Survey Methodology, December 1993
Vol. 19, No. 2, pp. 127–136
Statistics Canada

Estimation of Measurement Bias Using a Model Prediction Approach

PAUL P. BIEMER and DALE ATKINSON[1]

ABSTRACT

Methods for estimating response bias in surveys require "unbiased" remeasurements for at least a subsample of observations. The usual estimator of response bias is the difference between the mean of the original observations and the mean of the unbiased observations. In this article, we explore a number of alternative estimators of response bias derived from a model prediction approach. The assumed sampling design is a stratified two-phase design implementing simple random sampling in each phase. We assume that the characteristic, y, is observed for each unit selected in phase 1 while the true value of the characteristic, μ, is obtained for each unit in the subsample selected at phase 2. We further assume that an auxiliary variable x is known for each unit in the phase 1 sample and that the population total of x is known. A number of models relating y, μ and x are assumed which yield alternative estimators of $E(y - \mu)$, the response bias. The estimators are evaluated using a bootstrap procedure for estimating variance, bias, and mean squared error. Our bootstrap procedure is an extension of the Bickel-Freedman single phase method to the case of a stratified two-phase design. As an illustration, the methodology is applied to data from the National Agricultural Statistics Service reinterview program. For these data, we show that the usual difference estimator is outperformed by the model-assisted estimator suggested by Särndal, Swensson and Wretman (1991), thus indicating that improvements over the traditional estimator are possible using the model prediction approach.

KEY WORDS: Reinterview; Repeated measures; Response error; Bootstrap.

1. INTRODUCTION

It is well-known in the survey literature that when responses are obtained from respondents in sample surveys, the observed values of measured characteristics may differ markedly from the true values of the characteristics. Evidence of these so-called measurement errors in surveys has been collected in a number of ways. For example, the recorded response may be checked for accuracy against administrative records or legal documents within which the true (or at least a more accurate) value of the characteristic is contained. An alternative approach relies on revised reports from respondents via reinterviews. In a reinterview, a respondent is recontacted for the purpose of conducting a second interview regarding the same characteristics measured in the first interview. Rather than simply repeating the original questions in the interview, there may be extensive probes designed to elicit more accurate responses, or the respondent may be instructed to consult written records for the "book values" of the characteristics. For some reinterview surveys, descrepancies between the first and second interviews are reconciled with the respondent until the interviewer is satisfied that a correct answer has been obtained. Forsman and Schreiner (1991) provide an overview of the literature for these types of reinterviews. Other means of checking the accuracy of survey responses include: (a) comparing the survey statistics (*i.e.*, means, totals, proportions, *etc.*) to statistics from external sources that are more accurate; (b) using experimental designs to estimate the effects on survey estimates of interviewers and other survey personnel; and (c) checking the results within the same survey for internal consistency.

The focus of the current work is on estimators of measurement bias from data collected in true value remeasurement studies, *i.e.*, record check and reinterview studies, where the objective is to obtain the true value of the characteristic at, perhaps, a much greater cost per measurement than that of the original observation.

Because of the high costs typically involved in conducting reinterview studies, repeated measurements are usually obtained for only a small fraction of the original survey sample. While the sample size may be quite adequate for estimating biases at the national and regional levels, they may not be adequate for estimating the error associated with small subpopulations or rare survey characteristics. In this paper, our objective is to consider estimators of response bias having better mean squared error properties than the traditional estimators. The basic idea behind our approach can be described as follows.

In a typical remeasurement study, a random subsample of the survey respondents is selected and, through some means, the true values of the characteristics of interest are ascertained. Let n_1 denote the number of respondents to

[1] Paul P. Biemer, Principal Scientist, Center for Survey Research, Research Triangle Institute, Research Triangle Park, NC 27709; Dale Atkinson, Supervisory Mathematical Statistician, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Va 22030.

the first survey and let n_2 denote the number selected for the subsample or evaluation sample. The usual estimator of response bias is the net difference rate, computed for the n_2 respondents in the evaluation sample as

$$\text{NDR} = \bar{y}_2 - \bar{\mu}_2, \tag{1.1}$$

where $\bar{y}_2$ is the sample mean of original responses and $\bar{\mu}_2$ is the sample mean of the true measurements. A disadvantage of the NDR is that it excludes information on the $n_1 - n_2$ units in the original survey who were not included in the remeasurement study. Further, the estimator does not incorporate information on auxiliary variables, x, which may be combined with the information on y and μ available from the survey to provide a more precise estimator of response bias.

Given that we have a stratified, two-phase sample design and resulting data (y, μ, x), our objective is to determine the "best" estimator of measurement bias given these data. Our basic approach is to identify a model for the true value, μ_i, which is a function of the observed values, y_i, $i = 1, \ldots, n_1$, and any auxiliary information, x, that may be available for the population. The model is then used to predict μ_i for all units in the population for which μ_i is unknown. These predictions can then be used to obtain estimates of the true population mean, total, or proportion. Thus, estimators of the response bias for these parameters can be derived from the main survey. Since the approach provides a prediction equation for μ_i which is a function of the observations, estimators of response bias can be computed for areas having small sample sizes. In this case, the prediction equation for μ_i may be augmented by other respondent variables such as demographic characteristics, type of unit, unit size, geographic characteristics, and so on.

The basic estimation and evaluation theory for a prediction approach to the estimation of response bias is presented in the following sections. Under stratified random sampling, estimators of means and totals, their variances and their mean squared errors are provided. Results from application to National Agricultural Statistics Service (NASS) data are also presented.

2. METHODOLOGY FOR ESTIMATION AND EVALUATION

2.1 The Measurement Error Model

To fix the ideas, we shall consider the case of simple random sampling without replacement (SRSWOR) from a single population. Generalizations to stratified random sampling are straightforward and will be considered subsequently.

Let $U = \{1, 2, \ldots, N\}$ denote the label set for the population and let $S_1 = \{1, 2, \ldots, n_1\}$, without loss of generality, denote the label set for the first phase SRSWOR sample of n_1 units from U.

For y_i, $i \epsilon S_1$, assume the model

$$y_i = \gamma_0 + \gamma\mu_i + \epsilon_i, \tag{2.1}$$

where μ_i is the true value of the measured characteristic, γ_0 and γ are constants, and ϵ_i is an independent error term having zero expectation and conditional variance, $\sigma^2_{\epsilon i}$.

Since the focus of our investigation is on the bias associated with the measurements y_i, consider the expectation of y_i. Let $\mathrm{E}(y_i \mid i)$ denote the conditional expectation of y_i over the distribution of the ϵ_i holding the unit i fixed and let $\mathrm{E}(y_i) = \mathrm{E}_i[\mathrm{E}(y_i \mid i)]$ denote the expectation of $\mathrm{E}(y_i \mid i)$ over the sampling distribution. Then, for a given unit, i,

$$\mathrm{E}(y_i \mid i) = \gamma_0 + \gamma\mu_i \tag{2.2}$$

and, hence, the unconditional expectation is

$$\mathrm{E}(y_i) = \gamma_0 + \gamma\bar{M}, \tag{2.3}$$

where $\bar{M} = \sum_{i=1}^{N} \mu_i/N$. Thus, the measurement bias is

$$\bar{B} = \mathrm{E}(y_i - \mu_i) = \gamma_0 + (\gamma - 1)\bar{M}. \tag{2.4}$$

The parameter, γ_0, is a constant bias term that does not depend upon the magnitude of $\bar{M}$. Note that this definition of γ_0 is consistent with the usual definition of measurement bias obtained from the simple model

$$y_i = \mu_i + \epsilon_i, \tag{2.5}$$

with $\epsilon_i \sim (\gamma_0, \sigma^2_{\epsilon i})$. (See, for example, Biemer and Stokes 1991.)

Consider the estimation of $\bar{B}$. Assume that a subsample of size n_2 of the original n_1 sample units is selected and the true value, μ_i, is measured for these n_2 units. The true value may be ascertained either by a reinterview, a record check, interviewer observation, or some other means. Let $S_2 \subseteq S_1$ denote this so-called second phase sample. The usual estimator of the measurement bias is the NDR defined in (1.1). If the assumption that "the true value, μ_i, is observed in phase 2, for all $i \epsilon S_2$" is satisfied, then the NDR is an unbiased estimator of $\bar{B}$. It may further be shown that the variance of the NDR is

$$\mathrm{E}\left\{\left(1 - \frac{n_2}{n_1}\right)\frac{s^2_\mu}{n_2}\left(1 - \frac{s^2_{\mu y}}{s^2_y s^2_\mu}\right) + \left(1 - \frac{n_2}{n_1}\right)\frac{s^2_y}{n_2}(1 - r)^2\right\}, \tag{2.6}$$

where $s_\mu^2 = \sum_{j\in S_2} (\mu_j - \bar{\mu}_2)^2/(n_2 - 1)$ with analogous definitions for s_y^2 and $s_{\mu y}$, and $r = s_{\mu y}/s_y^2$.

The NDR may be suboptimal in a number of situations which occur with some frequency. To see this, consider estimators of $\bar{B}$ of the form

$$\bar{b}_{ga} = \bar{y}_g - \bar{\mu}_{Ra}, \tag{2.7}$$

where $\bar{y}_g = \sum_{j\in s_g} y_j/n_g$, $g = 1, 2$,

$$\bar{\mu}_{Ra} = \bar{\mu}_2 + a(\bar{y}_1 - \bar{y}_2) \tag{2.8}$$

and $\bar{\mu}_2 = \sum_{j\in s_2} \mu_j/n_2$, for a a constant given the subsample, S_1. It can be shown that the value of a that minimizes $\mathrm{Var}(\bar{b}_{ga})$ is

$$a = r \quad \text{for} \quad g = 1,$$
or
$$a = r - 1 \quad \text{for} \quad g = 2. \tag{2.9}$$

Thus, for $g = 1$ or 2, the "optimal" choice of $\bar{b}_{ga}$ is

$$\bar{b}_{\text{opt}} = \bar{y}_1 - [\bar{\mu}_2 + r(\bar{y}_1 - \bar{y}_2)], \tag{2.10}$$

which differs from the NDR by the term $(r-1)(\bar{y}_1 - \bar{y}_2)$. Since, in general, $\bar{y}_1 \neq \bar{y}_2$, NDR is optimal only if $r = 1$. It can be shown that this corresponds to the case where γ_1 in (2.1) is 1.

In this paper we shall explore alternatives to the NDR which incorporate information on y for units in the set $S_1 \sim S_2$ as well as information on some auxilliary variable, x. To illustrate the concepts, we shall restrict ourselves to "no-intercept" linear models initially, *i.e.*, models for which $\gamma_0 = 0$ in (2.1). This important class of models includes the difference estimator as well as ratio estimators.

2.2 Model Prediction Approaches To Estimation

Model prediction approaches to the estimation of population parameters in finite population sampling are well-documented in the literature. Cochran (1977) and other authors have demonstrated the model-based foundations of the ubiquitous ratio estimator. There is also considerable literature on the choice between using weights that are derived from explicit model assumptions in estimation for complex surveys or eliminating the sample weights. Proponents of so-called model-based estimation recommend against the use of weights in parameter estimation (see, for example, Royall and Herson 1973; and Royall and Cumberland 1981). They contend that the probabilities of selection in finite population sampling, whether equal or unequal, are irrelevant once the sample is produced. The reliability criteria used by model-based samples are derived from the model distributional assumptions rather than sampling distributions. If an appropriate model is chosen to describe the relationship between the response variable and other measured survey variables, "model-unbiased" estimators of the population parameters may be obtained which have greater reliability than estimators which incorporate weights.

On the other side of the controversy are the design-based samplers. Instead of the model-based assumptions, design-based samplers assume that an estimator from a survey is a single realization from a large population of potential realizations of the estimator, where each potential realization depends upon the selected sample. The distribution of the values of the estimator when all possible samples that may be selected by the sampling scheme are considered is referred to as the sampling distribution of the estimator. Criteria for evaluating estimators under the design-based approach then consider the properties of the sampling distributions of the estimators. Under this approach, weighting of the estimators is required to achieve unbiasedness if unequal probability sampling is used.

Although the estimators of $\bar{B}$ considered here represent all three classes of estimators, the objective of this paper is not necessarily to compare design-based, model-assisted, and model-based estimators. Rather, we first seek to develop a systematic approach for evaluating alternative estimators for a given two-phase sample design. The problem considered is the following: Given a two-phase sample design and estimators of $B = N\bar{B}$ denoted by $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_p$, how does an analyst identify which estimator minimizes the mean squared error? A second objective of the article is to specify a number of alternative estimators, and apply a systematic approach for evaluating the estimators. As an illustration, the methodology will be applied to data from the National Agricultural Statistics Service's December 1990 Agricultural Survey.

2.3 The Estimators Considered in Our Study

Extending the previously developed notation to stratified, two-phase designs, let N_h denote the size of the hth stratum, for $h = 1, \ldots, L$. A two-phase sample is selected in each stratum using simple random sampling at each phase. Let n_{1h} and $n_{2h} \leq n_{1h}$ denote the phase 1 and phase 2 sample sizes, respectively, in stratum h. Let S_{1h} and $S_{2h} \subseteq S_{1h}$ denote the label sets for the phase 1 and phase 2 samples, respectively, in stratum h. Assume the following data are either observed or otherwise known:

outcome variables: $y_i \ \forall \ i \in S_{1h}$

true values: $\mu_i \ \forall \ i \in S_{2h}$

auxilliary variables: $x_i \ \forall \ i \in S_{1h}$.

Further assume that $X_h = \sum_{i\in U_h} x_i$ is known for $h = 1, \ldots, L$ where U_h is the label set for the hth stratum.

2.3.1 Weighted Estimators of M and B

As a matter of convenience, we shall consider the estimation of the bias for an estimator of a population total denoted by M. The usual estimator of $M = N\bar{M}$ is the unbiased stratified estimator given by

$$\hat{M}_{2st} = \sum_h N_h \bar{\mu}_{2h}, \tag{2.11}$$

where $\bar{\mu}_{2h} = \sum_{i \in S_{2h}} \mu_i / n_{2h}$. The corresponding estimator of $B = N\bar{B}$ is N times the NDR defined in (1.1). For stratified samples, it is

$$\hat{B}_{2st} = \hat{Y}_{2st} - \hat{M}_{2st}, \tag{2.12}$$

where $\hat{Y}_{2st} = \sum_h N_h \bar{y}_{2h}$ and $\bar{y}_{2h} = \sum_{i \in S_{2h}} y_i / n_{2h}$. Note that (2.12) does not incorporate the information on y for units with labels $i \in S_{1h} \sim S_{2h}$. An alternative estimator that uses all the data on y is

$$\hat{B}_{12st} = \hat{Y}_{1st} - \hat{M}_{2st}, \tag{2.13}$$

where $\hat{Y}_{1st} = \sum_h N_h \bar{y}_{1h}$ and $\bar{y}_{1h} = \sum_{i \in S_{1h}} y_i / n_{1h}$.

A number of model-assisted estimators can be specified for two-phase stratified designs. These may take the form of either separate or combined estimators (see, for example, Cochran 1977, pp. 327-330). Further, the ratio adjustments may be applied to either phase 1 or phase 2 stratum-level estimators. Because stratum sample sizes are typically small in two-phase samples, only combined estimators shall be considered here.

As the emphasis in this paper is on the development of the methodology for model-based estimates of measurement bias and their evaluation, we shall consider a simple, special case of the model (2.1); *viz.*, $\gamma_0 = 0$ or the no-intercept model. However, generalizations of the no-intercept methodology to multivariate intercept models do not afford any difficulties and will be considered in a subsequent paper. Thus, letting $\gamma_0 = 0$ in (2.1) we have

$$y_i = \gamma \mu_i + \epsilon_i, \tag{2.14}$$

where γ is an unknown constant and we assume $\epsilon_i \sim (0, \sigma_\epsilon^2 \mu_i)$. The least squares estimator of γ is $\hat{\gamma} = \bar{y}_{2st}/\bar{\mu}_{2st}$, where $\bar{y}_{2st} = \hat{Y}_{2st}/N$ and $\bar{\mu}_{2st} = \hat{M}_{2st}/N$. Thus, a model-assisted estimator of μ_i is $y_i/\hat{\gamma} = \bar{\mu}_{2st} y_i / \bar{y}_{2st}$ and of M is

$$\hat{M}_{2stR} = \frac{\hat{M}_{2st}}{\hat{Y}_{2st}} \hat{Y}_{1st}. \tag{2.15}$$

Using this estimator of M, two estimators of B corresponding to (2.12) and (2.13) are

$$\hat{B}_{2stR} = \hat{Y}_{2st} - \hat{M}_{2stR} \tag{2.16}$$

and

$$\hat{B}_{12stR} = \hat{Y}_{1st} - \hat{M}_{2stR}. \tag{2.17}$$

A third estimator of B can be obtained via the model

$$y_i = \beta x_i + e_i, \tag{2.18}$$

where β is a constant and $e_i \sim (0, \sigma_e^2 x_i)$. This leads to a ratio estimator of Y,

$$\hat{Y}_{xstR} = \frac{\bar{y}_{1st}}{\bar{x}_{1st}} X. \tag{2.19}$$

Thus, the corresponding estimator of B is

$$\hat{B}_{x2stR} = \hat{Y}_{xstR} - \hat{M}_{2stR}. \tag{2.20}$$

Finally, Särndal, Swensson and Wretman (1992, p. 360) suggest a general estimator of M in two-phase sampling. Applying their equation 9.7.2 to the model in (2.14) under stratified sampling yields

$$\hat{M}_{SSW} = \hat{M}_{2stR} + \frac{\bar{\mu}_{2st}}{\bar{x}_{2st}} (X - \hat{X}_{1st}). \tag{2.21}$$

Note that this estimator is simply (2.15) with the addition of the unbiased estimator of zero. The resulting estimator may have smaller variance than $\hat{M}_{2stR}$ if this term is negatively correlated with $\hat{M}_{2stR}$. Likewise, their estimator of Y reduces to $\hat{Y}_{xstR}$ defined in (2.19). Thus the corresponding estimator of B is

$$\hat{B}_{SSW} = \hat{Y}_{xstR} - \hat{M}_{SSW}, \tag{2.22}$$

which is identical to $\hat{B}_{SSW} = B_{x2stR}$ plus the second term of the right hand side of (2.21).

2.3.2 Unweighted Estimators of M and B

Rewrite M as

$$\begin{aligned} M &= \sum_{i \in S_2} \mu_i + \sum_{i \in S_1 \sim S_2} \mu_i + \sum_{i \in U \sim S_1} \mu_i \\ &= M_{(2)} + M_{(1 \sim 2)} + M_{(\sim 1)}, \end{aligned} \tag{2.23}$$

say, where $S_g = \cup_{h=1}^{L} S_{gh}$, $g = 1, 2$. The strategy for unweighted, model-based estimation is to replace μ_i in $M_{(1 \sim 2)}$ and $M_{(\sim 1)}$ by a prediction, $\hat{\mu}_i$, obtained from a model.

Using the model in (2.14), an estimator of μ_i is

$$\hat{\mu}_i = y_i / \bar{\gamma},$$

where now $\bar{\gamma} = \bar{y}_2 / \bar{\mu}_2$. Thus, an estimator of $M_{(1 \sim 2)}$ is

$$\hat{M}_{(1\sim 2)} = \frac{\bar{\mu}_2}{\bar{y}_2} \sum_{i \in S_1 \sim S_2} y_i = \frac{\bar{\mu}_2}{\bar{y}_2}\,(n_1 \bar{y}_1 - n_2 \bar{y}_2), \tag{2.24}$$

where $\bar{y}_g = \sum_{i \in S_g} y_i / n_g$, $\bar{\mu}_2 = \sum_{i \in S_2} \mu_i / n_2$, and $n_g = \sum_h n_{gh}$, for $g = 1, 2$. Further, using the model

$$\mu_i = \delta x_i + \xi_i, \tag{2.25}$$

where δ is a constant and $\xi_i \sim (0, \sigma_\xi^2 x_i)$, we obtain

$$\hat{M}_{(\sim 1)} = \frac{\bar{\mu}_2}{\bar{x}_2} X_{U \sim S_1}, \tag{2.26}$$

where $X_{U\sim S_1} = \sum_{i \in U \sim S_1} X_i$. Thus, a model based estimator of M is

$$\hat{M}_M = M_{(2)} + \hat{M}_{(1\sim 2)} + \hat{M}_{(\sim 1)} = \hat{M}_{(1)} + \hat{M}_{(\sim 1)}, \tag{2.27}$$

where $\hat{M}_{(1)} = n_1\, \bar{\mu}_2 \bar{y}_1 / \bar{y}_2$.

Likewise, Y can be rewritten as

$$Y = \sum_{i \in S_1} y_i + \sum_{i \in U \sim S_1} y_i = Y_{(1)} + Y_{(\sim 1)} \tag{2.28}$$

and we wish to predict y_i in $Y_{(\sim 1)}$. Using the model in (2.18) a model-based estimator of $Y_{(\sim 1)}$ is

$$\hat{Y}_{(\sim 1)} = \frac{\bar{y}_1}{\bar{x}_1} X_{U \sim S_1}$$

and, thus, an estimator of Y is

$$\hat{Y}_M = Y_{(1)} + \hat{Y}_{(\sim 1)}. \tag{2.29}$$

Thus, B is estimated as

$$\hat{B}_M = \hat{Y}_M - \hat{M}_M. \tag{2.30}$$

Versions of $\hat{B}_{2stR}$, $\hat{B}_{12stR}$, $\hat{B}_{x2stR}$ and $\hat{B}_M$ which are more robust to model outliers may also be constructed. The corresponding estimators, denoted by $\tilde{B}_{2stR}$, $\tilde{B}_{12stR}$, $\tilde{B}_{x2stR}$ and $\tilde{B}_M$, respectively, may be formed by eliminating those data points which deviate substantially from the model predictions and computing the model-based or model-assisted estimators using the remaining data. To illustrate, consider the estimator $\hat{M}_{2stR}$ in (2.15). For this estimator, let

$$(n_{2h} - 1) s_{res,h}^2 = \sum_{\mu_{hi} \neq 0} \frac{(y_{hi} - \hat{\gamma}\,\mu_{hi})^2}{\mu_{hi}}, \tag{2.31}$$

denote the sum of squares of residuals for the model (2.14). Then, in calculating the estimator of γ, only those units in $i \in \tilde{S}_{2h}$ where $\tilde{S}_{2h} = \{ i \in S_{2h} : | y_{ih} - \hat{\gamma}\mu_{ih} | \le 3 s_{res,h} \sqrt{\mu_{hi}} \}$ are used. Denoting this estimator of γ as $\tilde{\gamma}$, the estimator of M is $\tilde{M}_{2stR} = \hat{Y}_{1st}/\tilde{\gamma}$ where $\tilde{\gamma} = \tilde{y}_{2st}/\tilde{\mu}_{2st}$ and $\tilde{\mu}_{2st}$ and $\tilde{y}_{2st}$ are the stratified means of μ_i and y_i for $i \in \tilde{S}_{2h}$. The other robust model prediction estimators may be computed analogously.

Many other unweighted, model-based estimators may be explored in the context of our two-phase design. For example, an intercept term may be added to models (2.14), (2.18), and (2.25). Further, slope and intercept parameters may be specified separately for each stratum or combination of strata.

2.4 Estimation of Mean Squared Errors Using Bootstrap Estimators

Although it is possible, under the appropriate design-based or model-based assumptions, to derive closed form analytical estimates of the variance of the estimators we are considering in this study, we have elected instead to use a computer-intensive resampling method. First, we seek a method which is easy to apply since there are potentially many estimators which will be considered in our study. Secondly, it is important to evaluate each estimator using the same criteria and a consistent method of variance estimation is essential to achieving this objective. Thus, it is essential that we employ a variance estimation method which can be applied to estimators of any complexity, under assumptions which are consistent and which do not rely upon any model assumptions. It is well-known that model-based variance estimation approaches are quite sensitive to model failure (see, for example, Royall and Herson 1973; Royall and Cumberland 1978; and Hansen, Madow and Tepping 1983). Royall and Cumberland (1981) discuss several bias relevant alternatives including the jackknife variance estimator.

Our approach is similar to that of Royall and Cumberland except rather than using a jackknife estimator, we employ a bootstrap estimator of the variance. For independent and identically distributed observations, Efron and Gong (1983) show that the bootstrap and the jackknife variance estimators differ by a factor of $n/(n - 1)$ for samples of size n. Thus, the robustness properties Royall and Cumberland demonstrate for the jackknife estimator also hold for the bootstrap estimator.

Other properties of the bootstrap estimator have led us to choose it above other resampling methods. The jackknife and balance repeated replication (BRR) methods are not easily modified for the two-phase sampling design of our study. However, the bootstrap is readily adaptable to two-phase sampling. Further, Rao and Wu (1988) provide evidence from a simulation study that the coverage properties of bootstrap confidence intervals in complex sampling compare favorably to the jackknife and BRR.

Our general approach extends the method developed by Bickel and Freedman (1984) for single phase, stratified sampling, to two-phase stratified sampling. Since the bootstrap procedure is implemented independently for each stratum, we shall, for simplicity, describe the method for the single stratum case.

2.4.1 Estimation of Variance

Extending the bootstrap method to two-phase sampling is not simply a matter of subsampling the single phase bootstrap samples. Recall that true values are known only for the units in S_2 and, therefore, the bootstrap sampling scheme must necessarily confine the selection to units in S_2. Therefore, let S_1 and S_2 denote the phase 1 and phase 2 samples, respectively, selected from U using SRSWOR. Let $S_{1\sim 2}$ denote the label set, $S_1 \sim S_2$. Let $\hat{\Theta} = \hat{\Theta}(S_{1\sim 2}, S_2)$ denote an estimator of Θ which may be a function of the observations corresponding to units in both S_2 and $S_{1\sim 2}$. Define N, n_1, n_2 and $n_{1\sim 2}$ as the sizes of sets U, S_1, S_2 and $S_{1\sim 2}$, respectively. Consider how the bootstrap is applied to obtain estimates of $\text{Var}(\hat{\Theta})$.

The simplest case is when N/n_1 is an integer, say k. First, we form the psuedo-population label set

$$U_A^* = U_{A(2)}^* \cup U_{A(1\sim 2)}^*, \qquad (2.32)$$

where $U_{A(2)}^*$ consists of k copies of the units in S_2 and $U_{A(1\sim 2)}^*$ consists of k copies of the units in $S_{1\sim 2}$. We then perform the following three steps:

1. Draw a SRSWOR of size n_2 from $U_{A(2)}^*$ and denote this set by S_2^*.
2. Draw a SRSWOR of size $n_{1\sim 2}$ from $U_{A(1\sim 2)}^*$ and denote this set by $S_{1\sim 2}^*$.
3. Compute $\hat{\Theta}_1^* = \hat{\Theta}_1(S_{1\sim 2}^*, S_2^*)$ which has the same functional form as $\hat{\Theta}(S_{1\sim 2}, S_2)$, but is computed for the $n_1 = n_{1\sim 2} + n_2$ units in $S_1^* = S_{1\sim 2}^* \cup S_2^*$.

Repeat steps 1 to 3 some large number, Q, times to obtain $\Theta_1^*, \ldots, \Theta_Q^*$. Then, an estimator of $\text{Var}(\hat{\Theta})$ is

$$\text{var}_{BSS}(\hat{\Theta}) = \sum_{q=1}^{Q} \frac{(\hat{\Theta}_q^* - \hat{\Theta}_\cdot^*)^2}{Q-1}, \qquad (2.33)$$

where $\hat{\Theta}_\cdot^* = \sum_{q=1}^{Q} \hat{\Theta}_q^*/Q$.

Using the methods of Rao and Wu (1988), it can now be shown that $\text{var}_{BSS}(\hat{\Theta})$ is a consistent estimator of $\text{Var}(\hat{\Theta})$. If $N = kn_1 + r$, where $0 < r < n_1$, the procedure is modified as follows using the Bickel and Freedman procedure. First, form the pseudo-population U_A^* as above consisting of kn_1 units. In addition, form the pseudo population $U_B^* = U_{B(1\sim 2)} \cup U_{B(2)}^*$ of size $(k+1)n_1$ where $U_{B(1\sim 2)}^*$ and $U_{B(2)}$ consist of $k+1$ copies of the labels in $S_{1\sim 2}$ and S_2, respectively. Then, for αQ of the bootstrap samples, select $S_1^* = S_{1\sim 2}^* \cup S_2^*$ from U_A^* and for $(1-\alpha)Q$ samples, select S_1^* from the psuedo-population, U_B^* using the three-step procedure described above, where

$$\alpha = \left(1 - \frac{r}{n_1}\right)\left(1 - \frac{r}{N-1}\right). \qquad (2.34)$$

2.4.2 Estimation of Bias and MSE

The bootstrap procedure can also provide an estimate of estimator bias. The usual bootstrap bias estimator (see Efron and Gong 1983; Rao and Wu 1988) is $b(\hat{\Theta}) = \hat{\Theta}_\cdot^* - \hat{\Theta}$ where $\hat{\Theta}_\cdot^* = \sum_q \hat{\Theta}_q^*/Q$ and $\hat{\Theta}$ is the estimate computed from the full sample. Note that $\hat{\Theta}_q^*(q = 1, \ldots, Q)$ and $\hat{\Theta}$ have the same functional form and are based upon the same model assumptions. Thus $b(\hat{\Theta})$ does not reflect the contribution to bias due to model failure. We propose an alternative estimator of bias which we conjecture is an improvement over $b(\hat{\Theta})$.

Recall from (2.4) that $\bar{B} = \text{E}(y_i - \mu_i)$ where $\text{E}(\cdot)$ denotes expectation over both the measurement error and sampling error distributions. Thus, $\bar{B}$ may be rewritten as $\bar{B} = \sum_{i=1}^{N} (Y_i - \mu_i)/N$ where $Y_i = \text{E}(y_i \mid i)$. Since Y_i is unknown and unobservable for all $i \in U$, $\bar{B}$ is also unknown and unobservable. Therefore, we shall construct a pseudo population resembling U, denoted by U^*, such that $\bar{B}^* = \text{E}^*(y_i - \mu_i)$ is known, where $\text{E}^*(\cdot)$ is expected value with respect to both the measurement error and the sampling distributions associated with U^*.

Let $U^* = \cup_{h=1}^{L} U_h^*$ where U_h^* consists of $k_h = N_h/n_{1h}$ copies of the units in S_{1h}. Here we have assumed k_h is an integer, but we will subsequently relax the assumption. Further, denote by y_i^* the value of the characteristic for the unit $i \in U^*$. This value is equal to the y_i for the corresponding unit in S_1. Thus, the population total of the y_i^* is $Y^* = \sum_{i \in U^*} y_i^* = \hat{Y}_{1st}$ for $\hat{Y}_{1st}$ defined in (2.13). Analogously, define the true value for unit $i \in U^*$ as $\mu_i^* = \mu_j$ for $i \in U^*$ corresponding to $j \in S_2$. For $j \in S_{1\sim 2}$, μ_j is unknown; however, for our pseudo-population we could generate pseudo-values for the μ_i^* such that $M^* = \sum_{i \in U^*} \mu_i^* = \hat{M}_{2st}$ where $\hat{M}_{2st}$ is defined in (2.11). Thus, for U^*, $B^* = \hat{Y}_{1st} - \hat{M}_{2st} = \hat{B}_{12st}$ defined in (2.13). As we shall see, it is not necessary to generate the pseudo-values for μ_i^* in order to evaluate the bias in the estimators of B^*.

Note that under stratified sampling, $U^* = U_A^*$, as defined in Section 2.4. Further, the bootstrap procedure described in this section is equivalent to repeated sampling from U^* and the alternative estimators $\hat{\Theta}_1, \ldots, \hat{\Theta}_p$ of B

may also be considered estimators of B^*. Since B^* is known, the bias of $\hat{\Theta}$ as an estimator of B^* is $\hat{B}^* = \hat{\Theta} - B^*$ and the corresponding MSE may be estimated as

$$\widehat{\text{MSE}} = \sum_q (\hat{\Theta}_q - B^*)^2/Q$$
$$\doteq \text{var}_{BSS}(\hat{\Theta}) + (\hat{\Theta}^*_\cdot - B^*)^2, \qquad (2.35)$$

where $\text{var}_{BSS}(\hat{\Theta})$, $\hat{\Theta}_q$, and $\hat{\Theta}^*_\cdot$ are defined in Section 2.4. It can be easily verified that these results still hold when k_h is non-integer.

Thus, the bootstrap procedure provides a method for evaluating the MSE of alternative estimators for estimating B^*. Further, the pseudo-population U^* is a reconstruction of U based upon copies of the values for the units in S_1 and S_2. Thus, it is reasonable to use $\hat{B}^*$ and $\widehat{\text{MSE}}^*$ to evaluate alternative estimators of B.

3. APPLICATION TO THE AGRICULTURAL SURVEY

3.1 Description of the Survey

The National Agricultural Statistics Service (NASS) annually conducts a series of surveys which are collectively referred to as the Agricultural Survey (AS) program. The purpose of these surveys is to collect data related to specific agricultural commodities at the state and national levels. Each December in the years 1988-1990, reinterview studies designed to assess the measurement bias in the data collected by Computer Assisted Telephone Interviewing (CATI) were conducted in six states: Indiana, Iowa, Minnesota, Nebraska, Ohio, and Pennsylvania. The reinterview techniques employed in these three studies are very similar to those used by the U.S. Census Bureau (see, for example, Forsman and Schreiner 1991). However, unlike the Census Bureau's program, the major objective in the NASS studies is the estimation of measurement bias rather than interviewer performance evaluation.

As noted above, only AS responding units whose original interview was conducted by CATI were eligible for selection into the reinterview sample. The reasons for this restriction on sampling were primarily cost, timing, and convenience. However, a large proportion of the AS is conducted by CATI and, thus, information regarding AS measurement bias for this group would provide important information for the entire AS program.

For the NASS reinterview studies, the interviewing staff consisted of a mix of field supervisors and experienced field interviewers. This interviewing staff, which was a separate corps of interviewers from those used for CATI, conducted face-to-face reinterviews in a subsample of AS units for a subset of AS survey items. To minimize any problems that respondents may have with recall, the reinterviews were conducted within 10 days of the original interview. Differences between the original AS and reinterview responses were reconciled to determine the "true" value. Considerable effort was expended in procedural development, training, and supervision of the reinterview process to ensure that the final reconciled response was completely accurate. For the most part, the wording of the subset of AS questions asked in the reinterview was identical to that of the parent survey. The reinterviewers attempted to contact the most knowledgeable respondent in order to ensure the accuracy of the reconciled values.

In this report, only the 1990 data are analyzed. Table 1 presents the reinterview sample sizes for this study.

Table 1

Sample Sizes by Survey Item

Item	x	y	μ
	U	S_1	S_2
All wheat stocks	108,267	8,176	1,157
Corn planted acres	225,269	8,211	1,157
Corn stocks	225,269	7,990	1,115
Cropland acreage	278,045	8,274	1,141
Grain storage capacity	207,460	8,126	1,104
Soybean planted acreage	171,761	8,211	1,156
Soybean stocks	171,761	8,113	1,130
Total land in farm	276,450	8,309	1,159
Total hog/pig inventory	248,571	8,247	1,142
Winter wheat seedings	108,267	8,211	1,150

3.2 Comparison of the Estimators of M and B

Using the December 1990 Agricultural Survey and its corresponding reinterview survey data, the estimators developed in the previous section were compared. Estimates of standard errors and mean squared errors were computed using the Bickel-Freedman bootstrap procedure described in Section 2.4, with $Q = 300$ bootstrap samples. Table 2 displays the results for six of the estimators: $\hat{B}_{2st}$, the traditional difference estimator; $\hat{B}_{x2stR}$, the weighted ratio estimator; $\tilde{B}_{x2stR}$, the robust (outlier deletion) version of $\hat{B}_{x2stR}$; $\hat{B}_{SSW}$, the Särndal, Swensson and Wretman estimator; $\hat{B}_M$, the unweighted model-based estimator; and $\tilde{B}_M$, the robust (outlier deletion) version of $\hat{B}_M$.

3.3 Summary of Results

Table 2 presents a summary of the results from our study. The first data column is the known value of $B^* = \mathrm{E}(y_i^* - \mu_i^*)$, the bias parameter for the pseudo-population, U^*. The other data columns contain the values of the estimators with their standard errors in parentheses, where s.e. $(\hat{\Theta}) = \sqrt{\mathrm{var}_{BSS}(\hat{\Theta})}$. The last four rows of the table correspond, respectively, to:

(a) the number of items (out of 10) for which a 95% confidence interval contains B^*;
(b) the average coefficient of variation (C.V.);
(c) the average square root of $\widehat{\mathrm{MSE}}$ (RMSE); and
(d) the average absolute relative bias.

A striking feature of these results is the large disparity among the six estimators across all commodities; particularly for All Wheat Stocks. For this commodity, the range of estimates is −94.2 to 103.2. Also indicated (by the ‡ symbol) in Table 2 is whether a 95% confidence interval, *i.e.*, $[\hat{\Theta} - 2\,\mathrm{s.e.}\,(\hat{\Theta}), \hat{\Theta} + 2\,\mathrm{s.e.}\,(\hat{\Theta})]$, covers the parameter B^*. The best performer for parameter coverage is $\hat{B}_{SSW}$ which produced confidence intervals that covered B^* for eight out of ten commodities. $\hat{B}_{2st}$ was the next best with six and $\hat{B}_M$ was third with five. The traditional ratio estimator and its robust version were the worst performers with only one commodity having a confidence interval covering B^*.

Application of the mean squared error criterion presents a different picture. Here, $\tilde{B}_M$ emerged as the estimator having the smallest average root MSE. However, $\hat{B}_{SSW}$ and $\hat{B}_{2st}$ are not much greater. Further, $\hat{B}_{SSW}$ was the estimator having the smallest average absolute relative bias. Only two commodities were estimated with significant biases using this estimator. Thus, it appears from these results that $\hat{B}_{SSW}$ is the preferred estimator using overall performance as the evaluation criterion.

Table 2

Comparison of Estimators with, B^*, the Pseudo-Population Value of the Bias†

Characteristic	B^*	$\hat{B}_{2st}$	$\hat{B}_{x2stR}$	$\tilde{B}_{x2stR}$	$\hat{B}_{SSW}$	$\hat{B}_M$	$\tilde{B}_M$
All wheat stocks	42.3	−6.1 (12.3)	103.2 (17.6)	−94.2 (16.5)	−0.9‡ (24.8)	19.2‡ (16.5)	10.6‡ (16.7)
Corn planted acreage	−1.8	1.1‡ (1.1)	11.7 (1.3)	10.1 (1.1)	0.3‡ (1.2)	−4.7‡ (1.9)	−5.0 (1.5)
Corn stocks	−6.4	−5.4‡ (1.5)	2.4 (1.6)	0.2 (1.3)	−6.5‡ (1.6)	−7.9‡ (2.4)	−9.3‡ (2.2)
Cropland acreage	27.0	−19.6 (8.3)	−15.0 (8.3)	7.0 (3.1)	−19.6 (8.2)	−36.8 (11.0)	−12.8 (4.0)
Grain storage capacity	−3.37	1.4‡ (3.7)	32.3 (3.7)	29.5 (2.6)	−0.1‡ (3.9)	−6.9 (3.0)	−6.8 (2.5)
Soybean planted acreage	−4.4	0.8 (0.8)	13.0 (1.0)	9.9 (0.9)	−0.3 (1.0)	−2.9 (1.1)	−2.7 (1.0)
Soybean stocks	−0.01	2.8‡ (3.1)	21.3 (2.9)	5.0 (2.3)	0.2‡ (3.5)	−11.0 (3.6)	−8.8 (3.4)
Total land in farm	−20.0	−24.7‡ (10.4)	−18.8‡ (12.5)	−2.6 (7.6)	−25.7‡ (10.7)	−44.5‡ (13.4)	−21.2 (5.8)
Total hogs/pigs inventory	−0.1	−2.1 (0.9)	3.4 (1.1)	−0.0‡ (1.0)	−2.2‡ (1.1)	−2.5‡ (1.3)	−1.6‡ (1.0)
Winter wheat seedings	−0.6	−0.5‡ (0.4)	3.8 (0.6)	1.8 (0.5)	−1.2‡ (0.6)	1.1 (0.4)	1.1 (0.4)
Number of items where C.I. covers B^*		6	1	1	8	5	3
Average C.V.		1.01	.30	11.1	9.5	.41	.48
Average RMSE		13.2	22.4	25.2	12.9	14.9	10.8
Average | Relbias |		30.8	220.0	53.4	4.9	113.1	91.3

† Standard errors in parentheses.
‡ 95% confidence interval covers the pseudo population parameter.

4. CONCLUSIONS AND RECOMMENDATIONS

In this article, we developed a general methodology for constructing and evaluating weighted and unweighted model prediction estimators of measurement bias for stratified random, two-phase sample designs. The proposed estimators incorporate information on the observations, y, from the first phase sample, and an auxilliary variable, x. Model robust versions of the estimators were also considered and evaluated. The ultimate goal of model prediction estimation is to identify estimators which make "optimal" use of the data (y, μ, x). The general estimation and evaluation methodology for achieving this goal was illustrated for the ordinary regression model with no intercept. However, the methodology can be easily extended to multivariate, intercept models.

Our proposed evaluation criteria are based upon estimates of bias, variance, and mean squared error computed using a bootstrap resampling methodology. The method of Bickel and Freedman was extended to two-phase sampling for this purpose. It was shown both analytically and empirically that the usual NDR estimator is not optimal under the model prediction approach to estimating measurement bias. Our analyses found that, for the six estimators we considered, the estimator derived from the work of Särndal *et al.* (1992), was the best overall estimator by the bootstap evaluation criteria.

Incorporating auxiliary information into the estimation of measurement bias creates a number of practical problems which may increase the costs and reduce the timeliness of producing the estimates. First, the auxiliary variable, x, must be available, at least in aggregate form, for all socioeconomic and geographic domains for which model prediction estimates are desired. This could be a large data management task. Further, the complexity of the variance estimator using analytical methods increases with the complexity of the bias estimator. Although simpler, the bootstrap variance estimation method can be prohibitively expensive if computer time must be purchased. However, these difficulties are not insurmountable, especially if a high-powered microcomputer is available. Further, given the cost of reinterview surveys for estimating measurement bias, even moderate increases in precision in the bias estimators can result in substantial cost savings.

The model prediction approach has the potential for extracting the maximum information on response bias from reinterview surveys and thus model prediction estimators will usually be more efficient than the traditional net difference estimator. In addition, the model prediction approach may also offer a means for extrapolating estimates of bias to areas which were not sampled. As an example, in the NASS application, the reinterview sample was drawn only from the CATI areas for reasons of operational convenience and cost efficiency. However, by using prediction models which are functions of the original responses and other available characteristics, it may be possible to predict the measurement bias in the non-CATI survey areas from the local characteristics of these areas - a type of "synthetic" estimation. Although this application of model-based estimation was not considered in this paper, it is a natural extension of the methodology and one which will be evaluated in a subsequent study.

Also for future research, we intend to incorporate multivariate, intercept models in the estimation of measurement bias. Since the bootstrap evaluation criteria developed in this article are general, no changes in the evaluation methodology are required to handle the addition of variables in the estimation models. Further, the model assumptions and the methods for handling outliers will be refined and evaluated in a subsequent paper. Finally, we need to explore the effect on estimation of departures from the model assumptions, particularly the assumption that the reinterview observation is without error. As Fuller (1991) has shown, if the reinterview is fallible but unbiased, the variance of the predicted values increases but the predictions are still unbiased. Thus, under these assumptions, one could explore the relative precision of the alternative estimators of measurement bias in order to determine the robustness of the model prediction approach.

ACKNOWLEDGEMENTS

The authors thank Manuel Cárdenas for his valuable assistance and the referees for their helpful comments.

REFERENCES

BICKEL, P., and FREEDMAN, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, 12, 470-482.

BIEMER, P., and STOKES, L. (1991). Approaches to the modeling of measurement errors. In *Measurement Errors in Surveys*, (Eds. P. Biemer, *et al.*). New York: John Wiley and Sons.

COCHRAN, W. (1977). *Sampling Techniques*. New York: John Wiley and Sons.

EFRON, B., and GONG, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 31, 36-48.

FORSMAN, G., and SCHREINER, I. (1991). The design and analysis of reinterview: an overview. In *Measurement Errors in Surveys*, (Eds. P. Biemer, *et al.*). New York: John Wiley and Sons.

FULLER, W.A. (1991). Regression estimation in the presence of measurement error. In *Measurement Errors in Surveys*, (Eds. P.P. Biemer, *et al.)*. New York: John Wiley and Sons, 617-636.

HANSEN, M., MADOW W., and TEPPING, B. (1983). An evaluation of model-dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

RAO, J.N.K., and WU, C. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

ROYALL, R., and HERSON, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 68, 880-893.

ROYALL, R., and CUMBERLAND, W. (1978). Variance estimation in finite population sampling. *Journal of the American Statistical Association*, 73, 351-361.

ROYALL, R., and CUMBERLAND, W. (1981). The finite-population linear regression estimator and estimators of its variance - an empirical study. *Journal of the American Statistical Association*, 76, 924-930.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.