Regression estimators in forest inventories with two-phase sampling and partially exhaustive information with applications to small-area estimation

Daniel Mandallaz [1]

Chair of Land Use Engineering

ETH Zurich

CH 8092 Zurich, Switzerland

[1]Tel. ++41(0)44 6323186 e-mail daniel.mandallaz@env.ethz.ch

Abstract

We consider two-phase sampling schemes where one component of the auxiliary information is known in every point ("wall-to wall") and a second component is available only in the large sample of the first phase, whereas the second phase yields a sub-sample with the terrestrial inventory data based on general tree inclusion probabilities. We propose a generalized version of the classical two-phase regression estimator, for global and local estimation and derive its asymptotic design-based variance. Cluster and two-stage sampling procedures are also considered.

1 Introduction

The motivation for this work is due to the increasing need of using national or regional inventories for local estimation in order to meet tighter budgetary constraints, which is only feasible under extensive use of auxiliary information, provided e.g. by remote sensing. It is hoped that the proposed estimators will contribute to achieve this objective, particularly because they are easy to implement in software packages like SAS or R.

There is an extensive literature on the problem of small area estimation (or small domain estimation in general sampling). In this paper we shall investigate the properties of some estimators in the **model-assisted framework**, in which prediction models are used to improve the efficiency but are not assumed to be correct as in the **model-dependent approach**. The validity of the statistical procedures is ensured by the randomization principle: i.e. we are in the **design-based** inference framework, which has a definite advantage in official statistics. The reader is referred to (Koehl et al. (2006), section 3.8) for a good review of small-area estimation in forest inventory that presents alternative techniques, in particular Bayesian. Let us now define the sampling scheme.

The **first phase** draws a large sample s_1 of n_1 points $x_i \in s_1$ $(i = 1, 2 \ldots n_1)$ that are independently and uniformly distributed within the forest area F. At each of those points auxiliary information is collected, very often coding information of qualitative nature (e.g. following the interpretation of aerial photographs) or quantitative (e.g. timber volume estimates based on LIDAR measurements). We shall assume that the auxiliary information at point x is described by the row vector $\boldsymbol{Z}^t(x) = (\boldsymbol{Z}^{(1)t}(x), \boldsymbol{Z}^{(2)t}(x)) \in \Re^{p+q}$ (the upper index t denotes the transposition operator). The first component $\boldsymbol{Z}^{(1)}(x) \in \Re^p$ of this vector is known at all points $x \in F$, it is the **exhaustive** part of the auxiliary information, e.g. it could be given by thematic maps. The second component $\boldsymbol{Z}^{(2)}(x) \in \Re^q$ is known only at points $x \in s_1$.

The **second phase** draws a small sample $s_2 \subset s_1$ of n_2 points from s_1 according to

equal probability sampling without replacement. In the forested area F we consider a well defined population $\mathcal{P}$ of N trees with response variable $Y_i,\ i = 1, 2 \ldots$, e.g. the timber volume. The objective is to estimate the spatial mean $\bar{Y} = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i$, where $\lambda(F)$ denotes the surface area of F (usually in ha). For each point $x \in s_2$ trees are drawn from the population $\mathcal{P}$ with probabilities π_i, for instance with concentric circles or angle count techniques. The set of trees selected at point x is denoted by $s_2(x)$. From each of the selected trees $i \in s_2(x)$ one determines Y_i. The indicator variable I_i is defined as

$$[1] \qquad I_i(x) = \begin{cases} 1 \text{ if } i \in s_2(x) \\ \\ 0 \text{ if } i \notin s_2(x) \end{cases}$$

At each point $x \in s_2$ the terrestrial inventory provides the local density $Y(x)$

$$[2] \qquad Y(x) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{I_i(x) Y_i}{\pi_i} = \frac{1}{\lambda(F)} \sum_{i \in s_2(x)} \frac{Y_i}{\pi_i}$$

The term $\frac{1}{\lambda(F)\pi_i}$ is the tree extrapolation factor f_i with dimension ha^{-1}. Because of possible boundary adjustments $\lambda(F)\pi_i = \lambda(F \cap K_i)$, where K_i is the inclusion circle of the i-th tree. In the infinite population or Monte Carlo approach one samples the function $Y(x)$ (Mandallaz (2008)) for which the following important relation holds:

$$[3] \qquad \mathbb{E}_x(Y(x)) = \frac{1}{\lambda(F)} \int_F Y(x) dx = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i = \bar{Y}$$

Where $\mathbb{E}_x$ denotes the expectation with respect to a random point x uniformly distributed in F. This establishes the link between the infinite population (continuum) $\{x \in F \mid Y(x)\}$ and the finite population of trees $\{i = 1, 2 \ldots N \mid Y_i\}$.

We shall work with the following linear models (see Mandallaz (2008), Chapter 6, and Mandallaz (2012) for more details)

1. The large model M

$$Y(x) = \boldsymbol{Z}(x)^t\boldsymbol{\beta} + R(x) = \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\beta}^{(1)} + \boldsymbol{Z}^{(2)t}(x)\boldsymbol{\beta}^{(2)} + R(x)$$

with $\boldsymbol{\beta}^t = (\boldsymbol{\beta}^{(1)t}, \boldsymbol{\beta}^{(2)t})$. The intercept term is contained in $\boldsymbol{Z}^{(1)}(x)$ or it is a linear combination of its components.

The theoretical regression parameter $\boldsymbol{\beta}$ minimizes

$$\int_F (Y(x) - \boldsymbol{Z}^t(x)\boldsymbol{\beta})^2 dx$$

It satisfies the normal equation

$$\Big(\int_F \boldsymbol{Z}(x)\boldsymbol{Z}^t(x)dx\Big)\boldsymbol{\beta} = \int_F Y(x)\boldsymbol{Z}(x)dx$$

and the orthogonality relationship

$$\int_F R(x)\boldsymbol{Z}(x)dx = \mathbf{0}$$

in particular the zero mean residual property

$$\frac{1}{\lambda(F)}\int_F R(x)dx = 0$$

2. The reduced model M_1

$$Y(x) = \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\alpha} + R_1(x)$$

The theoretical regression parameter $\boldsymbol{\alpha}$ minimizes

$$\int_F (Y(x) - \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\alpha})^2 dx$$

It satisfies the normal equation

$$\Big(\int_F \boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)dx\Big)\boldsymbol{\alpha} = \int_F Y(x)\boldsymbol{Z}^{(1)}(x)dx$$

the orthogonality relationship

$$\int_F R_1(x)\boldsymbol{Z}^{(1)}(x)dx = \mathbf{0}$$

in particular the zero mean residual property

$$\frac{1}{\lambda(F)}\int_F R_1(x)dx = 0$$

.

Let us emphasize the fact that we do not assume that the regression models are correct: the inference is based on the sampling design, that is we are doing model-assisted (and not model-dependent or model-based) inference in the sense of Särndal (see Särndal et al. (2003)).

2 The generalized regression estimator

We need the following design-based least squares estimators of the regression coefficients of the reduced model, which are essentially solutions of sample copies of the normal equations

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_k &= \Big(\frac{1}{n_k}\sum_{x\in s_k}\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)}(x)^t\Big)^{-1}\frac{1}{n_k}\sum_{x\in s_k}Y(x)\boldsymbol{Z}^{(1)}(x) \\ &:= (\boldsymbol{A}_k^{(1)})^{-1}\frac{1}{n_k}\sum_{x\in s_k}Y(x)\boldsymbol{Z}^{(1)}(x) \quad k=1,2 \end{aligned} \tag{4}$$

For the large model we set

$$[5] \qquad \hat{\boldsymbol{\beta}}_k = \Big(\frac{1}{n_2}\sum_{x\in s_k}\boldsymbol{Z}(x)\boldsymbol{Z}(x)^t\Big)^{-1}\frac{1}{n_2}\sum_{x\in s_k}Y(x)\boldsymbol{Z}(x) := \boldsymbol{A}_k^{-1}\frac{1}{n_2}\sum_{x\in s_2}Y(x)\boldsymbol{Z}(x)$$

Note that only $\hat{\boldsymbol{\alpha}}_2$ and $\hat{\boldsymbol{\beta}}_2$ are observable and that in general the vector consisting of the first p components of $\hat{\boldsymbol{\beta}}_2$ is not equal to $\hat{\boldsymbol{\alpha}}_2$ (they are if the corresponding explanatory variables are orthogonal in the classical least squares sense).

The large model yields the predictions $\hat{Y}(x) = \boldsymbol{Z}^t(x)\hat{\boldsymbol{\beta}}_2$ and the reduced model the predictions $\hat{Y}_1(x) = \boldsymbol{Z}^{(1)t}(x)\hat{\boldsymbol{\alpha}}_2$.

The **generalized regression estimate** is defined as

$$[6] \qquad \hat{Y}_{greg} = \frac{1}{\lambda(F)}\int_F \hat{Y}_1(x)dx + \frac{1}{n_1}\sum_{x\in s_1}(\hat{Y}(x) - \hat{Y}_1(x)) + \frac{1}{n_2}\sum_{x\in s_2}(Y(x) - \hat{Y}(x))$$

This estimator is the Monte Carlo version of Särndal's regression estimator for two-phase sampling in finite population (see Särndal et al. (2003), equation 9.7.20). It is clear by the law of large numbers that $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\alpha}}_2$ are asymptotically design-unbiased estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. This implies at once that

$$\mathbb{E}_{1,2}\hat{Y}_{reg} = \mathbb{E}_1\mathbb{E}_{2|1}\hat{Y}_{reg} \approx \bar{Y}$$

$\mathbb{E}_{2|1}$ denotes the conditional expectation of the second phase given the first phase (i.e. simple random sampling without replacement in the population s_1) and $\mathbb{E}_1$ denotes the expectation with respect to uniformly distribution points of the first phase (i.e. to $\mathbb{E}_x$).

The generalized regression estimate is therefore asymptotically design-unbiased.

To understand the potential usefulness of $\hat{Y}_{greg}$ we shall assume for the time being that the model is **external**, i.e. not fitted by the inventory data, and that the regression

coefficients have given fixed values. Using the well known variance decomposition

$$[7] \qquad \mathbb{V}(\hat{Y}_{greg}) = \mathbb{V}_1\mathbb{E}_{2|1}(\hat{Y}_{greg}) + \mathbb{E}_1\mathbb{V}_{2|1}(\hat{Y}_{greg})$$

we get the design-based variance as

$$[8] \qquad \mathbb{V}(\hat{Y}_{greg}) = \frac{1}{n_1}\mathbb{V}_x(R_1(x)) + (1 - \frac{n_2}{n_1})\frac{1}{n_2}\mathbb{V}_x(R(x))$$

The variances $\mathbb{V}_x(\cdot)$ are calculated under the uniform distribution in F of the random point x. An unbiased estimate of the variance is given by

$$[9] \qquad \hat{\mathbb{V}}(\hat{Y}_{greg}) = \frac{1}{n_1}\frac{1}{n_2 - 1}\sum_{x\in s_2}(R_1(x) - \hat{\bar{R}}_1)^2 + (1 - \frac{n_2}{n_1})\frac{1}{n_2(n_2-1)}\sum_{x\in s_2}(R(x) - \hat{\bar{R}})^2$$

where $\hat{\bar{R}}_1 = \frac{1}{n_2}\sum_{x\in s_2} R_1(x)$ and likewise for $\hat{\bar{R}}$.

This should be compared with the standard result for the variance of the regression estimator $\hat{Y}_{reg}$ under the large model

$$[10] \qquad \hat{Y}_{reg} = \frac{1}{n_1}\sum_{x\in s_1}\hat{Y}(x) + \frac{1}{n_2}\sum_{x\in s_2}(Y(x) - \hat{Y}(x))$$

whose theoretical variance is given by

$$[11] \qquad \mathbb{V}(\hat{Y}_{reg}) = \frac{1}{n_1}\mathbb{V}_x(Y(x)) + (1 - \frac{n_2}{n_1})\frac{1}{n_2}\mathbb{V}_x(R(x))$$

Thus, by using also the exhaustive information the variance of the observations in [11] is replaced by the variance of the residuals under the reduced model, a very nice and intuitive result indeed. To have better insight consider the design-based coefficients of

determination

$$[12] \qquad R_1^2 = \frac{\mathbb{V}_x(\hat{Y}_1(x))}{\mathbb{V}_x(Y(x))} \quad \text{and} \quad R^2 = \frac{\mathbb{V}_x(\hat{Y}(x))}{\mathbb{V}_x(Y(x))}$$

According to (Mandallaz (2008), equation 5.4) the reduction in variance is easily found to be

$$[13] \qquad \mathbb{V}(\hat{Y}_{reg}) - \mathbb{V}(\hat{Y}_{greg}) = \mathbb{V}(Y(x))\Big(\frac{R^2 - R_1^2}{n_1} + \frac{1 - R^2}{n_2}\Big) > 0$$

We now give an alternative definition of $\hat{Y}_{greg}$ based on the estimated regression coefficients which is essential to derive the design-based variance with **internal models**, i.e. fitted with the inventory data at hand, and for future generalization to the small-area estimation problem. To this end we need the following mean values

$$[14] \quad \bar{\boldsymbol{Z}}^{(1)} = \frac{1}{\lambda(F)} \int_F \boldsymbol{Z}^{(1)}(x)dx, \quad \hat{\bar{\boldsymbol{Z}}}_1^{(1)} = \frac{1}{n_1} \sum_{x \in s_1} \boldsymbol{Z}^{(1)}(x), \quad \hat{\bar{\boldsymbol{Z}}}_k = \frac{1}{n_k} \sum_{x \in s_k} \boldsymbol{Z}(x), \ k = 1, 2$$

The regression estimate can be rewritten as

$$\begin{aligned} \hat{Y}_{greg} &= (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t \hat{\boldsymbol{\alpha}}_2 + (\hat{\bar{\boldsymbol{Z}}}_1 - \hat{\bar{\boldsymbol{Z}}}_2)^t \hat{\boldsymbol{\beta}}_2 + \frac{1}{n_2} \sum_{x \in s_2} Y(x) \\ [15] \qquad &= (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t \hat{\boldsymbol{\alpha}}_2 + \hat{\bar{\boldsymbol{Z}}}_1^t \hat{\boldsymbol{\beta}}_2 \end{aligned}$$

The last equations follows from the fact that the sum of the residuals is zero by construction. Note that it suffices to know the integral of $\boldsymbol{Z}^{(1)}(x)$ and not necessarily the values at all points $x \in F$.

3 Variance estimate

To obtain a first estimate of the variance we can treat the internal model as an external one and replace in [8] the theoretical residuals by their empirical versions $\hat{R}_1(x) = Y(x) - \hat{Y}_1(x) = Y(x) - \boldsymbol{Z}^{(1)t}(x)\hat{\boldsymbol{\alpha}}_2$ and $\hat{R}(x) = Y(x) - \hat{Y}(x) = Y(x) - \boldsymbol{Z}^t(x)\hat{\boldsymbol{\beta}}_2$, which have zero means, to obtain

$$[16] \qquad \hat{\mathbb{V}}(\hat{Y}_{greg}) = \frac{1}{n_1}\frac{1}{n_2}\sum_{x\in s_2}\hat{R}_1^2(x) + \frac{1}{n_2}(1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{x\in s_2}\hat{R}^2(x)$$

To derive better variance estimates we shall use the g-weights technique (for details see Mandallaz (2008), section 6.2, for the Monte Carlo approach and Särndal et al. (2003), sections 6.5 and 6.6 for finite populations). The g-weights are defined by

$$\begin{aligned} g_2(x) &= 1 + (\hat{\bar{\boldsymbol{Z}}}_1 - \hat{\bar{\boldsymbol{Z}}}_2)^t \boldsymbol{A}_2^{-1}\boldsymbol{Z}(x) = \hat{\bar{\boldsymbol{Z}}}_1^t \boldsymbol{A}_2^{\ -1}\boldsymbol{Z}(x) \\ [17] \qquad g_1^{(1)}(x) &= 1 + (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t (\boldsymbol{A}_1^{(1)})^{-1}\boldsymbol{Z}^{(1)}(x) = \bar{\boldsymbol{Z}}^{(1)t}(\boldsymbol{A}_1^{(1)})^{-1}\boldsymbol{Z}^{(1)}(x) \end{aligned}$$

That the two versions of the g-weights are equivalent is a consequence of the zero residual sum for any local density (see Mandallaz (2008), section 6.2). The g-weights are therefore of order $1 + O_p(n_k^{-\frac{1}{2}})$ in probability. Note that they depend not only on the point x but also on the entire sample s_2, though weakly. Straightforward algebra leads to the following important calibration properties

$$\begin{aligned} \frac{1}{n_2}\sum_{x\in s_2} g_2(x)\boldsymbol{Z}(x) &= \hat{\bar{\boldsymbol{Z}}}_1 \\ [18] \qquad \frac{1}{n_1}\sum_{x\in s_1} g_1^{(1)}(x)\boldsymbol{Z}^{(1)}(x) &= \bar{\boldsymbol{Z}}^{(1)} \end{aligned}$$

and also to

$$\begin{aligned} \frac{1}{n_1}\sum_{x\in s_1} g_1^{(1)}(x)Y(x) &= \frac{1}{n_1}\sum_{x\in s_1} Y(x) + (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})\hat{\boldsymbol{\beta}}_1^{(1)} \\ \frac{1}{n_2}\sum_{x\in s_2} g_2(x)Y(x) &= \frac{1}{n_2}\sum_{x\in s_2} Y(x) + (\hat{\bar{\boldsymbol{Z}}}_1 - \hat{\bar{\boldsymbol{Z}}}_2)^t\hat{\boldsymbol{\beta}}_2 \\ [19] \qquad &= \hat{\bar{\boldsymbol{Z}}}_1^t\hat{\boldsymbol{\beta}}_2 = \hat{Y}_{reg} \end{aligned}$$

$\hat{Y}_{reg}$ is the standard regression estimator based on the large model only (Mandallaz (2008), section 5.1). The last equation in [19] follows again from the fact that the residuals sum up to zero. Also, note that the first quantity in [19] is not observable.

Intuitively, because the g-weights provide perfect estimates for the means of the auxiliary variables, they must perform well for the response variables if the models are adequate. Let us define

$$[20] \qquad \Delta = (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t(\hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_1)$$

Then, using [19] and [15] we obtain after some algebra the following formal decomposition of the regression estimate

$$[21] \qquad \hat{Y}_{greg} = \frac{1}{n_1}\sum_{x\in s_1} g_1^{(1)}(x)Y(x) + \frac{1}{n_2}\sum_{x\in s_2} g_2(x)Y(x) - \frac{1}{n_1}\sum_{x\in s_1} Y(x) + \Delta$$

This is a purely formal identity because only the second term is observable. In the definition [20] the first factor is of order $O(n_1^{-\frac{1}{2}})$ in design-probability and likewise the second factor of order $O(n_2^{-\frac{1}{2}})$. Thus Δ is of order $= O(n_2^{-1})$ in design-probability and can be neglected with respect to the first three terms in [21], which are of order $O(1)$.

In many applications with categorical explanatory variables it can happen that the matrices occurring in [17] are singular, in which case generalized inverse must be used instead.

In this case any particular solution $\hat{\boldsymbol{\beta}}_2^*$ of the consistent normal equations can be used and all statistically relevant quantities like g-weights and predictions are independent of the particular solution chosen and equations [18] remain valid (see Renssen and Martinus (2002) for details and further references).

We introduce the theoretical residuals $R_1(x)$ and $R(x)$ by the relations

$$\begin{aligned} Y(x) &= R_1(x) + \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\beta}^{(1)} \\ Y(x) &= R(x) + \boldsymbol{Z}^{t}(x)\boldsymbol{\beta} \end{aligned} \qquad [22]$$

Substituting these equalities into [21] and using [18] we obtain following expression for the error term

$$\begin{aligned} \hat{Y}_{greg} - \bar{Y} &= \Big(\frac{1}{n_1}\sum_{x\in s_1} g_1^{(1)}(x)R_1(x) - \frac{1}{\lambda}\int_F R_1(x)dx\Big) \\ &+ \Big(\frac{1}{n_2}\sum_{x\in s_2} g_2(x)R(x) - \frac{1}{n_1}\sum_{x\in s_1} R(x)\Big) + \Delta \end{aligned} \qquad [23]$$

This is the Monte Carlo version of equation 9.7.18 given in Särndal et al. (2003).

According to general heuristic principles (described in Mandallaz (2008), section 6.2 and Särndal et al. (2003), section 6.6) it can be expected that the following variance estimate based on the g-weights has better performances

$$\hat{\mathbb{V}}(\hat{Y}_{greg}) = \frac{1}{n_1}\frac{1}{n_2}\sum_{x\in s_2}(g_1^{(1)}(x))^2\hat{R}_1^2(x) + \frac{1}{n_2}(1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{x\in s_2} g_2^2(x)\hat{R}^2(x) \qquad [24]$$

This is the perfect Monte Carlo analogy of equation 9.7.22 in Särndal et al. (2003).

We now propose a different technique to obtain the design-based variance of $\hat{Y}_{greg}$ which is better suited for the small-area estimation problem. The starting point is the following important result for the design-based variance of the regression coefficients based on the

Taylor linearization technique (for proofs see Mandallaz (2008), p. 125 and Mandallaz (2012)) which leads to the asymptotic covariance matrices

$$[25] \qquad \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_k} = \boldsymbol{A}^{-1}\Big(\frac{1}{n_k}\mathbb{E}R^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\Big)\boldsymbol{A}^{-1}$$

where $\boldsymbol{A} = \mathbb{E}_x\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)$ and

$$[26] \qquad \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_k} = (\boldsymbol{A}^{(1)})^{-1}\Big(\frac{1}{n_k}\mathbb{E}R_1^2(x)\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)\Big)(\boldsymbol{A}^{(1)})^{-1}$$

We use the variance decomposition [7] on [15]. One gets

$$\mathbb{E}_{2|1}\hat{Y}_{greg} = (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t\hat{\boldsymbol{\alpha}}_1 + \hat{\bar{\boldsymbol{Z}}}_1^t\hat{\boldsymbol{\beta}}_1 \approx \bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\alpha}}_1$$

because $\hat{\bar{\boldsymbol{Z}}}_1^{(1)t}\hat{\boldsymbol{\alpha}}_1$ and $\hat{\boldsymbol{Z}}_1^t\hat{\boldsymbol{\beta}}_1$ both tend to $\bar{Y}$ asymptotically. Therefore, one has

$$\mathbb{V}_1\mathbb{E}_{2|1}(\hat{Y}_{greg}) = \bar{\boldsymbol{Z}}^{(1)t}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_1}\bar{\boldsymbol{Z}}^{(1)} = \frac{n_2}{n_1}\bar{\boldsymbol{Z}}^{(1)t}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_2}\bar{\boldsymbol{Z}}^{(1)}$$

To calculate $\mathbb{V}_{2|1}(\hat{Y}_{greg})$ we note that

$$\hat{Y}_{greg} - \mathbb{E}_{2|1}Y_{greg} = (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t(\hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_1) + \hat{\bar{\boldsymbol{Z}}}_1^t(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \approx \hat{\bar{\boldsymbol{Z}}}_1^t(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)$$

because by the law of large numbers the first term is of order $O(n_2^{-1})$ and the second of order $O(n_2^{-\frac{1}{2}})$. Using the Taylor expansion given in Mandallaz (2008) (at point $(\boldsymbol{A}_1,\ \hat{\boldsymbol{\beta}}_1)$ instead of $(\boldsymbol{A},\ \boldsymbol{\beta})$) we get

$$\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1 \approx \boldsymbol{A}_1^{-1}\big(\frac{1}{n_2}\sum_{x\in s_2}R(x)\boldsymbol{Z}(x)\big)$$

By the properties of simple random sampling without replacement within s_1 and the approximations $\frac{n_2-1}{n_1-1} \approx \frac{n_2}{n_1}$, $\sum_{x\in s_1} R(x)\boldsymbol{Z}(x) = 0$ (orthogonality relationship), we obtain after some algebra

$$\mathbb{E}_1\mathbb{V}_{2|1}(\hat{Y}_{greg}) = (1-\frac{n_2}{n_1})\bar{\boldsymbol{Z}}^t\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_2}\bar{\boldsymbol{Z}}$$

To get asymptotic estimates of the covariances matrices one can replace $\boldsymbol{A}$ by $\boldsymbol{A}_1$ or $\boldsymbol{A}_2$, and likewise $\boldsymbol{A}^{(1)}$ by $\boldsymbol{A}_1^{(1)}$ or $\boldsymbol{A}_2^{(1)}$. To ensure the important calibration properties [18] we will use

$$\hat{\mathbb{V}}(\hat{Y}_{greg}) = \frac{n_2}{n_1}\bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2}\bar{\boldsymbol{Z}}^{(1)} + (1-\frac{n_2}{n_1})\hat{\bar{\boldsymbol{Z}}}_1^t\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2}\hat{\bar{\boldsymbol{Z}}}_1 \quad [27]$$

with

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2} = \boldsymbol{A}_2^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2}\hat{R}^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\Big)\boldsymbol{A}_2^{-1} \quad [28]$$

and

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} = (\boldsymbol{A}_1^{(1)})^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2}\hat{R}_1^2(x)\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)\Big)(\boldsymbol{A}_1^{(1)})^{-1} \quad [29]$$

are the estimated **design-based covariance matrices** of the regression coefficients under the large and reduced models, which have also been discussed in a totally different context, i.e. model-dependent least squares theory under non-standard conditions, by Huber (1967) and Gregoire and Dyer (1989), they are sometimes called **robust covariance matrices**. It is straightforward to see that [27] and [24] are equal.

To get further insight into $\hat{Y}_{greg}$ we note that one can write

$$\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{A}_2^{(1)} & \boldsymbol{A}_2^{(12)} \\ \boldsymbol{A}_2^{(12)t} & \boldsymbol{A}_2^{(2)} \end{bmatrix} \quad [30]$$

with $\boldsymbol{A}_2^{(k)} = \frac{1}{n_2}\sum_{x\in s_2}\boldsymbol{Z}^{(k)}(x)\boldsymbol{Z}^{(k)t}(x),\ k = 1,2$ and $\boldsymbol{A}_2^{(12)} = \frac{1}{n_2}\sum_{x\in s_2}\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(2)t}(x)$. Developing the normal equations accordingly one obtains after some algebra the well-known relation

$$\hat{\boldsymbol{\beta}}_2^{(1)} = \hat{\boldsymbol{\alpha}}_2 - (\boldsymbol{A}_2^{(1)})^{-1}\boldsymbol{A}_2^{(12)}\hat{\boldsymbol{\beta}}_2^{(2)} \qquad [31]$$

where $\hat{\boldsymbol{\beta}}_2^t = (\hat{\boldsymbol{\beta}}_2^{(1)t}, \hat{\boldsymbol{\beta}}_2^{(2)t})$. Substituting in [15] we obtain

$$\hat{Y}_{greg} = \bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\beta}}_2^{(1)} + \hat{\bar{\boldsymbol{Z}}}_1^{(2)t}\hat{\boldsymbol{\beta}}_2^{(2)} + \delta$$

where $\delta = (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}^{(1)})^t(\boldsymbol{A}_2^{(1)})^{-1}\boldsymbol{A}_2^{(1,2)}\hat{\boldsymbol{\beta}}_2^{(2)}$, which is of order $O(n_1^{-\frac{1}{2}})$ in design-probability. Hence, we have the asymptotic equivalence

$$\hat{Y}_{greg} \doteq \bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\beta}}_2^{(1)} + \hat{\bar{\boldsymbol{Z}}}_1^{(2)t}\hat{\boldsymbol{\beta}}_2^{(2)} \qquad [32]$$

which is intuitively very appealing: the exhaustive component $\boldsymbol{Z}^{(1)}(x)$ occurs with its known true mean and the non-exhaustive component $\boldsymbol{Z}^{(2)}(x)$ with its estimated mean from the large sample, as compared with the classical two-phase estimator

$$\hat{Y}_{reg} = \hat{\bar{\boldsymbol{Z}}}^t\hat{\boldsymbol{\beta}}_2 = \hat{\bar{\boldsymbol{Z}}}_1^{(1)t}\hat{\boldsymbol{\beta}}_2^{(1)} + \hat{\bar{\boldsymbol{Z}}}_1^{(2)t}\hat{\boldsymbol{\beta}}_2^{(2)} \qquad [33]$$

One could also consider

$$\hat{Y}_{gregmod} = \bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\beta}}_2^{(1)} + \hat{\bar{\boldsymbol{Z}}}_1^{(2)t}\hat{\boldsymbol{\beta}}_2^{(2)} \qquad [34]$$

as a further estimator in its own right. If $\boldsymbol{A}_2^{(1,2)} = \boldsymbol{0}$, i.e. if the exhaustive and non-exhaustive components are orthogonal ("independent") then we have exactly $\hat{Y}_{gregmod} = \hat{Y}_{greg}$. We do not advocate the use of $\hat{Y}_{gregmod}$ when $\boldsymbol{Z}^{(1)}(x)$ and $\boldsymbol{Z}^{(2)}(x)$ are non-orthogonal

because, as we found out by simulations, $\hat{Y}_{gregmod}$ can have a larger variance than $\hat{Y}_{reg}$ with moderate sample sizes. Beside, the resulting formulae for the asymptotic covariance are more cumbersome than [27]. For these reasons we shall not consider $\hat{Y}_{gregmod}$ any more and we now proceed to adapt the previous results to the important small-area estimation problem.

4 Generalized small-area estimators

We consider a small area $G \subset F$ and we want to estimate

$$\bar{Y}_G = \frac{1}{\lambda(G)} \sum_{i=1}^{N} I_G(i) Y_i = \frac{1}{\lambda(G)} \int_G Y(x) dx$$

where $I_G(i) = 1$ if the i-th tree is in G, otherwise $I_G(i) = 0$. Strictly speaking the last equality holds if boundary adjustments are performed in G, whereas they are in most instances only performed with respect to F. We shall need the following notation: $s_{1,G} = s_1 \cap G$, $s_{2,G} = s_2 \cap G$, $n_{k,G} = \sum_{s \in s_2} I_G(x)$, $k = 1, 2$ (restriction of the samples and sample sizes to G, note that the $n_{k,G}$ are random variables). The simplest solution is to restrict the generalized regression estimator [6] to G, i.e. to consider **the generalized small-area estimator**

$$[35] \quad \hat{Y}_{G,greg} = \frac{1}{\lambda(G)} \int_G \hat{Y}_1(x) dx + \frac{1}{n_{1,G}} \sum_{x \in s_{1,G}} (\hat{Y}(x) - \hat{Y}_1(x)) + \frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} (Y(x) - \hat{Y}(x))$$

and treat the internal model as an external one to obtain the estimated conditional variance (i.e. given the $n_{k,G}$)

[36]

$$\hat{\mathbb{V}}(\hat{Y}_{greg}) = \frac{1}{n_{1,G}} \frac{1}{n_{2,G} - 1} \sum_{x \in s_{2,G}} (\hat{R}_1(x) - \hat{\bar{R}}_{1,G})^2 + (1 - \frac{n_{2,G}}{n_{1,G}}) \frac{1}{n_{2,G}(n_{2,G} - 1)} \sum_{x \in s_2} (\hat{R}(x) - \hat{\bar{R}}_G)^2$$

where $\hat{\bar{R}}_{1,G} = \frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} R_1(x)$ and likewise for $\hat{\bar{R}}_G$. This variance estimate neglects the uncertainty of the regression coefficients but there is empirical evidence that this is acceptable in large samples (see Mandallaz (2012) for examples with $\hat{Y}_{reg}$).
We can rewrite $\hat{Y}_{G,greg}$ as

$$\hat{Y}_{G,greg} = (\bar{\boldsymbol{Z}}_G^{(1)} - \hat{\bar{\boldsymbol{Z}}}_G^{(1)})^t \hat{\boldsymbol{\alpha}}_2 + \hat{\bar{\boldsymbol{Z}}}_{1,G}^t \hat{\boldsymbol{\beta}}_2 + \frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} \hat{R}(x) \qquad [37]$$

where we have set

$$\bar{\boldsymbol{Z}}_G^{(1)} = \frac{1}{\lambda(G)} \int_G \boldsymbol{Z}^{(1)}(x) dx, \;\; \hat{\bar{\boldsymbol{Z}}}_{1,G} = \frac{1}{n_{1,G}} \sum_{x \in s_{1,G}} \boldsymbol{Z}(x) = (\hat{\bar{\boldsymbol{Z}}}_{1,G}^{(1)t}, \hat{\bar{\boldsymbol{Z}}}_{1,G}^{(2)t})^t$$

The essential difference with $\hat{Y}_{greg} = \hat{Y}_{F,greg}$ is that the mean residual term in [37] does no longer vanish in general, which makes the calculation of the variance very difficult. To bypass this difficulty we use the technique presented in Mandallaz (2012) by extending the model with the indicator variable $I_G(x)$ of the small area G, **which insures zero mean residual over** F **and** G. We can include $I_G(x)$ in $\boldsymbol{Z}^{(1)}(x)$ or $\boldsymbol{Z}^{(2)}(x)$. It seems more natural to include it in the first component so that the zero mean residual properties will hold for both the reduced and the large model. Also, it is reasonable to assume that the perimeter and consequently the surface area of G are known. We consider therefore the following extended models with auxiliary vectors: $\boldsymbol{\mathcal{Z}}^t(x) = (\boldsymbol{\mathcal{Z}}^{(1)t}(x), \boldsymbol{\mathcal{Z}}^{(2)t}(x))$, where $\boldsymbol{\mathcal{Z}}^{(1)t}(x) = (\boldsymbol{Z}^{(1)t}(x), I_G^t(x))$ and $\boldsymbol{\mathcal{Z}}^{(2)t}(x) = \boldsymbol{Z}^{(2)t}(x)$. To have a uniform notation throughout we also change the notation for the second component, i.e. we will use $\boldsymbol{\mathcal{Z}}^{(2)}(x)$ instead of $\boldsymbol{Z}^{(2)}(x)$ in this section. We have therefore the following set up

1. The large extended model M

$$Y(x) = \boldsymbol{\mathcal{Z}}(x)^t \boldsymbol{\theta} + \mathcal{R}(x) = \boldsymbol{\mathcal{Z}}^{(1)t}(x) \boldsymbol{\theta}^{(1)} + \boldsymbol{\mathcal{Z}}^{(2)t}(x) \boldsymbol{\theta}^{(2)} + \mathcal{R}(x)$$

with $\boldsymbol{\theta}^t = (\boldsymbol{\theta}^{(1)t}, \boldsymbol{\theta}^{(2)t})$. The intercept term is contained in $\boldsymbol{\mathcal{Z}}^{(1)}(x)$ or it is a linear combination of its components.

The theoretical regression parameter $\boldsymbol{\theta}$ minimizes

$$\int_F (Y(x) - \boldsymbol{\mathcal{Z}}^t(x)\boldsymbol{\theta})^2 dx$$

It satisfies the normal equation

$$\Big(\int_F \boldsymbol{\mathcal{Z}}(x)\boldsymbol{\mathcal{Z}}^t(x)dx\Big)\boldsymbol{\theta} = \int_F (Y(x)\boldsymbol{\mathcal{Z}}(x)dx)$$

and the orthogonality relationship

$$\int_F \mathcal{R}(x)\boldsymbol{\mathcal{Z}}(x)dx = \mathbf{0}$$

in particular the zero mean residual properties

$$\frac{1}{\lambda(F)}\int_F \mathcal{R}(x)dx = \frac{1}{\lambda(G)}\int_G \mathcal{R}(x)dx = 0$$

2. The reduced extended model M_1

$$Y(x) = \boldsymbol{\mathcal{Z}}^{(1)t}(x)\boldsymbol{\gamma} + \mathcal{R}_1(x)$$

The theoretical regression parameter $\boldsymbol{\gamma}$ minimizes

$$\int_F (Y(x) - \boldsymbol{\mathcal{Z}}^{(1)t}(x)\boldsymbol{\gamma})^2 dx$$

It satisfies the normal equation

$$\Big(\int_F \boldsymbol{\mathcal{Z}}^{(1)}(x)\boldsymbol{\mathcal{Z}}^{(1)t}(x)dx\Big)\boldsymbol{\gamma} = \int_F Y(x)\boldsymbol{\mathcal{Z}}^{(1)}(x)dx$$

and the orthogonality relationship

$$\int_G R_1(x)\boldsymbol{\mathcal{Z}}^{(1)}(x)dx = \mathbf{0}$$

in particular the zero mean residual properties

$$\frac{1}{\lambda(F)}\int_F \mathcal{R}_1(x)dx = \frac{1}{\lambda(G)}\int_G \mathcal{R}_1(x)dx = 0$$

We can obviously apply mutatis mutandis all the previous results. The estimated regression coefficients are

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_2 &= \Big(\frac{1}{n_2}\sum_{x\in s_2}\boldsymbol{\mathcal{Z}}^{(1)}(x)\boldsymbol{\mathcal{Z}}^{(1)t}(x)\Big)^{-1}\frac{1}{n_2}\sum_{x\in s_2}Y(x)\boldsymbol{\mathcal{Z}}^{(1)}(x)\\ &:= (\boldsymbol{\mathcal{A}}_2^{(1)})^{-1}\frac{1}{n_2}\sum_{x\in s_2}Y(x)\boldsymbol{\mathcal{Z}}^{(1)}(x)\end{aligned} \tag{38}$$

and

$$\hat{\boldsymbol{\theta}}_2 = \Big(\frac{1}{n_2}\sum_{x\in s_2}\boldsymbol{\mathcal{Z}}(x)\boldsymbol{\mathcal{Z}}^t(x)\Big)^{-1}\frac{1}{n_2}\sum_{x\in s_2}Y(x)\boldsymbol{\mathcal{Z}}(x) := \boldsymbol{\mathcal{A}}_2^{-1}\frac{1}{n_2}\sum_{x\in s_2}Y(x)\boldsymbol{\mathcal{Z}}(x) \tag{39}$$

The estimated covariance matrices are according to [28] and [29]

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_2} &= \boldsymbol{\mathcal{A}}_2^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2}\hat{\mathcal{R}}^2(x)\boldsymbol{\mathcal{Z}}(x)\boldsymbol{\mathcal{Z}}^t(x)\Big)\boldsymbol{\mathcal{A}}_2^{-1}\\ \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_2} &= (\boldsymbol{\mathcal{A}}_1^{(1)})^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2}\hat{\mathcal{R}}_1^2(x)\boldsymbol{\mathcal{Z}}(x)\boldsymbol{\mathcal{Z}}^t(x)\Big)(\boldsymbol{\mathcal{A}}_1^{(1)})^{-1}\end{aligned} \tag{40}$$

where $\hat{\mathcal{R}}(x) = Y(x) - \boldsymbol{\mathcal{Z}}^t(x)\hat{\boldsymbol{\theta}}_2$ and $\hat{\mathcal{R}}_1(x) = Y(x) - \boldsymbol{\mathcal{Z}}^{(1)t}(x)\hat{\boldsymbol{\gamma}}_2$ are the residuals.

Because the sum of the residuals over $s_{2,G}$ is now zero we can write the new small-area estimator $\hat{\hat{Y}}_{G,greg}$ as in [15]

[41] $$\hat{\hat{Y}}_{G,greg} = (\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)} - \hat{\bar{\boldsymbol{\mathcal{Z}}}}_G^{(1)})^t\hat{\boldsymbol{\gamma}}_2 + \hat{\bar{\boldsymbol{\mathcal{Z}}}}_G^t\hat{\boldsymbol{\theta}}_2$$

where we have set

$$\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)} = \frac{1}{\lambda(G)}\int_G \mathcal{Z}(x)dx,\ \hat{\bar{\boldsymbol{\mathcal{Z}}}}_G^{(1)} = \frac{1}{n_{1,G}}\sum_{x\in s_{1,G}} \boldsymbol{\mathcal{Z}}^{(1)}(x)$$

To get an estimate of the design-based variance we use mutatis mutandis [27]

[42] $$\hat{\mathbb{V}}(\hat{\hat{Y}}_{greg}) = \frac{n_2}{n_1}\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_2}\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)} + (1 - \frac{n_2}{n_1})\hat{\bar{\boldsymbol{\mathcal{Z}}}}_G^t\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_2}\hat{\bar{\boldsymbol{\mathcal{Z}}}}_G$$

5 Generalization to cluster sampling

We follow the description of cluster sampling as defined in Mandallaz (2008) (especially section 5.5) and Mandallaz (2012). A cluster is identified by its origin x, uniformly distributed in $\tilde{F} \supset F$. The geometry of the cluster is given by M vectors $e_1, \dots e_M$ defining the random cluster $x_l = x + e_l$. $M(x) = \sum_{l=1}^M I_F(x_l)$ is the random number of points of the cluster falling into the forest area F. We define the local density at the cluster level by $Y_c(x) = \frac{\sum_{l=1}^M I_F(x_l)Y(x_l)}{M(x)}$, likewise we set $\boldsymbol{Z}_c(x) = \frac{\sum_{l=1}^M I_F(x_l)\boldsymbol{Z}(x_l)}{M(x)}$. The set $\tilde{F}$ above can be mathematically defined as the smallest set $\{x \in \mathcal{R}^2 \mid M(x) \neq 0\}$. In the first phase we have n_1 clusters identified by $x \in s_1$ and in the second phase n_2 clusters

with $x \in s_2$, obtained by simple random sampling from s_1.

We shall use the model-based approach, in which the regression coefficient $\boldsymbol{\beta}_c$ at the cluster level, under the large model with $\boldsymbol{Z}^t = (\boldsymbol{Z}^{(1)t}(x), \boldsymbol{Z}^{(2t)}(x))$, minimizes

$$\int_F M(x)(Y_c(x) - \boldsymbol{\beta}_c^t \boldsymbol{Z}_c(x))^2 dx$$

In the pure design-based approach the weights will be $M^2(x)$ but this leads to non-zero mean residual (thought close zero in practice), and the definitions of the regression estimator and of the normal equation are slightly different (see Mandallaz (2008), section 5.5 for details). The choice of $M(x)$ rather than $M^2(x)$ as weights is suggested by the model-dependent approach. When $Y_c(x)$ is the mean of the $M(x)$ observations, its variance can be expected to be inversely proportional to $M(x)$. This procedure leads to the normal equation

$$\Big(\int_F M(x) \boldsymbol{Z_c}(x) \boldsymbol{Z_c}^t(x) dx \Big) \boldsymbol{\beta}_c = \int_F M(x) Y_c(x) \boldsymbol{Z_c}(x) dx$$

and to $\int_F M(x) R_c(x) = 0$. An asymptotically design-unbiased estimate $\hat{\boldsymbol{\beta}}_{c,2}$ for $\boldsymbol{\beta}_c$ can be obtained by taking a sample copy of the above equation, i.e.

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{c,2} &= \Big(\frac{1}{n_2} \sum_{x \in s_2} M(x) \boldsymbol{Z}_c(x) \boldsymbol{Z}_c^t(x) \Big)^{-1} \Big(\frac{1}{n_2} \sum_{x \in s_2} M(x) Y_c(x) \boldsymbol{Z}_c(x) \Big) \\ &:= \boldsymbol{A}_{c,s_2}^{-1} \Big(\frac{1}{n_2} \sum_{x \in s_2} M(x) Y_c(x) \boldsymbol{Z}_c(x) \Big) \end{aligned} \qquad [43]$$

The empirical residuals at the cluster level are

$$\hat{R}_c(x) = Y_c(x) - \boldsymbol{Z}_c^t(x) \hat{\boldsymbol{\beta}}_{c,2}$$

which satisfy the orthogonality relation

$$\sum_{x\in s_2} M(x)\hat{R}_c(x)\boldsymbol{Z}_c(x) = 0$$

and in particular the zero mean residual property

$$\frac{\sum_{x\in s_2} M(x)\hat{R}_c(x)}{\sum_{x\in s_2} M(x)} = 0$$

With obvious notational changes we get the corresponding results under the reduced model with $\boldsymbol{Z}^{(1)}$ alone, where the regression coefficient $\boldsymbol{\alpha}_c$ minimizes

$$\int_F M(x)(Y_c(x) - \boldsymbol{\alpha}_c^t \boldsymbol{Z}_c^{(1)}(x))^2 dx$$

with estimate

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{c,2} &= \Big(\frac{1}{n_2}\sum_{x\in s_2} M(x)\boldsymbol{Z}_c^{(1)}(x)\boldsymbol{Z}_c^{(1)t}(x)\Big)^{-1}\Big(\frac{1}{n_2}\sum_{x\in s_2} M(x)Y_c(x)\boldsymbol{Z}_c(x)\Big) \\ [44] \qquad &:= (\boldsymbol{A}_{c,2}^{(1)})^{-1}\Big(\frac{1}{n_2}\sum_{x\in s_2} M(x)Y_c(x)\boldsymbol{Z}_c^{(1)}(x)\Big) \end{aligned}$$

The residuals for the reduced model are $\hat{R}_{1,c}(x) = Y_c(x) - \boldsymbol{Z}_c^{(1)t}(x)\hat{\boldsymbol{\alpha}}_{c,2}$ and enjoy the same properties as $\hat{R}_c(x)$.

Using mutatis mutandis exactly the same arguments as in simple random sampling we get the asymptotic robust design-based estimated covariance matrices

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,2}} &= \boldsymbol{A}_{c,2}^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2} M^2(x)\hat{R}_c^2(x)\boldsymbol{Z}_c(x)\boldsymbol{Z}_c^t(x)\Big)\boldsymbol{A}_{c,2}^{-1} \\ [45] \qquad \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{c,2}} &= (\boldsymbol{A}_{c,1}^{(1)})^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2} M^2(x)\hat{R}_{1,c}^2(x)\boldsymbol{Z}_c(x)\boldsymbol{Z}_c^t(x)\Big)(\boldsymbol{A}_{c,1}^{(1)})^{-1} \end{aligned}$$

where $\boldsymbol{A}_{c,1}^{(1)} = \frac{1}{n_1}\sum_{x\in s_1} M(x)\boldsymbol{Z}_c^{(1)}(x)\boldsymbol{Z}_c^{(1)t}(x)$.

We define the **generalized regression estimator in cluster sampling** directly by adapting [15]

$$\hat{Y}_{c,greg} = (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_{c,1}^{(1)})^t\hat{\boldsymbol{\alpha}}_{c,2} + \hat{\bar{\boldsymbol{Z}}}_{c,1}^t\hat{\boldsymbol{\beta}}_{c,2} \quad [46]$$

where the mean is now defined as

$$\hat{\bar{\boldsymbol{Z}}}_{c,1}^t = \frac{\sum_{x\in s_1} M(x)\boldsymbol{Z}_c(x)}{\sum_{x\in s_1} M(x)}$$

and similarly for $\hat{\bar{\boldsymbol{Z}}}_{c,1}^{(1)}$.

With the same technique as in simple random sampling we obtain the estimated design-based variance

$$\hat{\mathbb{V}}(\hat{Y}_{c,greg}) = \frac{n_2}{n_1}\bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{c,2}}\bar{\boldsymbol{Z}}^{(1)} + (1-\frac{n_2}{n_1})\hat{\bar{\boldsymbol{Z}}}_{c,1}^t\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,2}}\hat{\bar{\boldsymbol{Z}}}_{c,1} \quad [47]$$

For small-area estimation we shall work with the extended model

$$\boldsymbol{\mathcal{Z}}^t(x) = (\boldsymbol{\mathcal{Z}}^{(1)t}(x), \boldsymbol{\mathcal{Z}}^{(2)t}(x))$$

with $\boldsymbol{\mathcal{Z}}^{(1)t}(x) = (\boldsymbol{Z}^{(1)t}(x), I_G^t(x))$, $\boldsymbol{\mathcal{Z}}^{(2)t}(x) = \boldsymbol{Z}^{(2)t}(x)$ and $I_G(x)$ is the indicator of the small area G. At the cluster level we have $\boldsymbol{\mathcal{Z}}_c^{(1)t}(x) = (\boldsymbol{Z}_c^{(1)t}(x), I_{c,G}^t(x))$ where $I_{c,G}(x) = \frac{\sum_{l=1}^M I_G(x_l)}{M(x)}$. In extensive inventories we can reasonably assume that all the points of a cluster lying in the forest area F will belong to the same small area G so that in fact $I_{c,G}(x) \equiv 1$ for all $x \in \tilde{G} = \{x \mid \sum_{l=1}^M I_G(x_l) > 0\}$. This ensures again that we will have zero mean residual over F and G. For the regression estimates in the extended model we

have

$$\hat{\boldsymbol{\gamma}}_{c,2} = (\boldsymbol{\mathcal{A}}_{c,2}^{(1)})^{-1}\Big(\frac{1}{n_2}\sum_{x\in s_2} M(x)\boldsymbol{\mathcal{Z}}_c^{(1)}(x)Y(x)\Big) \quad [48]$$

with $\boldsymbol{\mathcal{A}}_{c,2}^{(1)} = \frac{1}{n_2}\sum_{x\in s_2} M(x)\boldsymbol{\mathcal{Z}}_c^{(1)}(x)\boldsymbol{\mathcal{Z}}_c^{(1)t}(x)$. Likewise we get

$$\hat{\boldsymbol{\theta}}_{c,2} = \boldsymbol{\mathcal{A}}_{c,2}^{-1}\Big(\frac{1}{n_2}\sum_{x\in s_2} M(x)\boldsymbol{\mathcal{Z}}_c(x)Y(x)\Big) \quad [49]$$

with $\boldsymbol{\mathcal{A}}_{c,2} = \frac{1}{n_2}\sum_{x\in s_2} M(x)\boldsymbol{\mathcal{Z}}_c(x)\boldsymbol{\mathcal{Z}}_c^{t}(x)$.

We define as in [41] the generalized small-area estimator by

$$\hat{\hat{Y}}_{c,G,greg} = (\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)} - \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{c,G}^{(1)})^t\hat{\boldsymbol{\gamma}}_{c,2} + \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{c,G}^{t}\hat{\boldsymbol{\theta}}_{c,2} \quad [50]$$

where we have set

$$\bar{\boldsymbol{\mathcal{Z}}}_G^{(1)} = \frac{1}{\lambda(G)}\int_G \mathcal{Z}(x)dx,\ \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{c,G}^{(1)} = \frac{\sum_{x\in s_{1,G}} M(x)\boldsymbol{\mathcal{Z}}_c^{(1)}(x)}{\sum_{x\in s_{1,G}} M(x)},\ \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{c,G} = \frac{\sum_{x\in s_{1,G}} M(x)\boldsymbol{\mathcal{Z}}_c(x)}{\sum_{x\in s_{1,G}} M(x)}$$

The estimated design-based covariance matrix are now

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_{c,2}} &= (\boldsymbol{\mathcal{A}}_{c,1}^{(1)})^{-1}\Big(\frac{1}{n_2^2}\sum_{\in s_2} M^2(x)\hat{\mathcal{R}}_{1,c}^2(x)\boldsymbol{\mathcal{Z}}_c^{(1)}(x)\boldsymbol{\mathcal{Z}}_c^{(1)t}(x)\Big)(\boldsymbol{\mathcal{A}}_{c,1}^{(1)})^{-1} \\
\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_{c,2}} &= \boldsymbol{\mathcal{A}}_{c,2}^{-1}\Big(\frac{1}{n_2^2}\sum_{\in s_2} M^2(x)\hat{\mathcal{R}}_c^2(x)\boldsymbol{\mathcal{Z}}_c(x)\boldsymbol{\mathcal{Z}}_c^{t}(x)\Big)\boldsymbol{\mathcal{A}}_{c,2}^{-1} \quad [51]
\end{aligned}$$

with the residuals in the extended models $\hat{\mathcal{R}}_{1,c}(x) = Y_c(x) - \boldsymbol{\mathcal{Z}}_c^{(1)t}(x)\hat{\boldsymbol{\gamma}}_{c,2}$ and $\hat{\mathcal{R}}_c(x) = Y_c(x) - \boldsymbol{\mathcal{Z}}_c^{t}(x)\hat{\boldsymbol{\theta}}_{c,2}$ and $\boldsymbol{\mathcal{A}}_{c,1} = \frac{1}{n_1}\sum_{x\in s_1} M(x)\boldsymbol{\mathcal{Z}}_c(x)\boldsymbol{\mathcal{Z}}_c^{t}(x)$.

We obtain as in [27] the estimated design-based variance

[52] $$\hat{\mathbb{V}}(\hat{Y}_{c,G,greg}) = \frac{n_2}{n_1}\bar{\boldsymbol{z}}_G^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_{c,2}}\bar{\boldsymbol{z}}_G^{(1)} + (1-\frac{n_2}{n_1})\hat{\bar{\boldsymbol{z}}}_{c,G}^{t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_{c,2}}\hat{\bar{\boldsymbol{z}}}_{G,1}$$

As shown in Mandallaz (2012) it is straightforward to consider simultaneously several small areas by extending the model with as many small area indicator variables.

6 Generalization to two-stage sampling

In many applications costs to measure the response variable Y_i are high. For instance, a good determination of the volume may require that one records DBH, as well as the diameter at $7m$ above ground and total height in order to utilize a three-way volume function. However, one could rely on a coarser, but cheaper, approximation of the volume based only on DBH. Nonetheless, it may be most sensible to assess those three parameters only on a sub-sample of trees. We now briefly formalize this simple idea, which is used in the Swiss National Forest Inventory. The reader is referred to (Mandallaz (2008), section 4.4, 4.5, 5.4 and 9.5) for details. For each point $x \in s_2$ trees are drawn with probabilities π_i. The set of selected trees is denoted by $s_2(x)$. From each of the selected trees $i \in s_2(x)$ one gets an approximation Y_i^* of the exact value Y_i. From the finite set $s_2(x)$ one draws a sub-sample $s_3(x) \subset s_2(x)$ of trees by Poisson sampling. For each tree $i \in s_3(x)$ one then measures the exact variable Y_i. Let us now define the second stage indicator variable

[53] $$J_i(x) = \begin{cases} 1 \text{ if } i \in s_3(x) \\ 0 \text{ if } i \notin s_3(x) \end{cases}$$

To construct a good point estimate, we must have the residual $R_i = Y_i - Y_i^*$ which is known only for trees $i \in s_3(x)$. The generalized local density $Y^*(x)$ is defined according

to

$$Y^*(x) = \frac{1}{\lambda(F)} \left(\sum_{i=1}^{N} \frac{I_i(x) Y_i^*}{\pi_i} + \sum_{i=1}^{N} \frac{I_i(x) J_i(x) R_i}{\pi_i p_i} \right)$$

$$= \frac{1}{\lambda(F)} \left(\sum_{i \in s_2(x)} \frac{Y_i^*}{\pi_i} + \sum_{i \in s_3(x)} \frac{R_i}{\pi_i p_i} \right) \quad [54]$$

where the p_i are the conditional inclusion probabilities for the the second stage sampling, i.e. $p_i = \mathbb{P}(J_i(x) = 1 \mid I_i(x) = 1)$. It follows from general principles presented in (Mandallaz (2008), sections 4.4 and 4.5) that one can use all the previous results by replacing everywhere the exact local densities $Y(x)$, or $Y(x_l)$ in cluster sampling, by the corresponding generalized local densities $Y^*(x)$ or $Y^*(x_l)$. The second-stage variance is automatically taken into account.

7 Example: Post-stratification

We consider the forested area F embedded in a subset $\tilde{F}$, which in many national inventories is simply the whole country or part of it. Hence, $\tilde{F} \setminus F = F_0$ is the non-forested area. The forested area itself is partitioned in L strata F_k, i.e. $F = \cup_{k=1}^{L} F_k$. We assume that the surface areas $\lambda(\tilde{F})$ and $\lambda(F_0)$ are known exactly, **whereas the surface areas of the strata $\lambda(F_k)$ for $k \geq 1$ are not**. The first-phase sampling consists of n_1 points, uniformly distributed in $\tilde{F}$ (set s_1), out of which n_2 points (set s_2) are selected by equal probability sampling without replacement (in practice one uses grids and sub-grids). The local density $Y(x)$ is set to zero whenever $x \in F_0$ (even if trees in F could be selected from $x \in F_0$), and is calculated according to [2] if $x \in F$ (i.e. with boundary adjustments if necessary). This ensures that $\int_{\tilde{F}} Y(x)dx = \int_F Y(x)dx = \sum_{i=1}^{N} Y_i$. We define the (random) sample sizes $n_{1k} = \sum_{x \in s_1} I_{F_k}(x)$, $n_{2k} = \sum_{x \in s_2} I_{F_k}(x)$ for $k \geq 0$ and $n_{1F} = \sum_{k=1}^{L} n_{1k}$, $n_{2F} = \sum_{k=1}^{L} n_{2k}$.

For the reduced model we define the one dimensional explanatory vector according to $\boldsymbol{Z}^{(1)}(x) = I_F(x)$. Trivial calculations lead to $\hat{\boldsymbol{\alpha}} = \hat{\bar{Y}}_F^t = \frac{1}{n_{2,F}} \sum_{x \in s_2} Y(x)$ (i.e. the empirical mean of $Y(x)$ in the forested area F). We get $\bar{\boldsymbol{Z}}^{(1)} = p_F = \frac{\lambda(F)}{\lambda(\tilde{F})}$. For the g-weight one obtains $g^{(1)}(x) = p_F \frac{n_1}{n_{1F}} I_F(x)$. For the predictions one obtains $\hat{Y}_1(x) = \hat{\bar{Y}}_F I_F(x)$.

For the large model, we define $\boldsymbol{Z}^{(2)}(x)$ as the L dimensional vector defined by the indicator variables of all strata in the forested area, i.e. $\boldsymbol{Z}^{(2)}(x) = (Z_1(x), Z_2(x), \ldots Z_L(x))^t$, $Z_k(x) = I_{F_k}(x)$ for $k = 1, 2 \ldots L$. Note that all the components of $\boldsymbol{Z}(x)$ are zero for $x \notin F$. The $(L+1, L+1)$ matrix $\boldsymbol{A} = \sum_{x \in s_1} \boldsymbol{Z}(x) \boldsymbol{Z}(x)^t$ is almost diagonal

$$\boldsymbol{A} = \begin{bmatrix} n_{2F} & n_{21} & n_{22} & n_{23} & ... & n_{2L} \\ n_{21} & n_{21} & 0 & 0 & ... & 0 \\ n_{22} & 0 & n_{22} & 0 & ... & 0 \\ ... & & ... & ... & ... & 0 \\ n_{2k} & ... & 0 & n_{2k} & ... & 0 \\ ... & ... & ... & ... & 0 & \\ n_{2L} & 0 & 0 & ... & 0 & n_{2L} \end{bmatrix}$$

but it is singular because the first column is the sum of the last L columns. We set $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots \beta_L)^t$. A particular solution of the singular normal equation is easily found to be $\hat{\beta}_0 = \hat{\bar{Y}}_F$ and $\hat{\beta}_k = \hat{\bar{Y}}_k - \hat{\bar{Y}}_F$ where $\hat{\bar{Y}}_k = \frac{1}{n_{2,k}} \sum_{x \in F_k} Y(x)$ (this corresponds to the solution of the standard one-way ANOVA). This leads to the intuitively obvious predictions $\hat{Y}(x) = 0$ for $x \in F_0$ and $\hat{Y}(x) = \hat{\bar{Y}}_k$ for $x \in F_k$. Note also that all empirical residuals are zero outside F.

The standard regression estimate (with respect to $\tilde{F}$) based on the large model alone is

easily found to be [19]

$$\hat{Y}_{reg} = \hat{Y}_{post,standard} = \sum_{k=1}^{L} \frac{n_{1k}}{n_1} \hat{\bar{Y}}_k = \frac{1}{n_2} \sum_{x \in s_2} g_2(x) Y(x)$$

Since this algebraic identity is true for an arbitrary density $Y(x)$ we have

[55] $$g_2(x) = \frac{n_2}{n_1} \frac{n_{1k}}{n_{2k}} \quad \text{for} \quad x \in F_k,\ k = 0, 1, 2, \ldots$$

Simple algebra yields then for [15] the result

[56] $$\hat{Y}_{greg} = \hat{Y}_{post,new} = (p_F - \hat{p}_F)\hat{\bar{Y}}_F + \sum_{k=1}^{L} \hat{p}_k \hat{\bar{Y}}_k$$

where we have set $p_F = \frac{\lambda(F)}{\lambda(\tilde{F})}$, $\hat{p}_F = \frac{n_{1F}}{n_1}$ and $\hat{p}_k = \frac{n_{1k}}{n_1}$. For $n_1, n_2 \to \infty$ this is obviously a consistent estimate of the density with respect to $\tilde{F}$, i.e. of $\bar{Y}_{\tilde{F}} = \frac{1}{\lambda(\tilde{F})} \sum_{i=1}^{N} Y_i$.
The standard post-stratified estimate, i.e. with respect to $\tilde{F}$ but without knowledge of the exact surface area $\lambda(F)$ and $\lambda(\tilde{F})$, is given by the second term only (see Mandallaz (2008), section 5.2.1):

[57] $$\hat{Y}_{post,standard} = \sum_{k=1}^{L} \hat{p}_k \hat{\bar{Y}}_k$$

The new estimate [56] is intuitively appealing: if the number of points falling in the forested area F is above or below its expected value, then the classical estimate is corrected accordingly. Tedious but simple calculations and the approximations $n_{2k} - 1 \approx n_{2k}$ lead to the following estimated variances

[58] $$\hat{\mathbb{V}}(\hat{Y}_{post,new}) = \frac{1}{n_1 n_2} \frac{p_F^2}{\hat{p}_F^2} (n_{2F} - 1)\hat{\sigma}_F^2 + (1 - \frac{n_2}{n_1}) \sum_{k=1}^{L} \hat{p}_k^2 \frac{\hat{\sigma}_k^2}{n_{2k}}$$

where $p_k = \frac{\lambda(F_k)}{\lambda(\tilde{F})}$, $\hat{p}_k = \frac{n_{1k}}{n_1}$, $\hat{\sigma}_F^2 = \frac{1}{n_{2F}-1}\sum_{x\in s_2\cap F}(Y(x) - \hat{\bar{Y}}_F)^2$.

The main advantage of the g-weight variance estimate is that the strata weights p_k are estimated from the large sample and the contribution of the strata to the variance is inversely proportional to the sample sizes n_{2k}. Result [58] is similar to previous findings given in (Mandallaz (2008), pp 84 and 107-108).

The asymptotic variance ($n_1,\ n_2 \to \infty$) is then

$$[59] \qquad \hat{\mathbb{V}}(\hat{Y}_{post,new}) = \frac{1}{n_1}p_F\sigma_F^2 + (1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{k=1}^{L} p_k\sigma_k^2$$

where $\sigma_F^2 = \frac{1}{\lambda(F)}\int_F (Y(x) - \bar{Y}_F)^2 dx$ is the overall variance within F and the $\sigma_k^2 = \frac{1}{\lambda(F_k)}\int_{F_k}(Y(x) - \bar{Y}_k)^2 dx$ (with the strata means $\bar{Y}_k = \frac{1}{\lambda(F_k)}$) are the variances within strata.

To calculate the variance of $\hat{Y}_{post,standard}$ we use the external model assumption and equation [5.5.1] in Mandallaz (2008) to get

$$\mathbb{V}(\hat{Y}_{post,standard}) = \frac{1}{n_1}\mathbb{V}_{x\in\tilde{F}}(Y(x)) + (1-\frac{n_2}{n_1})\frac{1}{n_2}\mathbb{V}_{x\in\tilde{F}}(R(x))$$

where $R(x) = 0$ for $x \notin F$ and $R(x) = Y(x) - \bar{Y}_k$ for $x \in F_k$.

Noting that $\bar{Y}_{\tilde{F}} = \bar{Y} = p_F\bar{Y}_F,\ \bar{R}_{\tilde{F}} = \bar{R}_F = 0$ and writing $(Y(x)-\bar{Y}) = (Y(x)-\bar{Y}_F+\bar{Y}_F - p_F\bar{Y}_F)$ we obtain after some algebra the asymptotic variance of the standard post-stratified estimate with respect to $\tilde{F}$ as:

$$[60] \quad \mathbb{V}(\hat{Y}_{post,standard}) = \frac{1}{n_1}(p_F\sigma_F^2 + p_F(1-p_F)\bar{Y}_F^2) + (1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{k=1}^{L} p_k\sigma_k^2 > \mathbb{V}(\hat{Y}_{post,new})$$

Hence, as expected, using the exhaustive information $I_F(x)$ reduces the variance.

In practice forest inventories occasionally present totals over the entire region $\tilde{F}$ based on post-stratification with the non forested area $F_0 = \tilde{F} \setminus F$ viewed as a further ordinary

stratum, that is $\lambda(\tilde{F})$ is known, but neither $\lambda(F_0)$, not the $\lambda(F_k)$ are known. We have therefore to consider the following estimates of totals and their asymptotic variances:

$$\begin{aligned}
\hat{T}_{post,standard} &= \lambda(\tilde{F})\hat{Y}_{post,standard} \\
\mathbb{V}(\hat{T}_{post,standard}) &= \lambda^2(\tilde{F})\Big(\frac{1}{n_1}(p_F\sigma_F^2 + p_F(1-p_F)\bar{Y}_F^2) + (1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{k=1}^{L} p_k\sigma_k^2\Big) \\
\hat{T}_{post,new} &= \lambda(\tilde{F})\hat{Y}_{post,new} \\
\mathbb{V}(\hat{T}_{post,new}) &= \lambda^2(\tilde{F})\Big(\frac{1}{n_1}p_F\sigma_F^2 + (1-\frac{n_2}{n_1})\frac{1}{n_2}\sum_{k=1}^{L} p_k\sigma_k^2\Big) \qquad [61]
\end{aligned}$$

In the present framework, because F is assumed to be known, we can also consider the conditional estimate based only on the n_{1F} and n_{2F} points falling into F, that is

$$\begin{aligned}
\hat{Y}_{post,cond} &= \sum_{k=L}^{L} \hat{\tilde{p}}_k \hat{\bar{Y}}_k \\
\mathbb{V}(\hat{Y}_{post,cond}) &= \frac{1}{n_{1F}}\sigma_F^2 + (1-\frac{n_{2F}}{n_{1F}})\frac{1}{n_{2F}}\sum_{k=1}^{L}\tilde{p}_k\sigma_k^2 \\
\hat{T}_{post,cond} &= \lambda(F)\hat{Y}_{post,cond} \\
\mathbb{V}(\hat{T}_{post,cond}) &= \lambda^2(F)\mathbb{V}(\hat{Y}_{post,cond}) \qquad [62]
\end{aligned}$$

where we have set $\hat{\tilde{p}}_k = \frac{n_{1k}}{n_{1F}}$ and $\tilde{p}_k = \frac{\lambda(F_k)}{\lambda(F)}$. Using the facts that $\tilde{p}_k = \frac{p_k}{p_F}$, $\lambda(F) = p_F\lambda(\tilde{F})$, $\mathbb{E}(n_{1F}) = n_1 p_F$, $\mathbb{E}(n_{2F}) = p_F n_2$), we see that in large samples $\hat{Y}_{post,cond}$ and $\hat{T}_{post,new}$ are equivalent and better than $\mathbb{V}(\hat{T}_{post,standard})$, in the sense that

$$\mathbb{V}(\hat{T}_{post,new}) \approx \mathbb{V}(\hat{T}_{post,cond}) < \mathbb{V}(\hat{T}_{post,standard})$$

This fact is surprisingly not widely known and many national inventories are using estimators related somehow to $\hat{T}_{post,standard}$. Usually the decision for $x \notin F$ is easy (implying necessarily $Y(x) = 0$), whereas the delineation of the forested area can be more problem-

atic (i.e. to decide for $x \in F$, even if $Y(x) = 0$ is still possible). With the technological advances in remote sensing it is only a matter of time until we can assume that F and $\lambda(F)$ are known with the same accuracy as $\tilde{F}$ and $\lambda(\tilde{F})$. From a pragmatic point of view and as far as the estimation of total is concerned a coarse delineation by polygons defining a set F containing the "true" forest should suffice (so that p_F is close to 1 and the extra variance term $p_F(1 - p_f)\bar{Y}_F^2$ is small).

We emphasize again the fact that the surface areas of the strata within the forest area need not be known. In any case, the conditional estimator $\hat{Y}_{post,cond}$ and the new estimator $\hat{T}_{post,new}$ should be preferred to the occasionally used practice with $\hat{T}_{post,standard}$, particularly if $\tilde{F}$ is much larger than F.

References

Gregoire, T. and Dyer, M. (1989). Model fitting under patterned heterogeineity of variance. *Forest Science.*, **35**:pp. 105–125.

Huber, P. J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistic*, **1**:pp. 221–233.

Koehl, M., Magnussen, S., and Marchetti, M. (2006). *Sampling Methods, Remote Sensing and GIS Multisource Forest Inventory*. Springer, Berlin Heidelberg.

Mandallaz, D. (2008). *Sampling Techniques for Forest Inventories*. Chapman and Hall, Boca Raton FL.

Mandallaz, D. (2012). Design-based properties of small-area estimators in forest inventory with two phase sampling. Technical report, ETH Zurich, Department of Environmental Systems Science, technical report, http://e-collection.ethb.ethz.ch.

Renssen, R. and Martinus, G. (2002). On the Use of Generalized Inverse Matrices in Sampling Theory. *Statistics Canada, Catalogue No.12-001*, **28**:pp. 209–212.

Särndal, C., Swenson, B., and Wretman, J. (2003). *Model Assisted Survey Sampling*. Springer Series in Statistics, New York.