On Small Area Estimation under Informative Sampling

Danny Pfeffermann[1] and Michael Sverchkov[2]

[1]Hebrew University and University of Southampton, [2]Bureau of Labor Statistics and User Technology Associates, Inc.
[2]Bureau of Labor Statistics, Room 1950, 2 Massachusetts Avenue, NE, Washington, DC 20212

SUMMARY

Classical small area estimation techniques assume either that all the areas are represented in the sample or that the selection of the areas to the sample is noninformative. When the areas are sampled with unequal selection probabilities that are related to the values of the response variable, the classical estimators are biased; the magnitude of the bias depends on the sampling fraction and the covariance between the sampling weights and the response variable. We illustrate this point using very simple models employing the notions of the sample distribution and sample-complement distribution. We suggest simple unbiased estimators based on these distributions.

Key words: Sample distribution, Sample-complement distribution

1. The sample and sample-complement distributions

Consider a finite population U consisting of *N* units belonging to *M* areas, with N_i units in area i, $\sum_{i=1}^{M} N_i = N$. Let *y* define the study variable with value y_{ij} for unit *j* in area i and denote by x_{ij} the values of auxiliary (covariate) variables that are possibly known for that unit. In what follows we consider the population *y*-values as random realizations of the following two level stochastic process:

First level- values (*random effects*) $\{u_1 ... u_M\}$ are generated independently from some distribution with probability density function (*pdf*) $f_p(u_i)$ for which $E_p(u_i) = 0$; $E_p(u_i^2) = \sigma_u^2$, where E_p defines the expectation operator; *Second level*- values $\{y_{i1} ... y_{iN_i}\}$ are generated from some conditional distribution with *pdf* $f_p(y_{ij} \mid x_{ij}, u_i)$, for $i = 1 ... M$. We assume a two-stage sampling scheme which in the first stage selects m areas with inclusion probabilities $\pi_i = \Pr(i \in s)$ and in the second step n_i units are sampled from area i selected in the first step with inclusion probabilities $\pi_{j|i} = \Pr(j \in s_i \mid i \in s)$. Note that the sample inclusion probabilities at both stages may depend in general on all the population or area values of *y, x* and possibly design variables *z*, used for the sample selection but not included in the working model. Denote by I_i and I_{ij} the sample indicator variables at the two stages ($\mathrm{I}_i = 1$ iff $i \in s$ and similarly for I_{ij}) and by $w_i = 1/\pi_i$ and $w_{j|i} = 1/\pi_{j|i}$ the corresponding first and second stage sampling weights.

Following Pfeffermann *et. al* (1998), we define the conditional *sample pdf* of u_i, i.e., the first level conditional *pdf* of u_i for area $i \in s$ as,

$$f_s(u_i) \overset{def}{=} f(u_i \mid \mathrm{I}_i = 1) \overset{Bayes}{=} \frac{\Pr(\mathrm{I}_i = 1 | u_i) f_p(u_i)}{\Pr(\mathrm{I}_i = 1)} \qquad (1.1)$$

Similarly, the conditional *sample-complement pdf*, i.e., the conditional *pdf* of u_i for area $i \notin s$ is defined in Sverchkov and Pfeffermann (2001) as,

$$f_c(u_i) \overset{def}{=} f(u_i \mid \mathrm{I}_i = 0) \overset{Bayes}{=} \frac{\Pr(\mathrm{I}_i = 0 | u_i) f_p(u_i)}{\Pr(\mathrm{I}_i = 0)} \qquad (1.2)$$

Notice that the *population*, *sample* and *sample-complement pdf*s of u_i are the same iff $\Pr(\mathrm{I}_i = 1 \mid u_i) = \Pr(\mathrm{I}_i = 1)\ \forall i$, in which case the sampling of areas is *noninformative*.

The second level *sample pdf* and *sample-complement pdf* of y_{ij} are defined similarly to (1.1) and (1.2) as,

$$f_s(y_{ij} \mid x_{ij}, u_i) \stackrel{def}{=} f(y_{ij} \mid x_{ij}, u_i, \mathrm{I}_{ij} = 1) = \frac{\Pr(\mathrm{I}_{ij} = 1 | y_{ij}, \mathbf{x}_{ij}, u_i) f_p(y_{ij} | \mathbf{x}_{ij}, u_i)}{\Pr(\mathrm{I}_{ij} = 1 | \mathbf{x}_{ij}, u_i)} \quad (1.3)$$

$$f_c(y_{ij} \mid x_{ij}, u_i) \stackrel{def}{=} f(y_{ij} \mid x_{ij}, u_i, \mathrm{I}_{ij} = 0) = \frac{\Pr(\mathrm{I}_{ij} = 0 | y_{ij}, \mathbf{x}_{ij}, u_i) f_p(y_{ij} | \mathbf{x}_{ij}, u_i)}{\Pr(\mathrm{I}_{ij} = 0 | \mathbf{x}_{ij}, u_i)} \quad (1.4)$$

The model defined by (1.1) and (1.3) defines the two-level *sample model* analogue of the population model defined by $f_p(u_i \mid z_i)$ and $f_p(y_{ij} \mid x_{ij}, u_i)$; see also Pfeffermann *et. al* (2001).
The following relationships are established in Pfeffermann and Sverchkov (1999) and Sverchkov and Pfeffermann (2001) for general pairs of random variables v_1, v_2 measured for elements $i \in U$ where E_p, E_s and E_c denote expectations under the *population*, *sample* and *sample-complement* distributions and (π_i, w_i) define the sample inclusion probability and the sampling weight.

$$f_s(v_{1i} | v_{2i}) = f(v_{1i} \mid v_{2i}, i \in s) = \frac{E_p(\pi_i | v_{1i}, v_{2i}) f_p(v_{1i} | v_{2i})}{E_p(\pi_i | v_{2i})} \quad (1.5)$$

$$E_p(v_{1i} | v_{2i}) = \frac{E_s(w_i v_{1i} | v_{2i})}{E_s(w_i | v_{2i})} \ ; \ E_p(\pi_i | v_{2i}) = \frac{1}{E_s(w_i | v_{2i})} \quad (1.6)$$

$$\begin{aligned} f_c(v_{1i} | v_{2i}) = f(v_{1i} | v_{2i}, i \notin s) &= \frac{E_p[(1-\pi_i) | v_{1i}, v_{2i}] f_p(v_{1i} | v_{2i})}{E_p[(1-\pi_i) | v_{2i}]} \\ &= \frac{E_s[(w_i - 1) | v_{1i}, v_{2i}] f_s(v_{1i} | v_{2i})}{E_s[(w_i - 1) | v_{2i}]} \end{aligned} \quad (1.7)$$

$$E_c(v_{1i} | v_{2i}) = \frac{E_p[(1-\pi_i) v_{1i} | v_{2i}]}{E_p[(1-\pi_i) | v_{2i}]} = \frac{E_s[(w_i - 1) v_{1i} | v_{2i}]}{E_s[(w_i - 1) | v_{2i}]} \quad (1.8)$$

Defining $v_{1i} = u_i, v_{2i}$=*constant* yields the relationships holding for the random area effects u_i. Defining $v_{1ij} = y_{ij}$; $v_{2ij} = (x_{ij}, u_i)$ and substituting $\pi_{j|i}$ and $w_{j|i}$ for π_i and w_i respectively yields the relationships holding for the observations y_{ij}.

2. Optimal Small Area Predictors

The target estimated population parameters are the small area means $\bar{Y}_i = \sum_{j=1}^{N_i} y_{ij} / N_i$ for $i = 1...M$. Let $D_s = \{(y_{ij}, \pi_{j|i}, \pi_i), (i,j) \in s; (\mathrm{I}_{kl}, \mathrm{I}_k, x_{kl}), (k,l) \in U\}$ define the known data. The MSE of a predictor $\hat{\bar{Y}}_i$ given D_s with respect to the *population pdf* is,

$$\begin{aligned} MSE(\hat{\bar{Y}}_i \mid D_s) &= E_p[(\hat{\bar{Y}}_i - \bar{Y}_i)^2 \mid D_s] = E_p\{[\hat{\bar{Y}}_i - E_p(\bar{Y}_i \mid D_s)]^2 \mid D_s\} + V_p(\bar{Y}_i \mid D_s) \\ &= [\hat{\bar{Y}}_i - E_p(\bar{Y}_i \mid D_s)]^2 + V_p(\bar{Y}_i \mid D_s) \end{aligned} \quad (2.1)$$

The variance $V_p(\bar{Y}_i \mid D_s)$ does not depend on the form of the predictor and hence the MSE is minimized when $\hat{\bar{Y}}_i = E_p(\bar{Y}_i \mid D_s)$. In what follows we distinguish between *sampled areas* ($\mathrm{I}_i = 1$) and *nonsampled areas* ($\mathrm{I}_i = 0$). Denote by s_i the sample of units in sampled area i. Then, for the sampled areas,

$$E_p(\overline{Y}_i \mid D_s, \mathrm{I}_i = 1) = \frac{1}{N_i}\{\sum\nolimits_{j\in s_i} E_p(y_{ij} \mid D_s) + \sum\nolimits_{l\notin s_i} E_p(y_{il} \mid D_s, \mathrm{I}_{il} = 0)$$
$$= \frac{1}{N_i}\{\sum\nolimits_{j\in s_i} y_{ij} + \sum\nolimits_{l\notin s_i} E_c(y_{il} \mid D_s) \tag{2.2}$$

For areas i not in the sample,

$$E_p(\overline{Y}_i \mid D_s, \mathrm{I}_i = 0) = \frac{1}{N_i}\sum\nolimits_{k=1}^{N_i} E_p(y_{ik} \mid D_s, \mathrm{I}_{ik} = 0)$$
$$= \frac{1}{N_i}\sum\nolimits_{k=1}^{N_i} E_c(y_{ik} \mid D_s) \tag{2.3}$$

The predictors in (2.2) and (2.3) can be written in a single equation as,

$$E_p(\overline{Y}_i \mid D_s) = \frac{1}{N_i}\{\sum\nolimits_{k=1}^{N_i} y_{ik}\mathrm{I}_{ik} + \sum\nolimits_{k=1}^{N_i} E_c[y_{ik}(1-\mathrm{I}_{ik}) \mid D_s]\}\mathrm{I}_i$$
$$+ \frac{1}{N_i}\{\sum\nolimits_{k=1}^{N_i} E_c[y_{ik} \mid D_s]\}(1-\mathrm{I}_i) \tag{2.4}$$

3. Bias of Small Area Predictors when ignoring the Sampling Scheme

Consider for convenience the case of a sampled area. Ignoring the sampling scheme implies an implicit assumption that the *sample-complement* model and the *sample* model are the same such that $\hat{\overline{Y}}_{i,IGN} = \sum_{j\in s_i} y_{ij} + \sum_{l\notin s_i} E_s(y_{il} \mid D_s)$. Hence,

$$E_p[(\hat{\overline{Y}}_{i,IGN} - \overline{Y}_i) \mid D_s, \mathrm{I}_i = 1] = \frac{1}{N_i}\sum\nolimits_{l\notin s_i}[E_s(y_{il} \mid D_s) - E_c(y_{il} \mid D_s)]$$
$$= -\frac{1}{N_i}\sum\nolimits_{l\notin s_i}\frac{Cov_s(y_{il}, w_{l|i} \mid D_s)}{E_s[(w_{l|i}-1) \mid D_s]} \tag{3.1}$$

with the second equality following from (1.8). Thus, unless the response values y_{il} and the 'within' sampling weights $w_{l|i}$ are uncorrelated, ignoring the sampling scheme results in biased predictors (see also the empirical results). A similar expression for the bias can be obtained for the nonsampled areas.

A simple Example. Let the *population model* be the "unit level random effects model"

$$y_{ij} = \mu + u_i + e_{ij} \ ; \ u_i \sim N(0,\sigma_u^2) \ , \ e_{ij} \sim N(0,\sigma_e^2) \tag{3.2}$$

with all the random effects and residual terms being mutually independent.

Let $\pi_i = c \times N_i$ where *c* is some constant and $\pi_{j|i} = n_0 / N_i$ (fixed sample size n_0 within the selected areas), such that $\pi_{ij} = \Pr[(i,j)\in s] = \pi_i\pi_{j|i} = const$. Note that the sample selection within the selected areas is *noninformative* in this case but if the area sizes N_i are correlated with the random effects u_i (say, the areas are *schools*, the study variable measures children's attainment, the large schools are in the poor areas), the selection of the areas is *informative*.

Suppose that the areas sizes can be modeled as $\log(N_i) \sim N(Au_i, \sigma_M^2)$, implying that $E_p(\pi_i \mid u_i) \prec \exp(Au_i + \frac{\sigma_M^2}{2})$ by familiar properties of the lognormal distribution. It follows that (see Pfeffermann *et al*. 1998, example 4.3),

$$f_s(u_i) = \frac{E_p(\pi_i | u_i) f_p(u_i)}{E_p(\pi_i)} = N(A\sigma_u^2, \sigma_u^2) \tag{3.3}$$

so that $E_s(u_i) = \gamma\sigma_u^2 \neq E_p(u_i) = 0$. The fact that the random effects in the sample have in this case a positive expectation is easily explained by the fact that the sampling scheme considered tends to select the areas with large positive random effects. Note, however, that by defining $\mu^* = \mu + A\sigma_u^2$ and $u_i^* = u_i - A\sigma_u^2$, the model holding for the sample data in sampled areas is $y_{ij} = \mu^* + u_i^* + e_{ij}$, $u_i^* \sim N(0,\sigma_u^2)$, $e_{ij} \sim N(0,\sigma_e^2)$, which is the same as the population model. Thus, the optimal predictors under the *population model* for the area means $\theta_i = \mu + u_i$ of the sampled areas ($\mathrm{I}_i = 1$) are still optimal under the *sample model*

Next consider *nonsampled areas*. By (1.7),

$$f_c(u_i) = \frac{E_p[(1-\pi_i) \mid u_i] f_p(u_i)}{E_p(1-\pi_i)} = \frac{f_p(u_i)}{E_p(1-\pi_i)} - \frac{E_p(\pi_i \mid u_i) f_p(u_i)}{E_p(1-\pi_i)} \qquad (3.4)$$

Let $E_p(m) = E_p[\sum_{l=1}^{M} \mathrm{I}_l] = E_p[E_p(\sum_{l=1}^{M} \mathrm{I}_l \mid \{N_i\})] = E_p[\sum_{l=1}^{M} \pi_i] = ME_p(\pi_i)$ define the expected number of sampled areas, such that $E_p(\pi_i) = E_p(m)/M$. If the number of sampled areas is fixed, $E_p(m) = m$. By (3.4) and (1.5),

$f_c(u_i) = [Mf_p(u_i) - E_p(m) f_s(u_i)]/[M - E_p(m)]$ and hence,

$$E_c(u_i) = -\frac{E_p(m)E_s(u_i)}{M - E_p(m)} = -\frac{E_p(m)A\sigma_u^2}{M - E_p(m)} \qquad (3.5)$$

Here again, the negative expectation of the random effects pertaining to *nonsampled areas* is easily explained by the tendency of the sampling scheme to select the areas with the large positive random effects. Thus, ignoring the sampling scheme underlying the selection of the areas and predicting the sample means in nonsampled areas by, say, the average of the predictors in the sampled areas yields in general biased predictors with a positive bias defined by the absolute value of the right hand side of (3.5).

4. On Small Area Estimation based on Sample Distribution

In order to illustrate the proposed approach, we suppose that the area level random effects model defined by (4.1) holds for the sampled areas, i.e., for $j \in s_i$

$$y_{ij} = \mu + u_i + e_{ij} \;;\; u_i \mid \mathrm{I} = 1_i \sim N(0,\sigma_u^2)\;,\; e_{ij} \mid \mathrm{I}_{ij} = 1 \sim N(0,\sigma_e^2) \qquad (4.1)$$

We mention in this respect that the sample model can be identified using conventional techniques, see, e.g., Rao (2003).

Suppose that in the first stage *m* areas are selected with inclusion probabilities π_i (*m* is fixed) and in the second stage n_i units are sampled from area i selected in the first stage with inclusion probabilities $\pi_{j|i}$ where again, we assume for convenience that the sample sizes n_i are fixed. Assume that,

$$E_s(w_{j|i} \mid y_{ij}, u_i) = E_s(w_{j|i} \mid y_{ij}) = c_i \exp(b y_{ij}) \qquad (4.2)$$

where $\theta_i = \mu + u_i$ and c_i>0 and b are fixed parameters. Notice that since n_i is fixed, $E_s(w_{j|i} \mid u_i) = N_i / n_i$.

<u>*Comment*</u>: As with the sample model (4.1), the expectation in (4.2) refers to the *sample distribution* within the areas. The relationship between the sampling weights and the observed data holding in the sample can be identified and estimated therefore from the sample data. See Pfeffermann and Sverchkov (1999, 2003) for discussion and examples. On the other hand, the relationship between the sampling weights w_i and the small area means $\theta_i = \mu + u_i$ is more difficult to detect since the area means are not observable and in what follows we do not model this relationship. See Pfeffermann *et al*. (2001) for an example of modeling the selection probabilities at both stages.

As established in Section 2, the optimal predictor for areas *in the sample* is,

$E_p(\overline{Y}_i \mid D_s, \mathrm{I}_i = 1) = [\sum_{j\in s_i} y_{ij} + \sum_{l\notin s_i} E_c(y_{il} \mid D_s, \mathrm{I}_i = 1)]/N_i$. In order to compute the expectations $E_c(y_{il} \mid D_s, \mathrm{I}_i = 1)$ we follow the following steps. First, by (1.7), (4.1) and (4.2),

$$f_c(y_{il} | \theta_i, \mathrm{I}_i = 1) = \frac{[E_s(w_{l|i} | y_{il}, \theta_i) - 1] f_s(y_{il} | \theta_i)}{E_s(w_{l|i} | \theta_i) - 1}$$

$$= [c_i \exp(by_{il}) - 1] \frac{1}{\sigma_e} \phi[\frac{y_{il} - \theta_i}{\sigma_e}] / [\frac{N_i}{n_i} - 1] \quad (4.3)$$

$$= \frac{n_i}{N_i - n_i} \{ c_i \exp(\theta_i b + \frac{\sigma_e^2 b^2}{2}) \frac{1}{\sigma_e} \phi[\frac{y_{il} - (\theta_i + b\sigma_e^2)}{\sigma_e}] - \frac{1}{\sigma_e} \phi[\frac{y_{il} - \theta_i}{\sigma_e}] \}$$

where ϕ is the standard normal *pdf*. Notice that if $b = 0$ (noninformative selection within the sampled areas with equal inclusion probabilities), $c_i = N_i / n_i$ and the *pdf* in (4.3) reduces to the conditional normal density defined by (4.1). Second, by (4.3),

$$E_c(y_{il} \mid \theta_i, \mathrm{I}_i = 1) = \frac{n_i}{N_i - n_i} \{ c_i \exp(\theta_i b + \frac{\sigma_e^2 b^2}{2}) [\theta_i + b\sigma_e^2] - \theta_i \} \quad (4.4)$$

Finally,

$$E_c(y_{il} \mid D_s, \mathrm{I}_i = 1) = E_s[E_c(y_{il} \mid D_s, \mathrm{I}_i = 1, \theta_i)] = E_s[E_c(y_{il} \mid \mathrm{I}_i = 1, \theta_i)] \quad (4.5)$$

where the exterior expectation is with respect to the distribution of $\theta_i \mid D_s, \mathrm{I}_i = 1$. Under the model (4.1), the latter distribution is normal with mean $\hat{\theta}_i = \gamma_i \overline{y}_i + (1-\gamma_i)\overline{y}$ and variance $v_i = \gamma_i \sigma_i^2 + (1-\gamma_i)^2 (\sum_{i=1}^m \gamma_i / \sigma_i^2)^{-1}$ where $\overline{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i$ is the sample mean in sampled area i, $\overline{y} = \sum_{i=1}^m n_i \overline{y}_i / \sum_{i=1}^m n_i$, $\sigma_i^2 = \sigma_e^2 / n_i = Var(\overline{y}_i \mid u_i)$ and $\gamma_i = \sigma_u^2 / [\sigma_u^2 + \sigma_i^2]$.

Thus, for the *sampled areas* $E_c(y_{il} \mid D_s, \mathrm{I}_i = 1)$ is obtained by computing the expectation of the right hand side of (4.4) with respect to the normal distribution of $\theta_i \mid D_s, \mathrm{I}_i = 1$. We find that,

$$E_c(y_{il} \mid D_s, \mathrm{I}_i = 1) = \frac{n_i}{N_i - n_i} \{ c_i [\hat{\theta}_i + b(v_i + \sigma_e^2)] \exp[\hat{\theta}_i b + \frac{b^2}{2}(\sigma_e^2 + v_i)] - \hat{\theta}_i \} \quad (4.6)$$

Notice that if b=0 (noninformative sampling within the areas with equal inclusion probabilities) $c_i = N_i / n_i$ and $E_c(y_{il} \mid D_s, \mathrm{I}_i = 1) = \hat{\theta}_i$.

<u>*Comment*</u>: The optimal predictor obtained for the case of noninformative sampling, $E_p(\overline{Y}_i \mid D_s, \mathrm{I}_i = 1) = [\sum_{j\in s_i} y_{ij} + (N_i - n_i)\hat{\theta}_i]/N_i$ (Eq. 2.2) is different from the common predictor, $\hat{\theta}_i$. This is so because the target parameter is defined to be the finite area mean $\overline{Y}_i$ rather than θ_i. See also Prasad and Rao (1990).

For the *nonsampled* areas the optimal predictor is defined in (2.3) to be, $E_p(\overline{Y}_i \mid D_s, \mathrm{I}_i = 0) = \sum_{k=1}^{N} E_c(y_{ik} \mid D_s, \mathrm{I}_i = 0)/N_i$. In order to compute the expectations $E_c(y_{ik} \mid D_s, \mathrm{I}_i = 0)$ we note first that

$$f_p(y_{ij} \mid \theta_i, \mathrm{I}_i = 1) = f_p(y_{ij} \mid \theta_i, \mathrm{I}_i = 0) = f_p(y_{kl} \mid \theta_k) \quad (4.7)$$

signifying that conditionally on the area means θ_i, the *population pdf* is the same for all the areas irrespective of whether the areas are sampled or not. The *pdf* $f_p(y_{il} | \theta_i)$ is obtained from (1.5), (1.6) and (4.2) similarly to the derivation of $f_c(y_{il} | \theta_i, \mathrm{I}_i = 1)$ in (4.3) as,

$$f_p(y_{il}|\theta_i) = \frac{E_s(w_{l|i}|y_{il},\theta_i) f_s(y_{il}|\theta_i)}{E_s(w_{l|i}|\theta_i)} \qquad (4.8)$$

$$= \frac{c_i n_i}{N_i}\exp(\theta_i b + \frac{\sigma_e^2 b^2}{2})\frac{1}{\sigma_e}\phi[\frac{y_{il}-(\theta_i + b\sigma_e^2)}{\sigma_e}]$$

Notice that the population *pdf* is different from the sample *pdf* defined by (4.1) unless the sampling scheme within the areas is noninformative (b=0).

By (4.8),

$$E_p(y_{il} \mid \theta_i) = \frac{c_i n_i}{N_i}\exp(\theta_i b + \frac{\sigma_e^2 b^2}{2})(\theta_i + b\sigma_e^2) \qquad (4.9)$$

Now,

$$E_c(y_{ik} \mid D_s, \mathrm{I}_i = 0) \overset{Def}{=} E_p(y_{ik} \mid D_s, \mathrm{I}_i = 0)$$

$$= E_p[E_p(y_{ik} \mid \theta_i, D_s, \mathrm{I}_i = 0) \mid D_s, \mathrm{I}_i = 0] \qquad (4.10)$$

$$= E_p[E_p(y_{ik} \mid \theta_i) \mid D_s, \mathrm{I}_i = 0] \overset{Def}{=} E_c[E_p(y_{ik} \mid \theta_i) \mid D_s]$$

where the exterior expectations in the last row are with respect to the conditional distribution $f_p(\theta_i \mid D_s, \mathrm{I}_i = 0) = f_c(\theta_i \mid D_s)$.

Finally, by (1.8) and (4.10),

$$E_c(y_{ik} \mid D_s, \mathrm{I}_i = 0) = E_c[E_p(y_{ik} \mid \theta_i) \mid D_s] = E_s[\frac{(w_i - 1)E_p(y_{ik} \mid \theta_i)}{E_s(w_i \mid D_s) - 1} \mid D_s] \qquad (4.11)$$

Denoting $\theta_{i,p} = E_p(y_{ik} \mid \theta_i)$, an estimator of the expectation $E_c(y_{ik} \mid D_s, \mathrm{I}_i = 0)$ is obtained from (4.11) as,

$$\hat{E}_c(y_{ik} \mid D_s, \mathrm{I}_i = 0) = \frac{1}{m}\sum_{i\in s}\frac{(w_i - 1)\hat{\theta}_{i,p}}{\hat{E}_s(w_i \mid D_s) - 1} = \sum_{i\in s}\frac{(w_i - 1)\hat{\theta}_{i,p}}{\sum_{i\in s}(w_i - 1)} \qquad (4.12)$$

where $\hat{E}_s(w_i \mid D_s) = \frac{1}{m}\sum_{i\in s} w_i$ and $\hat{\theta}_{i,p}$ is obtained by substituting θ_i in (4.9) by $\hat{\theta}_i = \gamma_i \bar{y}_i + (1-\gamma_i)\bar{y} = E_s(\theta_i \mid D_s, \mathrm{I}_i = 1)$ or by use direct Hajek estimator of $\theta_{i,p} = E_p(y_{ik} \mid \theta_i)$. Notice that the right hand side of (4.12) defines the predictor of the mean θ_i's in the *nonsampled* areas.

References

Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8, 1087-1114.

Pfeffermann, D., Moura, F., Silva, P.N. (2002). Fitting multilevel models under informative probability sampling. *Multilevel Modeling Newsletter*, 14, 8-17.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data, *Sankhya*, B, 61, Pt.1, 166 – 186

Pfeffermann, D. and Sverchkov, M. (2003). Fitting generalised linear models under informative sampling. Chapter 12 in R. L. Chambers and C. Skinner (eds.), *Analysis of Survey Data,* Chichester: Wiley , pp. 175 - 195

Rao, J. N. K. (2003). *Small Area Estimation*, Wiley.

Prasad, N. G. N. and Rao, J. N. K. (1999). On robust small area estimation using a simple random effect model. *Survey Methodology*, 25, 67-72

Sverchkov, M. and Pfeffermann, D. (2000). Prediction of finite population totals under informative sampling utilizing the sample distribution, *American Statistical Association Proceedings of the Section on Survey Research Methods*, 0, 41 - 46