Statistica Sinica **24** (2014), 395-414
doi:http://dx.doi.org/10.5705/ss.2012.064

EFFICIENCY OF MODEL-ASSISTED REGRESSION ESTIMATORS IN SAMPLE SURVEYS

Jun Shao[1,2] and Sheng Wang[3]

[1]*East China Normal University,* [2]*University of Wisconsin-Madison and* [3]*Mathematica Policy Research*

Abstract: Model-assisted regression estimators are popular in sample surveys for making use of auxiliary information and improving the Horvitz-Thompson estimators of population totals. In the presence of strata and unequal probability sampling, however, there are several ways to form model-assisted regression estimators: regression within each stratum or regression by combining all strata, and a separate ratio adjustment for population size, or a combined ratio adjustment, or no adjustment. In the literature, there is no comprehensive theoretical comparison of these regression estimators. We compare the asymptotic efficiencies of six model-assisted regression estimators under two asymptotic settings. When there are a fixed number of strata with large stratum sample sizes, our result shows that one of the six regression estimators is a clear winner in terms of asymptotic efficiency. When there are a large number of strata with small stratum sample sizes, however, the story is different. Some comparisons in special cases are also made. Some simulation results are presented to examine finite sample performances of regression estimators and their variance estimators.

Key words and phrases: Asymptotic efficiency, bootstrap, combined regression estimators, separate regression estimators, unequal probability without replacement sampling, variance estimation.

1. Introduction

Auxiliary information from sources such as administrative records is often available in sample surveys. Model-assisted regression estimators are the most popular estimators that utilize auxiliary information to gain efficiency in estimating population totals. Although regression estimators are constructed using a regression model between the variables of interest and some covariates, they are consistent and asymptotically normal under the traditional design-based framework in which the randomness of survey estimators is from repeated sampling. Thus they are robust against violation of the assumed regression model. But, when the regression model is correct, they are more efficient than the estimators that do not use the auxiliary information.

Let U denote the finite population of interest with N units stratified into H strata, $U = U_1 \cup \cdots \cup U_H$, where U_h contains N_h units and $N_1 + \cdots +$

$N_H = N$. For unit $i \in U$, let y_i be a nonconstant variable of interest and x_i be a nonconstant auxiliary variable associated with y_i. Although we consider a univariate x_i throughout, all results can be generalized to the case of multivariate x_i in a straightforward manner. Consider the estimation of the population total $Y = \sum_{i\in U} y_i$ based on a sample $S = S_1 \cup \cdots \cup S_H$ from U, where each S_h is a sample from U_h according to some probability sampling plan and $S_1, \ldots, S_H$ are independently selected. Let n_h be the size of S_h and $n = n_1 + \cdots + n_H$. For $i \in S$, y_i is observed. The Horvitz-Thompson estimator of the total Y is

$$\hat{Y} = \sum_h \hat{Y}_h, \qquad \hat{Y}_h = \sum_{i\in S_h} w_i y_i,$$

where w_i^{-1} is the first-order inclusion probability of unit i in S_h, a known quantity from the sampling design. Under the traditional design-based framework, y_i's are fixed values and the only randomness is from the repeated selection of S from U. Let E_s be the expectation with respect to repeated sampling. Then, it is well known that $E_s(\hat{Y}) = Y$.

For the auxiliary variable, the value of x_i is observed for $i \in S$, and the value of $X_h = \sum_{i\in U_h} x_i$ is known (e.g., from administrative records) for each h. To utilize the auxiliary variable x_i, the model-assisted approach (e.g., Särndal, Swensson, and Wretman (1992)) can be adopted to form regression estimators. In the presence of strata, however, there are two ways to apply the regression: regression within each stratum or regression by combining all strata. Also, one can apply a separate ratio adjustment for population size, or a combined ratio adjustment, or no adjustment. This leads to six regression estimators of Y:

$$\hat{Y}_{C0} = \sum_h \left[\hat{Y}_h + \hat{\beta}(X_h - \hat{X}_h)\right] = \hat{Y} + \hat{\beta}(X - \hat{X}), \tag{1.1}$$

$$\hat{Y}_{S0} = \sum_h \left[\hat{Y}_h + \hat{\beta}_h(X_h - \hat{X}_h)\right] = \hat{Y} + \sum_h \hat{\beta}_h(X_h - \hat{X}_h), \tag{1.2}$$

$$\hat{Y}_{CC} = \sum_h \left[\frac{N\hat{Y}_h}{\hat{N}} + \hat{\beta}\left(X_h - \frac{N\hat{X}_h}{\hat{N}}\right)\right] = \frac{N\hat{Y}}{\hat{N}} + \hat{\beta}\left(X - \frac{N\hat{X}}{\hat{N}}\right), \tag{1.3}$$

$$\hat{Y}_{SC} = \sum_h \left[\frac{N\hat{Y}_h}{\hat{N}} + \hat{\beta}_h\left(X_h - \frac{N\hat{X}_h}{\hat{N}}\right)\right] = \frac{N\hat{Y}}{\hat{N}} + \sum_h \hat{\beta}_h\left(X_h - \frac{N\hat{X}_h}{\hat{N}}\right), \tag{1.4}$$

$$\hat{Y}_{CS} = \sum_h \left[\frac{N_h\hat{Y}_h}{\hat{N}_h} + \hat{\beta}\left(X_h - \frac{N_h\hat{X}_h}{\hat{N}_h}\right)\right] = \sum_h \frac{N_h\hat{Y}_h}{\hat{N}_h} + \hat{\beta}\left(X - \sum_h \frac{N_h\hat{X}_h}{\hat{N}_h}\right), \tag{1.5}$$

$$\hat{Y}_{SS} = \sum_h \left[\frac{N_h\hat{Y}_h}{\hat{N}_h} + \hat{\beta}_h\left(X_h - \frac{N_h\hat{X}_h}{\hat{N}_h}\right)\right], \tag{1.6}$$

where $X = \sum_{i \in U} x_i$, $\hat{X} = \sum_{i \in S} w_i x_i$, $\hat{X}_h = \sum_{i \in S_h} w_i x_i$, $\hat{N} = \sum_{i \in S} w_i$, $\hat{N}_h = \sum_{i \in S_h} w_i$,

$$\hat{\beta} = \frac{\sum_h \sum_{i \in S_h} w_i y_i (x_i - \hat{X}_h/\hat{N}_h)}{\sum_h \sum_{i \in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2} \tag{1.7}$$

is the slope estimator assuming a single regression line over all strata, and

$$\hat{\beta}_h = \frac{\sum_{i \in S_h} w_i y_i (x_i - \hat{X}_h/\hat{N}_h)}{\sum_{i \in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2} \tag{1.8}$$

is the slope estimator for the regression line within stratum h. The first subscript u in $\hat{Y}_{uv}$ given by (1.1)−(1.6) indicates whether a combined regression line over all strata is assumed ($u = C$) or a separate regression line for each stratum is considered ($u = S$). The second subscript v in $\hat{Y}_{uv}$ indicates whether no ratio adjustment for population size is applied ($v = 0$), a combined (over strata) ratio adjustment $N/\hat{N}$ is applied ($v = C$), or a separate ratio adjustment $N_h/\hat{N}_h$ within each stratum is applied ($v = S$). Some of the estimators in (1.1)−(1.6) are the same or similar to those proposed by previous researchers (e.g., Isaki and Fuller (1982), Wright (1983), Montanari (1987), Särndal, Swensson, and Wretman (1992) and Fuller (2009)) and are frequently used in practice; the others are included for comparison. When combined regression is applied, knowing X, instead of each X_h, is enough for estimators $\hat{Y}_{C0}$, $\hat{Y}_{CC}$, and $\hat{Y}_{CS}$.

Intuitively, regression by combining strata is motivated by the thinking that it produces more efficient estimators when all regression models across strata are the same (which is not always true as we show in Section 2). Although some discussions on the relative efficiencies among the six estimators in (1.1)−(1.6) can be found in the literature (e.g., Cochran (1977), Särndal, Swensson, and Wretman (1992) and Fuller (2009)), a comprehensive theoretical comparison of these estimators is not available.

The purpose of this paper is to study the asymptotic relative efficiencies among these six estimators under the model-assisted approach that imposes a regression model between y_i and x_i in each stratum. We consider two types of sampling designs, one with a fixed number of strata and a large number of sampled units within each stratum and the other one with a large number of strata and a few sampled units within each stratum. Under the first type of design, we are able to draw some definite conclusions on the relative performance of estimators (1.1)−(1.6) provided the regression models are correct. Note that if we do not use any model, all six estimators are consistent and asymptotically normal under the design-based framework, but their asymptotic relative efficiencies cannot be precisely assessed unless some special designs are considered, for example, stratified simple random sampling. An example is given in Section 2.

Under the model-assisted framework, some asymptotic results for two types of designs are given in Sections 2 and 3, respectively. Some simulation results are presented in Section 4. The last section contains some discussion. All technical proofs are given in the Appendix.

2. Asymptotic Results for Large n_h's and Fixed H

To consider asymptotics, the population U is viewed as a member of a sequence of populations $\{U^{(k)}, k = 1, 2, \ldots\}$, where the number of units in $U^{(k)}$ increases to infinity as $k \to \infty$. All quantities, such as y_i, N_h, etc., depend on the index k but k is omitted for simplicity. All limiting processes are with respect to $k \to \infty$.

In this section, we assume that, in each stratum h, the population size N_h and the sample size $n_h \to \infty$ as $k \to \infty$. The number of strata, H, is fixed. The sampling design with a fixed H and large n_h's is commonly used in many surveys, for example, business surveys.

We can establish the asymptotic normality of estimators in (1.1)$-$(1.6) when the y_i's and x_i's are fixed values and the randomness is from the sample selection of S_h, $h = 1, \ldots, H$. A simple example, however, indicates that, under the design-based approach, we cannot generally tell which estimator in (1.1)$-$(1.6) is the most efficient one unless a special design is considered.

Consider that $H = 1$, each $i \in U$ has probability p_i to be selected, and sampling is with replacement. We compare the estimators $\hat{Y}_1 = \sum_{i\in S} y_i/np_i$ and $\hat{Y}_2 = N\hat{Y}_1/\hat{N}$, $\hat{N} = \sum_{i\in S} 1/np_i$, that are special cases of estimators in (1.1) and (1.2), respectively. Suppose there exists a positive constant M such that $M^{-1} \le Np_i \le M$ for all $i \in U$ and $N^{-1}\sum_{i\in U} y_i^2 \le M$. Then, by the Central Limit Theorem, $\sqrt{n}(\hat{Y}_1 - Y)/N\sqrt{v_1} \xrightarrow{d} N(0,1)$, where $\xrightarrow{d}$ denotes convergence in distribution and

$$v_1 = \frac{1}{N^2}\sum_{i\in U} p_i\left(\frac{y_i}{p_i} - Y\right)^2.$$

Applying the delta method, we obtain that $\sqrt{n}(\hat{Y}_2 - Y)/N\sqrt{v_2} \xrightarrow{d} N(0,1)$, where

$$v_2 = v_1 - \frac{2Y}{N^3}\sum_{i\in U} p_i\left(\frac{1}{p_i} - N\right)\left(\frac{y_i}{p_i} - Y\right) + \frac{Y^2}{N^4}\sum_{i\in U} p_i\left(\frac{1}{p_i} - N\right)^2.$$

Then, $\hat{Y}_2$ is asymptotically more efficient than $\hat{Y}_1$ if and only if

$$\frac{\bar{Y}^2}{N^2}\sum_{i\in U}\frac{1}{p_i} < \frac{2\bar{Y}}{N^2}\sum_{i\in U}\frac{y_i}{p_i} - \bar{Y}^2, \tag{2.1}$$

where $\bar{Y} = Y/N$. Under the design-based framework, however, y_i's are fixed values and we cannot tell whether (2.1) holds except for a few special cases. This is similar to the discussion in Section 6.6 of Cochran (1977).

If we assume the y_i's are iid with mean μ_y, then (2.1) holds for sufficiently large N unless $\mu_y = 0$ or $p_i = N^{-1}$ (simple random sampling). This can be seen as follows. Suppose $\mu_y = 0$ and p_i is not a constant. By the Law of Large Numbers and the condition $M^{-1} \le Np_i \le M$ for all $i \in U$,

$$\bar{Y} \to \mu_y \quad \text{and} \quad \sum_{i\in U} \frac{y_i}{p_i N^2} - \sum_{i\in U} \frac{\mu_y}{p_i N^2} \to 0 \;\; a.s..$$

This together with Jensen's inequality

$$\frac{1}{N}\sum_{i\in U}\frac{1}{p_i} > \Big(\frac{1}{N}\sum_{i\in U} p_i\Big)^{-1} = N$$

implies that (2.1) holds for large enough N.

We consider the model-assisted approach that views (y_i, x_i), $i \in U$, as random vectors following a model. To utilize the auxiliary variable x_i, we consider the regression model

$$y_i = \alpha_h + \beta_h x_i + \epsilon_i, \qquad i \in U_h, \quad h = 1, \ldots, H, \tag{2.2}$$

where α_h and β_h are, respectively, the unknown intercept and slope for the regression within stratum U_h, ϵ_i, $i \in U_h$, are iid with mean 0 and an unknown variance, x_i, $i \in U_h$, are iid with an unknown variance, and ϵ_i's are independent of x_i's. The estimator $\hat{\beta}_h$ in (1.8) is the least squares estimator of β_h, $h = 1, \ldots, H$. If we combine all regression lines into a single line, then the estimator $\hat{\beta}$ in (1.7) is the least squares estimator of the common slope. Although the combined regression is wrong when β_h's are unequal, estimators in (1.1)−(1.6) are still consistent and asymptotically normal since they are model-assisted estimators. This is even true when model (2.2) is incorrect.

A theorem establishes the asymptotic normality of estimators (1.1)−(1.6), with respect to the probability under regression model (2.2) and repeated sampling. Let P_s, E_s, and V_s be the probability, expectation, and variance with respect to repeated sampling, and let P_m, E_m, and V_m be the probability, expectation, and variance with respect to model (2.2). We require some conditions.

(C1) For any h, $n_h/N_h \to 0$ and there exists φ_h such that $N_h/N \to \varphi_h$.

(C2) For any h, $\sigma^2_{xh} = V_m(x_i) < \infty$, $\sigma^2_h = V_m(\epsilon_i) < \infty$, $i \in U_h$, and there exists $\delta > 0$ such that as $k \to \infty$,

$$\frac{\sum_{i\in S_h} E_m \left|R_h(w_i - 1)(x_i - \mu_{xh})\right|^{2+\delta} + \sum_{i\in U_h/S_h} E_m \left|x_i - \mu_{xh}\right|^{2+\delta}}{\left[V_m\left(R_h \sum_{i\in S_h} w_i x_i - \sum_{i\in U_h} x_i\right)\right]^{1+\delta/2}} \to 0,$$

$$\frac{\sum_{i\in S_h} E_m\left|R_h(w_i-1)\epsilon_i\right|^{2+\delta}+\sum_{i\in U_h/S_h} E_m\left|\epsilon_i\right|^{2+\delta}}{\left[V_m\left(R_h\sum_{i\in S_h} w_i\epsilon_i-\sum_{i\in U_h}\epsilon_i\right)\right]^{1+\delta/2}}\to 0,$$

where $R_h=1$, $\hat{N}/N$, or $\hat{N}_h/N_h$, and $\mu_{xh}=E_m(x_i)$ for $i\in U_h$.

(C3) There exists $M>1$ such that for any h, $N_h/(n_hM)<w_i<N_hM/n_h$.

Theorem 1. *If* (2.2) *and* (C1)$-$(C3) *hold, then*

$$\frac{\hat{Y}_{uv}-Y}{N\sqrt{\phi_{uv}}}\xrightarrow{d} N(0,1),$$

where $u=C$ *or* S, $v=0$, C, *or* S,

$$\begin{aligned}
\phi_{SS}&=N^{-2}\Big[\textstyle\sum_h\sigma_h^2\Big(\sum_{i\in U_h}w_i-N_h\Big)\Big],\\
\phi_{CS}&=\phi_{SS}+\tfrac{1}{N^2}\textstyle\sum_h\sigma_{xh}^2(\beta_h-\beta)^2\Big(\sum_{i\in U_h}w_i-N_h\Big),\\
\phi_{SC}&=\phi_{SS}+V_s\Big(\textstyle\sum_h\alpha_h\Big(\tfrac{\hat{N}_h}{\hat{N}}-\tfrac{N_h}{N}\Big)\Big),\\
\phi_{CC}&=\phi_{CS}+V_s\Big(\textstyle\sum_h[\alpha_h+\mu_{xh}(\beta_h-\beta)]\Big(\tfrac{\hat{N}_h}{\hat{N}}-\tfrac{N_h}{N}\Big)\Big),\\
\phi_{S0}&=\phi_{SS}+\tfrac{1}{N^2}V_s\Big(\textstyle\sum_h\alpha_h(\hat{N}_h-N_h)\Big),\\
\phi_{C0}&=\phi_{CS}+\tfrac{1}{N^2}V_s\Big(\textstyle\sum_h\Big[\alpha_h+\mu_{xh}(\beta_h-\beta)\Big](\hat{N}_h-N_h)\Big),
\end{aligned}$$

and

$$\beta=\frac{\sum_h N_h\beta_h\sigma_{xh}^2}{\sum_h N_h\sigma_{xh}^2}.$$

The proof is given in the Appendix. The condition $n_h/N_h\to 0$ is not needed when sampling is with replacement or simple random sampling without replacement.

The following are some conclusions for the relative efficiencies among estimators in (1.1)$-$(1.6). For two estimators $\hat{Y}_u$ and $\hat{Y}_v$, $\hat{Y}_u\succeq\hat{Y}_v$ denotes that $\hat{Y}_u$ is asymptotically as efficient as $\hat{Y}_v$, $\hat{Y}_u\succ\hat{Y}_v$ denotes that $\hat{Y}_u$ is asymptotically more efficient than $\hat{Y}_v$, and $\hat{Y}_u\cong\hat{Y}_v$ denotes that $\hat{Y}_u$ and $\hat{Y}_v$ are asymptotically equivalent.

1. Asymptotically, no other estimator is more efficient than $\hat{Y}_{SS}$, even when regression lines across strata are the same. Some other estimators may be asymptotically as efficient as $\hat{Y}_{SS}$ under some special situations.

2. When $\alpha_h = \alpha \neq 0$ and $\beta_h = \beta$ for all h, $\hat{Y}_{CC} \cong \hat{Y}_{CS} \cong \hat{Y}_{SC} \cong \hat{Y}_{SS} \succ \hat{Y}_{C0} \cong \hat{Y}_{S0}$. When $\alpha_h = 0$ and $\beta_h = \beta$ for all h, all estimators in (1.1)$-$(1.6) are asymptotically equivalent.
3. When $\beta_h = \beta$ for all h, $\hat{Y}_{C0} \cong \hat{Y}_{S0}$, $\hat{Y}_{CC} \cong \hat{Y}_{SC}$, $\hat{Y}_{CS} \cong \hat{Y}_{SS}$, $\hat{Y}_{SS} \succeq \hat{Y}_{S0}$, and $\hat{Y}_{SS} \succeq \hat{Y}_{SC}$. When $\hat{N} = N$, $\hat{Y}_{S0} \cong \hat{Y}_{SC}$. No general conclusion can be made about $\hat{Y}_{S0}$ and $\hat{Y}_{SC}$.
4. When there exist h_1 and h_2 such that $\beta_{h_1} \neq \beta_{h_2}$, $\hat{Y}_{S0} \succ \hat{Y}_{C0}$, $\hat{Y}_{SC} \succ \hat{Y}_{CC}$ and $\hat{Y}_{SS} \succ \hat{Y}_{CS}$. Thus, applying the same ratio, the estimators using separate lines are more efficient than those using a single line.
5. When $\alpha_h = \alpha \neq 0$ for all h, $\hat{Y}_{SC} \cong \hat{Y}_{SS} \succ \hat{Y}_{S0}$. When $\alpha_h = 0$ for all h, $\hat{Y}_{SC} \cong \hat{Y}_{SS} \cong \hat{Y}_{S0}$.
6. When there exist h_1 and h_2 such that $\alpha_{h_1} \neq \alpha_{h_2}$, $\hat{Y}_{SS} \succ \hat{Y}_{SC}$ and $\hat{Y}_{SS} \succ \hat{Y}_{S0}$ unless $\hat{N}_{h_1} = N_{h_1}$ and $\hat{N}_{h_2} = N_{h_2}$.

There is a definite conclusion. Asymptotically, $\hat{Y}_{SS}$ is the winner among all estimators in (1.1)$-$(1.6) even when the regression lines across strata are the same. This phenomenon has been noticed by previous researchers (see, e.g., Fuller (2009)). One explanation is that, when all n_h's are large, the information about the equality of regression lines does not help to improve model-assisted estimators of the forms (1.1)$-$(1.6) that are robust against model violation. Some model-based estimators may be more efficient than $\hat{Y}_{SS}$ when all regression lines are the same, but they are not robust against model violation. Furthermore, the situation is quite different if some n_h's are small, as the results in the next section indicate.

Another point is that the result in Theorem 1 still holds if $\hat{\beta}_h$ in (1.8) is replaced by a different consistent estimator of β_h. Fuller (2009) indicates that, given a sampling design, an optimal slope estimator can be derived (e.g., (2.4.2) in Fuller (2009)) that is asymptotically more efficient than $\hat{\beta}_h$ in (1.8) in terms of estimating β_h. For the estimation of Y, however, the efficiency of $\hat{\beta}_h$ is a second-order asymptotic effect. It does not improve the asymptotic efficiency of $\hat{Y}_{SS}$.

3. Asymptotic Results for Large H and Small n_h's

To increase efficiency, many surveys involve a large number of strata. To reduce the cost, a few units are sampled from each stratum. See, for example, Krewski and Rao (1981). We consider a different asymptotic setting: all n_h's are bounded by a fixed constant and $H \to \infty$.

For estimators based on separate regressions, $\hat{Y}_{SS}$, $\hat{Y}_{SC}$, and $\hat{Y}_{S0}$, each n_h has to be sufficiently large so that a regression within stratum h can be fitted. Since model (2.2) involves a two term regression function, we require $n_h > 2$ for

$\hat{Y}_{SS}$, $\hat{Y}_{SC}$, and $\hat{Y}_{S0}$. If x_i is multivariate with dimension p, then $n_h > p+1$ for each h is required.

When n_h is small, it is not possible to obtain a consistent estimator for a parameter related only with stratum U_h. However, quantities such as totals over all strata can be consistently estimated. Consider $\hat{Y}_{SS}$ as an example. Let $E_{m|x}$ be the expectation under (2.2), conditional on the x_i's. Then $E_{m|x}(y_i) = \alpha_h + \beta_h x_i$, $i \in U_h$, and

$$\begin{aligned} E_{m|x}(\hat{Y}_{SS}) &= \sum_h \Big[\frac{N_h E_{m|x}(\hat{Y}_h)}{\hat{N}_h} + E_{m|x}(\hat{\beta}_h)\Big(X_h - \frac{N_h \hat{X}_h}{\hat{N}_h}\Big)\Big] \\ &= \sum_h \Big[\frac{N_h \alpha_h \hat{N}_h}{\hat{N}_h} + \frac{N_h(\beta_h \hat{X}_h)}{\hat{N}_h} + \beta_h\Big(X_h - \frac{N_h \hat{X}_h}{\hat{N}_h}\Big)\Big] \\ &= \sum_h (N_h\alpha_h + \beta_h X_h) \\ &= \sum_h E_{m|x}(Y_h) \\ &= E_{m|x}(Y). \end{aligned}$$

Thus, $\hat{Y}_{SS}$ is unbiased with respect to model (2.2). Since $\hat{Y}_{SS}$ is a sum (over h) of independent random variables having finite means and variances (under some conditions), we can apply Liapounov's Central Limit Theorem to establish the asymptotic normality of $\hat{Y}_{SS}$. This is the approach considered by Krewski and Rao (1981). We need some conditions.

(D1) There exists M_1 such that $|\alpha_h| + |\beta_h \mu_{xh}| < M_1$ for any h.

(D2) There exists M_2 such that $1/M_2 < \sigma^2_{xh} = V_m(x_i) < M_2$ and $1/M_2 < \sigma^2_h = V_m(\epsilon_i) < M_2$ for any h, and there exists $\delta > 0$ such that, as $H \to \infty$,

$$\frac{\sum_{h=1}^{H} E_m \Big| R_h \hat{Y}_h + Q_h(X_h - R_h \hat{X}_h) - E_m[R_h \hat{Y}_h + Q_h(X_h - R_h \hat{X}_h)]\Big|^{2+\delta}}{\Big\{\sum_{h=1}^{H} V_m[R_h \hat{Y}_h + Q_h(X_h - R_h \hat{X}_h)]\Big\}^{2+\delta}} \to 0,$$

where $R_h = 1$, $\hat{N}/N$, or $\hat{N}_h/N_h$, and $Q_h = \hat{\beta}_h$ or $\hat{\beta}$.

(D3) As $H \to \infty$,

$$\frac{\sum_h \beta_h \sum_{i \in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2}{\sum_h \sum_{i \in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2} \to_p \beta,$$

$$E_s\Big\{\frac{\sqrt{n}}{N} \sum_{h=1}^{H} [\alpha_h + (\beta_h - \beta)\mu_{xh}]\Big(\frac{N\hat{N}_h}{\hat{N}} - N_h\Big)\Big\} \to 0.$$

Theorem 2. *If model* (2.2), (C3), *and* (D1)−(D3) *hold, then*

$$\frac{\hat{Y}_{uv}-Y}{N\sqrt{\phi_{uv}}}\xrightarrow{d}N(0,1),$$

where $u=C$ *or* S, $v=0$, C, *or* S, *and*

$$\phi_{C0}=\frac{1}{N^2}\sum_h V_s\Big(\Lambda_h\hat{N}_h\Big)+\frac{1}{N^2}\Big[\sum_h\Gamma_h\Big(\sum_{i\in U_h}w_i-N_h\Big)\Big],$$

$$\phi_{S0}=\frac{1}{N^2}E_sE_m\Big[\sum_h\sum_{i\in S_h}\sigma_h^2\Big(\zeta_{hi}+w_i-1\Big)^2\Big]+\frac{1}{N^2}\sum_h\sigma_h^2(N_h-n_h)+\frac{1}{N^2}V_s\Big(\sum_h\alpha_h\hat{N}_h\Big),$$

$$\phi_{CC}=V_s\Big(\sum_h\Lambda_h\frac{\hat{N}_h}{\hat{N}}\Big)+\frac{1}{N^2}\Big[\sum_h\Gamma_h\Big(\sum_{i\in U_h}w_i-N_h\Big)\Big],$$

$$\phi_{SC}=\frac{1}{N^2}E_sE_m\Big[\sum_h\sum_{i\in S_h}\sigma_h^2\Big(\zeta_{hi}+w_i-1\Big)^2\Big]+\frac{1}{N^2}\sum_h\sigma_h^2(N_h-n_h)+V_s\Big(\sum_h\alpha_h\frac{\hat{N}_h}{\hat{N}}\Big),$$

$$\phi_{CS}=\frac{1}{N^2}E_s\Big[\sum_h\Gamma_h\Big(\frac{N_h^2}{\hat{N}_h^2}\sum_{i\in S_h}w_i^2-N_h\Big)\Big],$$

$$\phi_{SS}=\frac{1}{N^2}\sum_h\sigma_h^2E_sE_m\Big[\sum_{i\in S_h}\Big(\psi_{hi}+\frac{N_h}{\hat{N}_h}w_i-1\Big)^2\Big]+\frac{1}{N^2}\sum_h\sigma_h^2(N_h-n_h),$$

with $\Lambda_h=\alpha_h+(\beta_h-\beta)\mu_{xh}$, $\Gamma_h=\sigma_h^2+(\beta_h-\beta)^2\sigma_{xh}^2$, $\bar{x}_h=\hat{X}_h/\hat{N}_h$,

$$\zeta_{hi}=\frac{-(x_i-\bar{x}_h)w_i}{\sum\limits_{i\in S_h}(x_i-\bar{x}_h)^2w_i}\Big(\hat{X}_h-X_h\Big),\quad \textit{and}\quad \psi_{hi}=\frac{-(x_i-\bar{x}_h)w_i}{\sum\limits_{i\in S_h}(x_i-\bar{x}_h)^2w_i}\Big(\frac{N_h\hat{X}_h}{\hat{N}_h}-X_h\Big).$$

The proof is given in the Appendix. Note that $\sum_{i\in S_h}\psi_{hi}=0$ and $\sum_{i\in S_h}\zeta_{hi}=0$ for any h. If $\sum_{i\in S_h}(x_i-\bar{x}_h)w_i^2=0$ for any h, which occurs for example when w_i is a constant within stratum h, then

$$\phi_{SS}=\frac{1}{N^2}\Big\{\sum_h\sigma_h^2E_sE_m\sum_{i\in S_h}\psi_{hi}^2\Big\}-\sum_h(\beta_h-\beta)^2\sigma_{xh}^2E_s\Big(\frac{N_h^2}{\hat{N}_h^2}\sum_{i\in S_h}w_i^2-N_h\Big)+\phi_{CS}. \tag{3.1}$$

As for the relative efficiencies among estimators in (1.1)−(1.6), we have the following discussion.

1. With simple random sampling (SRS) within each stratum, $\hat{N}_h=N_h$, $\hat{N}=N$, $\hat{Y}_{C0}=\hat{Y}_{CC}=\hat{Y}_{CS}$, and $\hat{Y}_{S0}=\hat{Y}_{SC}=\hat{Y}_{SS}$. Thus, we compare $\hat{Y}_{CS}$ and $\hat{Y}_{SS}$. Under SRS within each stratum, $\sum_{i\in S_h}\psi_{hi}w_i=0$ for any h. If $\beta_h=\beta$ for

any h, then it follows from (3.1) that

$$\phi_{SS} = \frac{1}{N^2}\Big\{\sum_h \sigma_h^2 E_s E_m \sum_{i\in S_h} \psi_{hi}^2\Big\} + \phi_{CS} > \phi_{CS} \tag{3.2}$$

and, hence, $\hat{Y}_{CS} \succ \hat{Y}_{SS}$. However, it also follows from (3.1) that, when $(\beta_h - \beta)^2$ increases to ∞, the ratio ϕ_{SS}/ϕ_{CS} increases to infinity, so $\hat{Y}_{SS} \succ \hat{Y}_{CS}$ if $(\beta_h - \beta)^2$ is sufficiently large for some h.

2. Under a general sampling design, there is no definite conclusion about the relative efficiency among $\hat{Y}_{C0}$, $\hat{Y}_{CC}$, and $\hat{Y}_{CS}$, because ϕ_{C0}, ϕ_{CC}, and ϕ_{CS} involve expectations and variances with respect to sampling. For example, there is no definite conclusion on which of

$$\frac{1}{N^2}\sum_h V_s\left(\Lambda_h \hat{N}_h\right) \quad \text{and} \quad V_s\Big(\sum_h \Lambda_h \frac{\hat{N}_h}{\hat{N}}\Big)$$

is larger. Similarly, we are not able to draw a definite conclusion about the relative efficiency among $\hat{Y}_{S0}$, $\hat{Y}_{SC}$, and $\hat{Y}_{SS}$.

3. If $\beta_h = \beta$ and $\sum_{i\in S_h} \psi_{hi} w_i = 0$ for any h, then $\hat{Y}_{CS} \succ \hat{Y}_{SS}$. This is because (3.1) holds and, if $\beta_h = \beta$, then (3.2) also holds. Similarly, if $\sum_{i\in S_h} \zeta_{hi} w_i = 0$ for any h, then $\hat{Y}_{CC} \succ \hat{Y}_{SC}$ and $\hat{Y}_{C0} \succ \hat{Y}_{S0}$. If we further assume that $\alpha_h = \alpha \neq 0$ for any h, then $\hat{Y}_{CC} \succ \hat{Y}_{C0}$ and $\hat{Y}_{SC} \succ \hat{Y}_{S0}$. When $\alpha_h = 0$ for any h, $\hat{Y}_{CC} \cong \hat{Y}_{C0}$ and $\hat{Y}_{SC} \cong \hat{Y}_{S0}$.

4. If the β_h's are not all the same, $(\beta_h - \beta)^2$ increases to ∞ for some h, ϕ_{C0}, ϕ_{CC}, and ϕ_{CS} all diverge to infinity. On the other hand, none of ϕ_{C0}, ϕ_{CC}, and ϕ_{CS} depends on $(\beta_h - \beta)^2$. This means that $\hat{Y}_{S0} \succeq \hat{Y}_{C0}$, $\hat{Y}_{SC} \succeq \hat{Y}_{CC}$, and $\hat{Y}_{SS} \succeq \hat{Y}_{CS}$ when β_h's are different and the differences are large enough.

The result is quite different from that in Section 2 where all n_h's are large and H is fixed. In Section 2 we found a clear winner in terms of asymptotic efficiency, $\hat{Y}_{SS}$, regardless of the scenario, whereas for small n_h's, there is no winner even if we use SRS within each stratum. The asymptotic relative efficiency between $\hat{Y}_{CS}$ and $\hat{Y}_{SS}$ depends on how the β_h's differ.

4. Empirical Results

In this section, we present some empirical results using simulated data sets. We also discuss variance estimation and examine related confidence intervals.

4.1. Simulation results for large n_h's and fixed H

We considered a population with $H = 4$ strata. In stratum h, we generated independent $x_i \sim \Gamma(\vartheta_h, \theta_h)$, $i = 1, \ldots, N_h = 2000$, where $\Gamma(\vartheta_h, \theta_h)$ is the gamma distribution with shape parameter ϑ_h and scale parameter θ_h, $y_i = \alpha_h + \beta_h x_i + \epsilon_i$ and $z_i = |30 + x_i + \nu_i|$, where ϵ_i's and ν_i's are independent and normally distributed with mean 0 and variance 50, and they are independent of x_i's. Values from different strata were independently generated. The values of the parameters of the gamma distribution in each stratum were as follows.

ϑ_1	θ_1	ϑ_2	θ_2	ϑ_3	θ_3	ϑ_4	θ_4
4	10	3	8	2	9	3	11

For each h, we adopted the Rao-Hartley-Cochran (RHC) sampling scheme to obtain a sample of size $n_h = 200$ without replacement from U_h using z_i's as the weights (see, e.g., Sampford (1967)).

1. Divide U_h into n_h subgroups $U_{h1}, \ldots, U_{hn_h}$, where each group consists of $k_h = N_h/n_h$ units.
2. Select unit j from each subgroup U_{hq} using $\{z_i : \ i \in U_{hq}\}$ as weights. Under this sampling scheme, $w_j^{-1} = z_j / \sum_{i \in U_{hq}} z_i$.
3. Samples across strata are independently sampled.

Table 1 reports the value of (α_h, β_h) and, based on 2000 simulations, the biases and standard deviations (SD) of six estimators, two estimated SD, $\widehat{\text{SD}}_S$ obtained by substituting asymptotic variance formulas derived in Section 3 and $\widehat{\text{SD}}_B$ based on the bootstrap variance estimator described in Bickel and Freedman (1984), Shao and Tu (1995), and Antal and Tillé (2011) with $B = 300$, together with the coverage probability, CP_u, of the approximate 95% confidence interval for the population total Y: [estimated total$-1.96\widehat{\text{SD}}_u$, estimated total$+1.96\widehat{\text{SD}}_u$], where $u = S$ or B.

The following observations summarize the results in Table 1.

1. All estimators have negligible biases.
2. The simulation results support the asymptotic theory in Theorem 1. When β_h's and α_h's are different, $\hat{Y}_{SS}$ has the best performance. When α_h's and/or β_h's are the same, other estimators are comparable with $\hat{Y}_{SS}$. If $\hat{Y}_{SS}$ is not the best, then the difference between the SD of $\hat{Y}_{SS}$ and the SD of the best estimator is negligible.
3. Substitution and bootstrap variance estimators perform reasonably well. When the estimator of SD performs well, so does the 95% approximate confidence interval. In a few cases the substitution estimator $\widehat{\text{SD}}_S$ overestimates substantially, which results in a too large CP. The bootstrap estimator $\widehat{\text{SD}}_B$ is more stable, but always overestimates in our simulation.

Table 1. Simulation Results for Six Estimators (Large n_h's and Fixed H).

h	α_h	β_h	Y		$\hat{Y}_{C0}$	$\hat{Y}_{S0}$	$\hat{Y}_{CC}$	$\hat{Y}_{SC}$	$\hat{Y}_{CS}$	$\hat{Y}_{SS}$
1	0	1	230530	Bias	-23.6	-21.6	-23.9	-22.1	-23.5	-19.0
2	0	1		SD	2055	2056	2057	2058	2058	2061
3	0	1		$\widehat{\text{SD}}_S$	1982	1980	1982	1979	1982	1978
4	0	1		$\widehat{\text{SD}}_B$	2077	2076	2077	2076	2076	2077
				CP_S	93.7%	93.2%	93.4%	93.3%	93.5%	93.1%
				CP_B	94.9%	94.8%	94.8%	94.6%	94.9%	94.6%
1	0	1	245886	Bias	17.5	15.1	20.7	15.7	36.1	18.0
2	0	1.5		SD	2043	1955	2046	1957	2234	1958
3	0	1.2		$\widehat{\text{SD}}_S$	2359	1982	2357	1982	2289	1980
4	0	0.8		$\widehat{\text{SD}}_B$	2190	2074	2195	2075	2358	2077
				CP_S	97.1%	95.1%	97.0%	94.9%	95.6%	95.1%
				CP_B	96.2%	96.1%	96.4%	96.1%	96.4%	96.1%
1	5	1	332923	Bias	-67.4	-67.1	-53.1	-52.8	-58.6	-57.8
2	15	1		SD	2289	2290	2015	2015	1971	1975
3	20	1		$\widehat{\text{SD}}_S$	2268	2265	2035	2032	1983	1980
4	25	1		$\widehat{\text{SD}}_B$	2374	2373	2133	2133	2080	2081
				CP_S	95.0%	95.0%	94.8%	95.0%	94.5%	94.8%
				CP_B	95.2%	95.4%	95.6%	96.1%	96.0%	95.7%
1	5	1	492270	Bias	55.6	-7.1	57.8	-6.2	142.5	10.9
2	10	1.5		SD	2431	2302	2259	2055	3374	1985
3	15	2		$\widehat{\text{SD}}_S$	4249	2274	4135	2039	3586	1980
4	20	2.5		$\widehat{\text{SD}}_B$	2717	2366	2547	2121	3595	2072
				CP_S	97.0%	95.0%	99.9%	94.3%	96.1%	93.9%
				CP_B	96.8%	95.2%	97.2%	94.7%	96.4%	95.1%

4.2. Simulation results for large H and small n_h's

We considered a population with $H = 600$ strata. In stratum h, we independently generated $x_i \sim \Gamma(\vartheta_h, \theta_h)$, $i = 1, \ldots, N_h = 60$, where ϑ_h and θ_h were independently generated from the uniform $U(6, 8)$. In stratum h, y_i's and z_i's were generated as in Section 4.2. Values from different strata were independently generated. In stratum h, a sample of size $n_h = 4$ was taken without replacement within U_h, using SRS or RHC as described in Section 4.1. Samples across strata were independently sampled. The regression parameters were chosen as follows: $(\alpha_h, \beta_h) = (a_1, b_1)$, $h = 1, \ldots, H/2$, and $(\alpha_h, \beta_h) = (a_2, b_2)$, $h = H/2+1, \ldots, H$, where values of a_k and b_k, $k = 1, 2$, are shown in Table 2.

Table 2 reports the same quantities as those in Table 1 (for SRS or RHC), except that there is only one estimated SD (and hence one CP) for each estimator. For $\hat{Y}_{C0}$, $\hat{Y}_{CC}$, and $\hat{Y}_{CS}$, the estimated SD, $\widehat{\text{SD}}$, is based on the bootstrap with $B = 300$ and bootstrap sample size $n_h - 1$ (see, for example, McCarthy and Snowden (1985), for the reason of using $n_h - 1$). This bootstrap method, however,

Table 2. Simulation Results for Six Estimators (Large H and Small n_h's).

k	a_k	b_k	Y	Sampling		$\hat{Y}_{C0}$	$\hat{Y}_{S0}$	$\hat{Y}_{CC}$	$\hat{Y}_{SC}$	$\hat{Y}_{CS}$	$\hat{Y}_{SS}$
1	20	1.5	3372400	SRS	Bias					-111.7	49.5
2	20	1.5			SD					5003	7199
					$\widehat{\text{SD}}$					5157	7350
					CP					95.6%	95.8%
				RHC	Bias	-39.0	20.8	-268.7	-216.8	-277.8	-225.1
					SD	6471	6803	5311	5789	5356	7182
					$\widehat{\text{SD}}$	6421	7315	5303	6517	5328	7146
					CP	95.2%	95.0%	94.8%	95.4%	95.0%	94.4%
1	20	1.5	4705570	SRS	Bias					-1787.0	-428.3
2	20	3			SD					11052	7317
					$\widehat{\text{SD}}$					11462	7298
					CP					96.8%	94.8%
				RHC	Bias	69.6	263.6	29.2	222.7	172.5	826.7
					SD	8042	6768	7125	5834	11489	7272
					$\widehat{\text{SD}}$	8176	7465	7361	6560	12073	7162
					CP	94.8%	95.6%	95.2%	97.2%	95.4%	94.6%
1	20	1.5	3555523	SRS	Bias					154.9	199.0
2	30	1.5			SD					5107	6901
					$\widehat{\text{SD}}$					5157	7358
					CP					95.2%	96.2%
				RHC	Bias	-676.8	-642.4	-596.9	-562.7	-554.5	-547.5
					SD	7108	7321	5622	5928	5621	7136
					$\widehat{\text{SD}}$	7022	7603	5394	6619	5342	7144
					CP	95.4%	94.8%	94.6%	96.8%	94.0%	95.2%
1	20	1.5	4852641	SRS	Bias					-608.9	-90.7
2	30	3			SD					10908	7328
					$\widehat{\text{SD}}$					11358	7342
					CP					96%	95.1%
				RHC	Bias	585.5	391.3	389.2	192.3	982.4	301.9
					SD	7547	7135	6327	5832	11339	7471
					$\widehat{\text{SD}}$	8393	7770	7105	6662	12010	7122
					CP	96.8%	95.0%	96.8%	96.3%	96.5%	94.2%

does not work for $\hat{Y}_{S0}$, $\hat{Y}_{SC}$, and $\hat{Y}_{SS}$, because it is highly possible that the bootstrap sample in stratum h contains only one or two points (y_i, x_i) from S_h when n_h is small, which prevents us from regression fitting. Thus, we consider substitution estimators based on the result in Theorem 2:

$$\hat{\phi}_{S0} = \frac{1}{N^2}\sum_h \sum_{i\in S_h} \hat{\sigma}_h^2 (\zeta_{hi}+w_i-1)^2 + \frac{1}{N^2}\sum_h \hat{\sigma}_h^2(N_h-n_h) + \frac{1}{N^2}\Big[\sum_h \hat{\alpha}_h(\hat{N}_h-N_h)\Big]^2,$$

$$\hat{\phi}_{SC} = \frac{1}{N^2}\sum_h \sum_{i\in S_h} \hat{\sigma}_h^2(\zeta_{hi}+w_i-1)^2 + \frac{1}{N^2}\sum_h \hat{\sigma}_h^2(N_h-n_h) + \Big[\sum_h \hat{\alpha}_h\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big)\Big]^2,$$

$$\hat{\phi}_{SS} = \frac{1}{N^2}\sum_h \hat{\sigma}_h^2 \sum_{i\in S_h}\Big(\psi_{hi} + \frac{N_h}{\hat{N}_h}w_i - 1\Big)^2 + \frac{1}{N^2}\sum_h \hat{\sigma}_h^2(N_h-n_h),$$

where ζ_{hi} and ψ_{hi} are given in Theorem 2.

Under SRS, $\hat{Y}_{C0} = \hat{Y}_{CC} = \hat{Y}_{CS}$ and $\hat{Y}_{S0} = \hat{Y}_{SC} = \hat{Y}_{SS}$. Thus, the results are shown for $\hat{Y}_{CS}$ and $\hat{Y}_{SS}$ only.

The following is a summary of the results in Table 2.

1. The biases of all estimators are negligible, although they are relatively large compared with those in Table 1.
2. In terms of SD, $\hat{Y}_{SS}$ is no longer always the best. When regression lines are all the same, the estimators using combined regression are much more efficient than those using separate regression lines, which is quite different from the situation with large n_h's. On the other hand, when the slopes of regression lines are different, the estimators using separate regression lines are more efficient than those using combined regression. Under RHC, among the three estimators using separate regression lines, $\hat{Y}_{SS}$ is no longer the best, $\hat{Y}_{SC}$ is.
3. The substitution variance estimators for $\hat{Y}_{S0}$, $\hat{Y}_{SC}$, and $\hat{Y}_{SS}$ and the bootstrap variance estimator for $\hat{Y}_{C0}$, $\hat{Y}_{CC}$, and $\hat{Y}_{CS}$ perform well and result in good CP of confidence intervals.

5. Discussion

We study the asymptotic efficiencies of estimators (1.1)−(1.6) in two different asymptotic settings under stratified sampling with H strata and model (2.2). In the case where H is fixed and all stratum sample sizes tend to infinity, the estimator $\hat{Y}_{SS}$ in (1.6) is asymptotically the most efficient estimator, regardless of whether regression models in different strata are the same or not. When all stratum sample sizes are small and H tends to infinity, however, no general conclusion can be made.

When H is fixed and all stratum sample sizes are large, it is not difficult to derive design-based consistent variance estimators for all six estimators in (1.1)−(1.6). In fact, it can be shown that the bootstrap variance estimators used in Section 4.1 are design-based consistent. For the case where all stratum sample sizes are small and H is large, it can still be shown that the bootstrap variance estimators for $\hat{Y}_{C0}$, $\hat{Y}_{CC}$ and $\hat{Y}_{CS}$ described in Section 4.2 are design-based consistent.

There are estimators of Y other than those in (1.1)−(1.6). For example, an anonymous referee commented that we could consider regression adjustments instead of sample size adjustments, which lead to an estimator of the form

$$\sum_{h=1}^{H}[N_h\hat{Y}_h - \hat{\beta}_h(X_h - \hat{X}_h) - \hat{\alpha}_h(N_h - \hat{N}_h)]$$

with some estimators $\hat{\alpha}_h$ and $\hat{\beta}_h$. Asymptotic results can be similarly derived.

Acknowledgements

We would like to thank three referees and an associate editor for providing helpful comments and suggestions. The research was partially supported by an NSF grant.

Appendix

Proof of Theorem 1. Let P, E, and V be the the probability, expectation and variance under both model (2.2) and design. By the fact that $y_i = \alpha_h + \beta_h x_i + \epsilon_i$, $i \in S_h$ and the definition of $\hat{Y}_{SS}$, we have

$$\frac{\hat{Y}_{SS} - Y}{N} = \sum_h \Big[(\beta_h - \hat{\beta}_h)\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big) + \Big(\frac{1}{\hat{N}_h}\sum_{i\in S_h} w_i\epsilon_i - \frac{\sum_{i\in U_h}\epsilon_i}{N_h}\Big)\Big]\frac{N_h}{N}. \quad \text{(A.1)}$$

First, notice that

$$E\Big[\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big] = E_s E_m\Big[\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big] = 0 \quad \text{and} \quad E\Big[\frac{1}{\hat{N}_h}\sum_{i\in S_h} w_i\epsilon_i - \frac{\sum_{i\in U_h}\epsilon_i}{N_h}\Big] = 0.$$

By (C2) and Liapounov's Central Limit Theorem,

$$\frac{1}{\sqrt{\psi_{1h}}}\Big[\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big] \xrightarrow{d} N(0,1),$$

where

$$\begin{aligned}
\psi_{1h} &= V\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big)\\
&= V_s E_m\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big) + E_s V_m\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big)\\
&= E_s V_m\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big)\\
&= \sigma_{xh}^2 E_s\Big[\sum_{i\in S_h}\Big(\frac{w_i}{\hat{N}_h} - \frac{1}{N_h}\Big)^2 + \sum_{i\in U_h/S_h}\frac{1}{N_h^2}\Big]
\end{aligned}$$

$$= \sigma_{xh}^2 \Big[E_s \Big(\sum_{i \in S_h} \frac{w_i^2}{\hat{N}_h^2} \Big) - \frac{1}{N_h} \Big],$$

and the third equality follows since $E_m\Big(\hat{X}_h/\hat{N}_h - X_h/N_h\Big) = 0$. By (C3), we have $\sum_{i\in S_h} w_i^2/\hat{N}_h^2 < 1/(n_h M)$. It is easy to show that $N_h/\hat{N}_h \to_p 1$. By the Dominated Convergence Theorem and Liapounov's Central Limit Theorem, we have

$$\frac{1}{\sqrt{\phi_{1h}}} \Big[\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h} \Big] \xrightarrow{d} N(0,1), \tag{A.2}$$

where

$$\phi_{1h} = \frac{\sigma_{xh}^2}{N_h^2} \Big[E_s \Big(\sum_{i \in S_h} w_i^2 \Big) - N_h \Big] = \frac{\sigma_{xh}^2}{N_h^2} \Big(\sum_{i \in U_h} w_i - N_h \Big),$$

and $\sigma_{xh}^2 = E[x_i]$, $i \in U_h$. Similarly, we have

$$\frac{1}{\sqrt{\phi_{2h}}} \Big[\frac{1}{\hat{N}_h} \sum_{i \in S_h} w_i \epsilon_i - \frac{\sum_{i \in U_h} \epsilon_i}{N_h} \Big] \xrightarrow{d} N(0,1),$$

where $\phi_{2h} = \sigma_h^2 (\sum_{i\in U_h} w_i - N_h)/N_h^2$. Therefore, by (A.1),

$$\frac{\hat{Y}_{SS} - Y}{N\sqrt{\phi_{SS}}} = \sum_h \frac{\sqrt{\phi_{2h}}}{\sqrt{\phi_{SS}}} \Big[\sqrt{\frac{\phi_{1h}}{\phi_{2h}}} \frac{\beta_h - \hat{\beta}_h}{\sqrt{\phi_{1h}}} \Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h} \Big) + \frac{1}{\sqrt{\phi_{2h}}} \Big(\sum_{i \in S_h} \frac{w_i \epsilon_i}{\hat{N}_h} - \sum_{i \in U_h} \frac{\epsilon_i}{N_h} \Big) \Big] \frac{N_h}{N},$$

where $\phi_{SS} = \sum_h \phi_{2h} N_h^2/N^2 = \Big[\sum_h \sigma_h^2 \Big(\sum_{i\in U_h} w_i - N_h \Big) \Big] / N^2$. Using $\phi_{1h}/\phi_{2h} = \sigma_{xh}/\sigma_h$, (A.2), and $\hat{\beta}_h - \beta_h \to_p 0$, we obtain

$$\frac{\hat{Y}_{SS} - Y}{N\sqrt{\phi_{SS}}} = \sum_h \frac{\sqrt{\phi_{2h}}}{\sqrt{\phi_{SS}}} \Big[\frac{1}{\sqrt{\phi_{2h}}} \Big(\frac{1}{\hat{N}_h} \sum_{i \in S_h} w_i \epsilon_i - \frac{\sum_{i \in U_h} \epsilon_i}{N_h} \Big) \Big] \frac{N_h}{N} + o_p(1).$$

Therefore, by Slutsky's Theorem,

$$\frac{\hat{Y}_{SS} - Y}{N\sqrt{\phi_{SS}}} \xrightarrow{d} N(0,1).$$

From the conditions, there exists b_h such that

$$\frac{\sum_{i\in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2}{\sum_h \sum_{i \in S_h} w_i (x_i - \hat{X}_h/\hat{N}_h)^2} \to_p b_h.$$

According to the definition of $\hat{\beta}$, we have $\hat{\beta} \to_p \beta$, where $\beta = \sum_h b_h \beta_h$. Similarly, we can obtain

$$\frac{\hat{Y}_{CS} - Y}{N\sqrt{\phi_{CS}}} \xrightarrow{d} N(0,1),$$

where $\phi_{CS} = \sum_h \tau_h N_h^2/N^2$ and $\tau_h = (\beta_h - \beta)^2\phi_{1h} + \phi_{2h}$. Therefore,

$$\phi_{CS} = \frac{1}{N^2}\sum_h \sigma_{xh}^2(\beta_h - \beta)^2\Big(\sum_{i\in U_h} w_i - N_h\Big) + \phi_{SS}.$$

For $\hat{Y}_{SC}$, we have

$$\frac{\hat{Y}_{SC}-Y}{N} = \sum_h \alpha_h\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big) + \sum_h(\beta_h - \hat{\beta}_h)\Big(\frac{\hat{X}_h}{\hat{N}} - \frac{X_h}{N}\Big) + \sum_h\Big(\sum_{i\in S_h}\frac{w_i\epsilon_i}{\hat{N}} - \sum_{i\in U_h}\frac{\epsilon_i}{N}\Big).$$

Similar to the previous proof, by the fact that $\hat{\beta}_h \to_p \beta_h$, the asymptotic distribution of $(\hat{Y}_{SC} - Y)/N$ is as same as T, where

$$T = \sum_h \alpha_h\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big) + \sum_h\Big(\frac{1}{\hat{N}}\sum_{i\in S_h} w_i\epsilon_i - \frac{1}{N}\sum_{i\in U_h}\epsilon_i\Big).$$

Therefore, under (C1)−(C3),

$$\frac{T - E[T]}{\sqrt{\phi_{SC}}} \xrightarrow{d} N(0,1),$$

where $E[T] = E_s[\sum_h \alpha_h(\hat{N}_h/\hat{N} - N_h/N)]$ and $\phi_{SC} = \phi_{SS} + V_s[\sum_h \alpha_h(\hat{N}_h/\hat{N} - N_h/N)]$. Under (C3), $N_h/M < \hat{N}_h < N_h M$, $N_h^2/(Mn_h) < \sum_{i\in S_h} w_i^2 < MN_h^2/n_h$. Therefore $N/M < \hat{N} < NM$ and $\sum_h(N_h^2/(n_h M)) < \sum_h\sum_{i\in S_h} w_i^2 < M\sum_h (N_h^2/n_h)$. Then

$$\Big|\sum_h \alpha_h\Big(\frac{N\hat{N}_h}{\hat{N}N_h} - 1\Big)\Big| < \sum_h |\alpha_h|\Big(M^2 + 1\Big). \tag{A.3}$$

Since $n_h/N_h \to 0$ then, for any fixed $\epsilon_0 > 0$, when n_h is large enough,

$$\sqrt{\sum_h \sigma_h^2\Big(\frac{N_h}{n_h M} - 1\Big)} > \epsilon_0,$$

$$\phi_{SC} \ge \phi_{SS} > \frac{1}{N^2}\sum_h \sigma_h^2\Big(\frac{N_h^2}{n_h M} - N_h\Big) > 0.$$

Then

$$\frac{|E[T]|}{\sqrt{\phi_{SC}}} \le \frac{E_s\Big[\Big|\sum_h \alpha_h\Big(N\hat{N}_h/\hat{N}N_h - 1\Big)\Big|\Big]}{\sqrt{\sum_h \sigma_h^2\Big(N_h/n_h M - 1\Big)}} \le \frac{1}{\epsilon_0}E_s\Big[\sum_h |\alpha_h|\Big|\frac{N\hat{N}_h}{\hat{N}N_h} - 1\Big|\Big] \to 0,$$

where the last inequality holds from $\hat{N}_h/N_h \to_p 1$, (A.3), and the Dominated Convergence Theorem. Therefore, $T/\sqrt{\phi_{SC}} \xrightarrow{d} N(0,1)$, which implies

$$\frac{\hat{Y}_{SC} - Y}{N\sqrt{\phi_{SC}}} \xrightarrow{d} N(0,1).$$

Similarly, we can get the asymptotic properties of $\hat{Y}_{S0}$, $\hat{Y}_{CC}$, and $\hat{Y}_{C0}$.

Proof of Theorem 2. By the Law of Large Numbers, we have $\hat{\beta} \to \beta$. Then, similar to the proof of Theorem 1, we have

$$\frac{\hat{Y}_{CS} - Y}{N} = \sum_h \Big[(\beta_h - \hat{\beta})\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big) + \Big(\frac{1}{\hat{N}_h}\sum_{i\in S_h} w_i\epsilon_i - \frac{\sum_{i\in U_h}\epsilon_i}{N_h}\Big)\Big]\frac{N_h}{N}. \quad \text{(A.4)}$$

By (D2), it is easy to show that $(\hat{Y}_{CS} - Y)/N$ has the same distribution as T_2, where

$$T_2 = \sum_h \Big[(\beta_h - \beta)\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big) + \Big(\frac{1}{\hat{N}_h}\sum_{i\in S_h} w_i\epsilon_i - \frac{\sum_{i\in U_h}\epsilon_i}{N_h}\Big)\Big]\frac{N_h}{N}.$$

For T_2, $E[T_2] = E_s E_m[T_2] = 0$. Therefore, by Liapounov's Central Limit Theorem, we have

$$\frac{T_2}{\sqrt{\phi_{CS}}} \xrightarrow{d} N(0,1),$$

where

$$\phi_{CS} = E_s\Big[\frac{1}{N^2}\Big(\sum_h(\beta_h - \beta)^2\sigma_{xh}^2 + \sigma_h^2\Big)\Big(\frac{N_h^2}{\hat{N}_h^2}\sum_{i\in S_h} w_i^2 - N_h\Big)\Big].$$

For $\hat{Y}_{CC}$, by Liapounov's Central Limit Theorem, we have

$$\frac{(\hat{Y}_{CC} - Y)/N - \sum_h E_s\Big[(\alpha_h + (\beta_h - \beta)\mu_{xh})(\hat{N}_h/\hat{N} - N_h/N)\Big]}{\sqrt{\phi_{CC}^*}} \xrightarrow{d} N(0,1),$$

where

$$\phi_{CC}^* = V_s\Big[\sum_h\Big(\alpha_h + (\beta_h - \beta)\mu_{xh}\Big)\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big)\Big] + \sum_h\Big[\sigma_h^2 + (\beta_h - \beta)^2\sigma_{xh}^2\Big]E_s\Big(\frac{\sum_{i\in S_h} w_i^2}{\hat{N}^2} - \frac{2\sum_{i\in S_h} w_i}{\hat{N}N} + \frac{N_h}{N^2}\Big).$$

It can be shown that $\sqrt{n}\Big(\hat{N}/N - 1\Big) = O_p(1)$ and $\sum_h N_h^2/(\sqrt{n}N^2) = o(1)$. Based on these facts, and similar to the proof of Theorem 1, we get

$$\frac{(\hat{Y}_{CC} - Y)/N - \sum_h E_s\Big[\Big(\alpha_h + (\beta_h - \beta)\mu_{xh}\Big)\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big)\Big]}{\sqrt{\phi_{CC}}} \xrightarrow{d} N(0,1).$$

Similar to the proof of Theorem 1, by (D1) and (D3),

$$\frac{\sum_h E_s\Big[\Big(\alpha_h + (\beta_h - \beta)\mu_{xh}\Big)\Big(\frac{\hat{N}_h}{\hat{N}} - \frac{N_h}{N}\Big)\Big]}{\sqrt{\phi_{CC}}} \to 0.$$

Therefore,

$$\frac{\hat{Y}_{CC} - Y}{N\sqrt{\phi_{CC}}} \xrightarrow{d} N(0,1).$$

For $\hat{Y}_{SS}$, we have

$$\begin{aligned}\frac{\hat{Y}_{SS} - Y}{N} &= \sum_h \Big[(\beta_h - \hat{\beta}_h)\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big) + \Big(\frac{1}{\hat{N}_h}\sum_{i\in S_h} w_i\epsilon_i - \frac{\sum_{i\in U_h}\epsilon_i}{N_h}\Big)\Big]\frac{N_h}{N},\\ &= \sum_h \Big[\sum_{i\in S_h}\Big(\psi_{hi} + \frac{w_i}{\hat{N}_h} - \frac{1}{N_h}\Big)\epsilon_i - \sum_{i\in U_h/S_h}\Big(\frac{1}{N_h}\Big)\epsilon_i\Big]\frac{N_h}{N},\end{aligned}$$

where

$$\psi_{hi} = \frac{-(x_i - \bar{x}_h)w_i}{\sum_{i\in S_h}(x_i - \bar{x}_h)^2 w_i}\Big(\frac{\hat{X}_h}{\hat{N}_h} - \frac{X_h}{N_h}\Big).$$

Therefore,

$$\frac{\hat{Y}_{SS} - Y}{N\sqrt{\phi_{SS}}} \xrightarrow{d} N(0,1),$$

where

$$\phi_{SS} = \sum_h \frac{N_h^2}{N^2}\sigma_h^2 E_s E_m \sigma_h^2\Big[\sum_{i\in S_h}\Big(\psi_{hi} + \frac{w_i}{\hat{N}_h} - \frac{1}{N_h}\Big)^2 + \frac{N_h - n_h}{N_h^2}\Big].$$

Similarly, we can get the asymptotic properties of $\hat{Y}_{C0}$, $\hat{Y}_{SC}$, and $\hat{Y}_{S0}$.

References

Antal, E. and Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *J. Amer. Statist. Assoc.* **106**, 534-543.

Bickel, P. J. and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Ann. Statist.* **12**, 470-482.

Cochran, W. G. (1977). *Sampling Techniques*. Third edition. Wiley, New York.

Fuller, W. A. (2009). *Sampling Statistics*. Wiley, Hoboken, New Jersey.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.* **77**, 89-96.

Krewski, D. and Rao, J. N. K. (1981). Inference from stratified samples: properties of the linearization, jackknife and balanced repeated replication methods. *Ann. Statist.* **9**, 1010-1019.

McCarthy, P. J. and Snowden, C. B. (1985). The bootstrap and finite population sampling. *Vital and Health Statistics*, 2-95, Public Health Service Publication 85-1369. U.S. Government Printing Office, Washington, D.C.

Montanari, G. E. (1987). Post-sampling efficient Q-R prediction in large-sample surveys. *Internat. Statist. Rev.* **55**, 191-202.

Sampford, M. R. (1967). On sampling without replacement with unequal probability of selections. *Bimetrika* **54**, 499-513.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer-Verlag, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer-Verlag, New York.

Wright, R. L. (1983). Finite population sampling with multivariate auxiliary information. *J. Amer. Statist. Assoc.* **78**, 879-884.

School of Finance and Statistics, East China Normal University, Shanghai 200241, China.

Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706, U.S.A.

E-mail: shao@stat.wisc.edu

Mathematica Policy Research, Princeton, NJ 08540, U.S.A.

E-mail: kingsun2002@gmail.com

(Received February 2012; accepted December 2012)