and (f). Specifically, (d) is the choice $\tilde{q}_i = \hat{q}_i \equiv 1 - \Sigma_k \, \hat{p}_{ik}$; (e) is the population-averaged version

$$\hat{q}_{Av} \;=\; 1 \,-\, |U|^{-1} \, \Sigma_U \Sigma_{k\le K} \, p_{ik} \tag{9}$$

and (f) is the choice of $\tilde{q}_i$ as the set of weighted-average estimated nonresponses

$$\tilde{q}_{mod} \;=\; \Sigma_{i\in U} \, y_i \, \hat{q}_i \,/\, \Sigma_{i\in U} \, y_i \tag{10}$$

3. DATA ANALYSIS FOR 1990 CENSUS

We first specialize the comparison between MSE's (4) to a case most relevant to the ACS, in which the attribute y_i of interest can be regarded as the indicator of a valid household enumeration for the index i on the Master Address File (MAF) if household

i were followed up without time-limitation. (Different, but also relevant, choices would be to view y_i as the number of valid enumerated persons within each household, or the household income or indicator of poverty or of participation in a program like Food Stamps, but we do not pursue these possibilities for now.) Then the population total $\sum_U y_i$ is be interpreted as the true count of households, and (2) is the probability within a sample survey like ACS that a household on the MAF would *not* be enumerated up to and including the final (K'th) response mode. Although undercount and refusal-rates are usually small — of the order of one percent up to a few percent — it may still happen that differences among q_i for addresses i with different demographic and geographical characteristics can differ by sizeable factors. When this is true, there is some hope that model-assisted weighting can improve the accuracy of estimates from the survey.

The data used in the present comparison are the 1990 decennial-census files previously used by Slud (1998, 1999, 2001) in modelling response to the census, by mail or by later 'modes' of personal response to followup enumerators within successive quantile intervals of followup time within ARA. These files include tallies by block-group of the numbers (of HU's) responding by mail, before the 50^{th} percentile followup time within the ARA, between the 50^{th} and 75^{th}, between the 75^{th} and 90^{th}, and after the 90^{th}. In addition, the files contain geographic and demographic information on the block-group, plus the one variable housing-type (*htyp*) which referred to individual HU's (and was known from the address-file, before enumeration). These predictive variables were used in Slud (1998, 1999, 2001) to fit logistic models for response-rates by state, with numbers of predictors ranging from about 20 to about 70, depending on the size of the state and the mode of response. (The greater numbers of variables arose, from BIC-like penalized-deviance model-fitting, in models in large states either for mail-response or for response before the 90^{th} percentile of enumerator check-in times, among HU's which had not responded by the 75^{th} percentile.) The fitted models are used here as the models $h_k(\mathbf{X}_i, \beta)$ specifying conditional response probabilities h_{ik} for individual HU's in the i'th *htyp*-by-block-group stratum within a state to respond in mode k, given that it had not responded earlier. (Then the estimated unconditional response probabilities $\hat{p}_{ik}$ are derived from h_{ik} and a_j specified below.) The response-modes chosen for illustration are $k = 1$ for Mail-response, and then for HU's which did *not* respond by mail, $k = 2$ for response before the 50^{th} percentile of ARA check-in time, $k = 3$ for response between the 50^{th} and 75^{th} percentiles of check-in time, and $k = 4$ for response between the 75^{th} and 90^{th} percentiles. For present purposes, we treat the HU's which did not respond by the 90^{th} percentile of check-in times as though they did not respond at all.

To mimic the sub-sampling scenario of ACS, we define *and fix throughout the rest of this paper* an inclusion probability of $\pi = 1/2$ followed by sub-selection probabilities $a_k = 0.7$ for later stages $k = 1, 2, 3$ among HU's which have not responded in mode $\leq k$. (In ACS, all mail nonrespondents are followed up for telephone (CATI) response, and CATI nonrespondents are subsampled with probability 1/3 for further (CAPI) follow-up. The overall subsampling fraction of 1/3 is what we imitate by choice of $a_j = 0.7,\ j = 1, 2, 3,$ since $(0.7)^3 = 0.343 \approx 1/3$.) We compare the elements and total of the MSE formula (4), for this choice of subsampling probabilities and inclusion-rate $\pi = 0.5$ with simple random sampling, for several states and several choices of weights based either on a 'correct' model (ie one fitted to the same state for the same census data), no model, or on an incorrect model, such as one based on a neighboring state. Recall that the attribute y_i of primary interest here is the indicator of response (in mode $k \leq 4$). We display for each state, model, and choice of $\tilde{q}_i$ used in weights $\tilde{w}_i$, the following quantities: *Bias* as given by the expression inside the parentheses of the first term in (4), the total *MSE* as given by (4), and also the root-MSE divided by the actual number of nonresponders, denoted *rMSE/Nrsp*. Always $MSE = Bias^2 + Var1 + Var2$, and *rMSE/Nrsp* $= \sqrt{MSE}/\sum_U(1 - y_i)$. Quality of estimation is assessed using *rMSE/Nrsp*, just as coefficients of variation are used in survey practice. Values of .01 to .03 are typical of state models which fit well, and more generally values up to .1 allow conservative confidence intervals with value 20% of the nonresponse numbers being estimated.

Consider first Delaware (DE), which had 219509 HU addresses in 1990 within the 727 *htyp* by block-group strata containing at least 21 HU's. In DE, 10662 HU's did not respond by the 90^{th} percentile of their ARA's check-in times. Estimation error in the count $\sum_{i \in U} y_i$ of responders is likely to range up to 2000, we expect MSE to range up to $4.e6$.

For the six plans (a)-(f) of choosing $\tilde{q}_i$ in (3), the MSE results on the DE data (using the DE model) are displayed in Table 1. For comparison, the Table provides analogous MSE's for MD and MN, which respectively had 84883 and 35321 non-

Table 1. MSE components for DE, MD and MN data of six plans (a)-(f) for weights $\tilde{w}_i$ in (3) based on choices for $\tilde{q}_i$. Model-based plans (d)-(f) used models fitted to same-state data.

State	Plan		Bias	MSE	**rMSE/Nrsp**
DE					
	(a)	q_i	0	85860	0.028
	(b)	q^*	1001	1080454	0.098
	(c)	$\tilde{q}_{opt}$	0	78358	0.026
	(d)	$\hat{q}_i$	158	108391	0.031
	(e)	$\hat{q}_{Av}$	699	567868	0.071
	(f)	$\tilde{q}_{mod}$	242	137074	0.035
MD					
	(a)	q_i	0	665393	0.010
	(b)	q^*	6337	40770674	0.075
	(c)	$\tilde{q}_{opt}$	0	613595	0.009
	(d)	$\hat{q}_i$	1478	2832153	0.020
	(e)	$\hat{q}_{Av}$	4742	23107527	0.057
	(f)	$\tilde{q}_{mod}$	1959	4453478	0.025
MN					
	(a)	q_i	0	357278	.017
	(b)	q^*	2387	6031873	.070
	(c)	$\tilde{q}_{opt}$	0	331459	.016
	(d)	$\hat{q}_i$	8421	71263655	.239
	(e)	$\hat{q}_{Av}$	9456	89750630	.268
	(f)	$\tilde{q}_{mod}$	8333	69773960	.236

responding HU's. Certainly it seems from the comparisons for DE and MD that it is worthwhile to use predictively modelled response probabilities to define weights, since the information to approach the truly optimal weights of settings (a) or (c) will never be available. Stratum-dependent weights $\tilde{q}_i$ derived from a good predictive model, as in (d), seem to provide nearly as good MSE's, but remarkably, constant weights $\tilde{q}_{mod}$ in (f) are virtually just as good. Note that the relatively small changes among constant weights in DE — $q^* = 0.14663$ in (b), to $\hat{q}_{Av} = 0.14786$ in (e), to $\tilde{q}_{opt} = 0.14378$ in (c), and finally to $\tilde{q}_{mod} = 0.14477$ in (f) — have sizeable consequences in MSE. However, the MN results in Table 1 show that, even in a model used in (3) on same-state data, unanticipated biases can results in unsatisfactory *rMSE/Nrsp* values up to 25%.

One way of trying to assess the robustness of the model-based weighting-adjustment plans (d)–(f) is to see how much noise is needed to corrupt the stage-wise PSU predictions before the derived MSE's for the recommended plan (f) become worse than for plan (b), in which the overall true response rate was used in raw form, not broken down by response mode. In each line of Table 2, we display MSE values derived from 'corrupted' predictive-model values p_{ik}^{aux} satisfying $logit(p_{ik}^{aux}) = logit(p_{ik}^{mod}) + \epsilon_{ik}$, where p_{ik}^{mod} are the fitted model values and ϵ_{ik} are *iid* simulated $\mathcal{N}(0, \sigma_e^2)$ deviates, with standard deviations σ_e shown in the first column. Recall, for purposes of comparison, that the MSE values under plans (a), (b), (c) are respectively 6.65e5, 4.08e7, and 6.14e5. Each MSE value in the (f) column which exceeds the plan-(b) MSE value is followed by an asterisk (∗). From Table 2, we can see that additive noise on *logit* scale of up to 0.3 or 0.4 in the model-predicted PSU response rates by mode will usually still leave the MSE's for plan (f) less than the MSE's for plan (b). This leaves a comfort zone: this standard deviation is only slightly less than the PSU random-effect standard deviation of .50 to .55 fitted (Slud 1998) in a mixed-effect (random-intercept) logistic regression for Mail-Response.

Table 2. MSE's of three model-based plans for DE data, with DE model predictors at each of 4 stages corrupted by additive noise on logit scale with standard deviation σ_e. Asterisks (∗) show cases with (f) MSE > $4.08e7$ = (b) MSE.

σ_e	(d) MSE	(e) MSE	(f) MSE	
0	2.82e6	2.31e7	4.44e6	
0.01	3.04e6	2.37e7	4.70e6	
0.1	6.57e6	3.01e7	7.15e6	
0.3	4.33e7	5.83e7	1.77e7	
0.4	9.20e7	7.78e7	2.40e7	
0.4	2.11e8	1.75e8	8.09e7	(*)
0.5	2.27e7	1.54e7	6.00e7	(*)
0.5	2.29e7	1.52e7	6.25e7	(*)
0.6	5.74e7	3.11e7	1.43e8	(*)

To make these calculations more realistic, we consider several cases of using reasonable but *wrong* models. Table 3 summarizes the relative performance of various neighboring states' predictive models versus models fitted and applied on the same states' data. Correlations are shown between predicted and actual rates of response, at each of the 4 (**MR**, **rsp50**, **rsp75**, and **rsp90**) check-in stages considered by Slud (1999, 2001), at the level of block-group-by-htyp stratum, for each of 5 pairs of states. These correlations vary considerably by stage, and are generally a little but sometimes a lot worse at each stage when based on a neighboring state's model.

In contrasting the MSE performance of model-based estimators when the model is that of the same versus a neighboring state, we restrict attention in Table 4 to the plans (b) from (7) and (f) from (10)) for assigning $\tilde{q}$ in the weights (3). Table 4 shows

Table 3. Correlations between data and model predictions for response rates at each of four checkin stages, at block-group-by-htyp level, in 1990 census.

Data	Model	MR	rsp50	rsp75	rsp90
OH	PA	.80	.04	.16	.12
OH	OH	.82	.10	.19	.23
PA	OH	.83	.12	.19	.17
PA	PA	.85	.18	.22	.15
MD	PA	.80	.16	.15	.11
MD	MD	.83	.22	.16	.14
MN	WI	.78	.13	.25	.21
MN	MN	.82	.10	.31	.30
OR	WA	.76	.06	.07	.26
OR	OR	.80	.12	.11	.28
WA	OR	.77	.04	.08	.17
WA	WA	.79	.13	.16	.28

Table 4. MSE's in 1990 data of two weighting plans [(b) as benchmark, and (f) as model-based choice] for various states, using models from either the same or a neighboring state.

Data	Model	Plan	**Bias**	**MSE**	$\frac{\textbf{rMSE}}{\textbf{Nresp}}$
MD	MD	(b)	6337	4.077e7	.075
MD	PA	(f)	-1794	2.093e6	.017
OH	OH	(b)	8606	7.520e7	.068
OH	PA	(f)	39663	1.574e9	.310
MN	MN	(b)	2387	6.032e6	.070
MN	WI	(f)	-2455	6.357e6	.071
OR	OR	(b)	2231	5.306e6	.018
OR	WA	(f)	-231	3.805e5	.019
WA	WA	(b)	5873	3.513e7	.081
WA	OR	(f)	4352	1.958e7	.060

the variety of biases and values *rMSE/Nrsp* achieved in the same set of state-pairs as in Table 3. (The numbers of non-responder HU's in these states are: 128011 in OH, 84883 in MD, 35321 in MN, 32875 in OR, and 73252 in WA.

The MSE's in Table 4 for model-adjusted weighting estimates (f) for MD based on the PA model are satisfactorily small. Based on the evidence of Table 3, greater degrees of model misspecification are very likely, and Table 4 shows that while *rMSE/Nrsp* always remains well under 0.1 for same-state models, it can become awful for misspecified-model-based weighting plan (f), as it did for the PA model on OH data. The table confirms that model-based weighting adjustment has risks ! But strangely, on MN data, while we saw in Table 2 that plan (f) yielded a very bad *rMSE/Nrsp* of 0.24 using its same-state model, the value for (f) on MN data under the WI-based model has the much better value 0.71.

4. CONCLUSIONS

This paper has studied sample-weighted estimation of population totals, based on survey data collected in several stages, in which nonresponders from earlier stages are sub-sampled at later stages. The theoretical and data-analytic comparisons among MSE's for weighting plans lead to two conclusions:

(i) Weights for analyzing such survey data should be constant across response modes, although vary across demographically distinct PSU's.

(ii) The weighting adjustment form (f), with weights $\tilde{w}_i \equiv 1/(\pi_i(1-\tilde{q}_{mod}))$, seems in all cases studied to be close to optimal. Although it appears robust to some degree of model misspecification, unacceptably high MSE's result in cases of extreme misspecification. But no method of estimating survey nonresponse can be free of such model-bias.

Further research is needed to check whether the magnitudes of model misspecification between decennial-census and ACS datasets, or of ACS datasets from one time-period to another, are of the order of those investigated here.

5. REFERENCES

Adeshiyan, S. (1998) A study of the weighting adjustment procedures for the American Community Survey. *Proc. Amer. Statist. Assoc. on Survey Research Methodology*, 178-183.

Dahl, S. (1998) Weighting the 1996 and 1997 American Community Surveys. *Proc. Amer. Statist. Assoc. on Survey Research Methodology*, 172-177.

Elliott, M., Little, R., and Lewitzky, S. (2000) Subsampling callbacks to improve survey efficiency. *Jour. Amer. Statist. Assoc.* **95**, 730-738.

Groves, R. (1999) **Survey Nonresponse**. New York: Wiley.

Särndal, C., Swensson, B., and Wretman, J., **Model-Assisted Inference in Survey Sampling**. Springer-Verlag: New York 1997.

Slud, E. (1998) Predictive models for decennial census household response. *Proc. Amer. Statist. Assoc. on Survey Research Methodology*, 272-277.

Slud, E. (1999) Analysis of 1990 decennial census checkin-time data. Proc. Fed. Comm. Statist. Methodology Res. Conf., Pt. 2, pp. 635-44, June 2000, *Statistical Policy Working Paper 30, OMB.*

Slud, E. (2001) Order selection, random effects and multilevel predictors in modelling Decennial Census response. *Proc. Amer. Statist. Assoc. on Survey Research Methodology.*