Regression estimators in forest inventories with three-phase sampling and two multivariate components of auxiliary information

Daniel Mandallaz,[1]

ETH Zurich

Department of Environmental Systems Science

Chair of Land Use Engineering

[1]Tel. ++41(0)44 6323186 e-mail daniel.mandallaz@env.ethz.ch, CHN K74.1, CH 8092 Zurich, Switzerland

Abstract

We consider three-phase sampling schemes where one multivariate component of the auxiliary information is known in the very large sample of the so-called null phase ("almost exhaustive") and the second multivariate component is available only in the large sample of the first phase, whereas the second phase provides the terrestrial inventory data. We extend the generalized regression estimator with partially exhaustive information to this three-phase set-up, for both global and local estimation and derive its asymptotic design-based variance. We also consider further extensions to cluster and two-stage sampling.

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 (aerial photographs or LiDAr data). The small-area estimation problem is in this context of the utmost importance. This paper adapts the so-called generalized regression estimator proposed in Mandallaz et al. (2013) to the case where the first component of the auxiliary information is no longer exhaustive, but is provided by a very large sample, the null-phase. The second component is available on a sub-sample of the null-phase, the first-phase, and the terrestrial inventory is performed on a sub-sample of the first phase, the second-phase. This set-up is particularly useful for national or regional inventories for two reasons: (1) the first component may not be available exhaustively (2) even if it were it may be computationally prohibitive for some of its variables, particularly those based on sophisticated algorithms requiring single tree identification (as in Mandallaz et al. (2013)). The methodology and terminology rests upon the design-based Monte-Carlo approach to sampling theory for forest theory. The reader unfamiliar with this topic should first consult Mandallaz (2013a) for a short bibliographical review and a thorough but terse analysis of the regression estimator in a new setting, particularly useful in the small-area estimation context. The book Mandallaz (2008), chapters 4 and 5, is recommended for a first perusal, as well as the on-line technical reports Mandallaz (2012, 2013b) for the mathematical details.

2 Methodology

The **null phase** draws a very large sample s_0 of n_0 points $x_i \in s_0$ $(i = 1, 2 \dots n_0)$ 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 column vector $\boldsymbol{Z}^{(1)}(x) \in \Re^p$. We shall call this component **pseudo-exhaustive** by analogy with the terminology used in Mandallaz (2013a), which deals with the case $n_0 = \infty$ (i.e. $\boldsymbol{Z}^{(1)}(x)$ is in this case the **exhaustive** component). The **first phase** draws a large sample $s_1 \subset s_0$ of $n_1 << n_0$ points by simple random sampling in s_0. Note that the points $x \in s_1$ are also uniformly independently distributed in F. For each point in the first phase a further component $\boldsymbol{Z}^{(2)}(x) \in \Re^q$ of the auxiliary information is available and hence also the 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 **second phase** draws a small sample $s_2 \subset s_1$ of n_2 points from s_1 by simple random sampling and consists of the terrestrial inventory. Note that we have used the terms null, first and second phases instead of first, second and third phases simply to ensure compatibility of the terminology used in previous work, in which the null-phase was exhaustive.

The reason for introducing the null phase sample s_0 is that the component $\boldsymbol{Z}^{(1)}(x)$ can be very computing intensive to calculate exhaustively (see Mandallaz (2013a) for a case study with LiDAR data).

In the forested area F we consider a well defined population $\mathcal{P}$ of N trees with response variable Y_i, $i = 1, 2 \dots$, 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]
$$I_i(x) = \begin{cases} 1 \text{ if } i \in s_2(x) \\ 0 \text{ if } i \notin s_2(x) \end{cases}$$

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

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

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 $Y(x)$ and we have $\mathbb{E}_x(Y(x)) = \frac{1}{\lambda(F)} \int_F Y(x)dx = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i = \bar{Y}$, where $\mathbb{E}_x$ denotes the expectation with respect to a random point x uniformly distributed in F.

3 The models

We shall work with the following linear models (see Mandallaz (2013b) for more details)

1. The large model M

   [3] $Y(x) = \mathbf{Z}^t(x)\boldsymbol{\beta} + R(x) = \mathbf{Z}^{(1)t}(x)\boldsymbol{\beta}^{(1)} + \mathbf{Z}^{(2)t}(x)\boldsymbol{\beta}^{(2)} + R(x) =: \hat{Y}_{theo}(x) + R(x)$

   with $\boldsymbol{\beta}^t = (\boldsymbol{\beta}^{(1)t}, \boldsymbol{\beta}^{(2)t})$ and the theoretical predictions $\hat{Y}_{theo}(x) = \mathbf{Z}^t(x)\boldsymbol{\beta}$.

   The intercept term is contained in $\mathbf{Z}^{(1)}(x)$ or it is a linear combination of its components.

   The theoretical regression parameter $\boldsymbol{\beta}$ minimizes $\int_F (Y(x) - \mathbf{Z}^t(x)\boldsymbol{\beta})^2 dx$, it satisfies

the normal equation $\left(\int_F \boldsymbol{Z}(x)\boldsymbol{Z}^t(x)dx \right)\boldsymbol{\beta} = \int_F Y(x)\boldsymbol{Z}(x)dx$ and the orthogonality relationship $\int_F R(x)\boldsymbol{Z}(x)dx = \boldsymbol{0}$, in particular the zero mean residual property $\frac{1}{\lambda(F)} \int_F R(x)dx = 0$.

2. The reduced model M_1

[4] $$ Y(x) = \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\alpha} + R_1(x) =: \hat{Y}_{theo,1}(x) + R_1(x) $$

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 $\left(\int_F \boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)dx \right)\boldsymbol{\alpha} = \int_F Y(x)\boldsymbol{Z}^{(1)}(x)dx$ and the orthogonality relationship $\int_F R_1(x)\boldsymbol{Z}^{(1)}(x)dx = \boldsymbol{0}$, in particular the zero mean residual property $\frac{1}{\lambda(F)} \int_F R_1(x)dx = 0$. $\hat{Y}_{theo,1}(x) = \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\alpha}$ are the theoretical predictions.

Let us emphasize the fact that in this paper we consider only the properties of estimators in the **design-based** paradigm and that we do not assume the above models to be correct in the sense of **model-dependent** inference (see Mandallaz (2013a)).

4 The generalized regression estimator

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

$$ \hat{\boldsymbol{\alpha}}_k = \left(\frac{1}{n_k} \sum_{x \in s_k} \boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x) \right)^{-1} \frac{1}{n_k} \sum_{x \in s_k} Y(x)\boldsymbol{Z}^{(1)}(x) $$

[5]
$$:= (\boldsymbol{A}_k^{(1)})^{-1} \frac{1}{n_k} \sum_{x \in s_k} Y(x)\boldsymbol{Z}^{(1)}(x) =: (\boldsymbol{A}_k^{(1)})^{-1}\boldsymbol{U}_k^{(1)}, \quad k = 0, 1, 2 $$

Likewise for the large large model we set

$$\hat{\boldsymbol{\beta}}_k = \left(\frac{1}{n_k}\sum_{x\in s_k}\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\right)^{-1}\frac{1}{n_k}\sum_{x\in s_k}Y(x)\boldsymbol{Z}(x)$$

[6]
$$= \boldsymbol{A}_k^{-1}\frac{1}{n_k}\sum_{x\in s_k}Y(x)\boldsymbol{Z}(x) =: \boldsymbol{A}_k^{-1}\boldsymbol{U}_k, \quad k=0,1,2$$

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** introduced and discussed by Mandallaz (2013b,a) is defined as

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

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 $\mathbb{E}_{1,2}\hat{Y}_{greg} = \mathbb{E}_1\mathbb{E}_{2|1}\hat{Y}_{greg} \approx \bar{Y}$ (with $\mathbb{E}_{2|1}$ denoting the conditional expectation of the second phase given the first phase, i.e. simple random sampling without replacement in the set s_1), and $\mathbb{E}_1$ denoting 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.

Under the **external model assumption** the design-based variance is given by

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

5

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

[9]
$$\hat{Y}_{F,reg} = \frac{1}{n_1}\sum_{x \in s_1}\hat{Y}(x) + \frac{1}{n_2}\sum_{x \in s_2}(Y(x) - \hat{Y}(x))$$

whose theoretical variance is given by

[10]
$$\mathbb{V}(\hat{Y}_{F,reg}) = \frac{1}{n_1}\mathbb{V}_x(Y(x)) + (1 - \frac{n_2}{n_1})\frac{1}{n_2}\mathbb{V}_x(R(x))$$

Thus, by using the exhaustive information the variance of the observations in [10] is replaced by the variance of the residuals under the reduced model, a very nice and intuitive result indeed.

The generalized regression estimator can be rewritten as

$$
\begin{aligned}
\hat{Y}_{F,greg} &= (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t\hat{\boldsymbol{\alpha}}_2 + (\hat{\bar{\boldsymbol{Z}}}_1 - \hat{\bar{\boldsymbol{Z}}}_2)^t\hat{\boldsymbol{\beta}}_2 + \frac{1}{n_2}\sum_{x \in s_2}Y(x) \\
\text{[11]} \qquad &= (\bar{\boldsymbol{Z}}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t\hat{\boldsymbol{\alpha}}_2 + \hat{\bar{\boldsymbol{Z}}}_1^t\hat{\boldsymbol{\beta}}_2
\end{aligned}
$$

Consistent estimates of the design-based variance of the regression coefficients are given by

[12]
$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2} = \boldsymbol{A}_2^{-1}\Big(\frac{1}{n_2^2}\sum_{x \in s_2}\hat{R}^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\Big)\boldsymbol{A}_2^{-1}$$

and

[13]
$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} = (\boldsymbol{A}_1^{(1)})^{-1}\Big(\frac{1}{n_2^2}\sum_{x \in s_2}\hat{R}_1^2(x)\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)\Big)(\boldsymbol{A}_1^{(1)})^{-1}$$

6

with the empirical residuals $\hat{R}(x) = Y(x) - \boldsymbol{Z}^t(x)\hat{\boldsymbol{\beta}}_2$ and $\hat{R}_1(x) = Y(x) - \boldsymbol{Z}^{1)t}(x)\hat{\boldsymbol{\alpha}}_2$ (see section A in the Appendix for a short review and Mandallaz (2008),pp 124-125 for details). Furthermore

[14]
$$\hat{\mathbb{V}}(\hat{Y}_{F,greg}) = \frac{n_2}{n_1}\bar{\boldsymbol{Z}}^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2}\bar{\boldsymbol{Z}}^{(1)} + (1 - \frac{n_2}{n_1})\hat{\bar{\boldsymbol{Z}}}_1^t\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2}\hat{\bar{\boldsymbol{Z}}}_1$$

is a consistent estimate of the **design-based variance** of $\hat{Y}_{F,greg}$ (see Mandallaz (2013b) for the proof).

The g-weights are defined as

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

They enjoy the calibration properties

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

The variance [14] can be rewritten as

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

Under the external model assumptions it is straightforward to obtain approximate estimates of the design-based variances of $\hat{Y}_{F,reg}$ and $\hat{Y}_{F,greg}$ by using the standard variance estimates in [8] and [10] after replacing the theoretical residuals by the empirical ones. The generalized regression estimator $\hat{Y}_{G,greg}$ for any small-area $G \subset F$ has been discussed in Mandallaz (2013b); Mandallaz et al. (2013). $\hat{Y}_{F,greg}$ and $\hat{Y}_{G,greg}$ can be viewed as limit

cases of the estimators presented in the next two sections.

5 The generalized regression estimator in three-phase sampling

If $\boldsymbol{Z}^{(1)}(x)$ is no longer exhaustive we replace the first term in [7] by its sample mean in the null phase sample and we define the new three-phase estimator as

$$[18] \qquad \hat{Y}_{F,g3reg} = \frac{1}{n_0} \sum_{x \in s_0} \hat{Y}_1(x) + \frac{1}{n_1} \sum_{x \in s_1} (\hat{Y}(x) - \hat{Y}_1(x)) + \frac{1}{n_2} \sum_{x \in s_2} (Y(x) - \hat{Y}(x))$$

By using the properties of conditional expectations and variances (see Mandallaz (2008), appendix B) we have

$$[19] \qquad \mathbb{E}_{0,1,2}(\hat{Y}_{F,g3reg}) = \mathbb{E}_0 \mathbb{E}_{1|0} \mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg})$$

and

$$\mathbb{V}_{0,1,2}(\hat{Y}_{F,g3reg}) = \mathbb{E}_0 \mathbb{E}_{1|0} \mathbb{V}_{2|0,1}(\hat{Y}_{F,g3reg}) + \mathbb{E}_0 \mathbb{V}_{1|0} \mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) + \mathbb{V}_0 \mathbb{E}_{1|0} \mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg})$$

$$[20] \qquad =: T_1 + T_2 + T_3$$

Under the external model assumption it follows from [20] after some algebra that $\hat{Y}_{F,g3reg}$ is exactly unbiased with variance

$$[21] \qquad \mathbb{V}_{0,1,2}(\hat{Y}_{F,g3reg}) = \frac{1}{n_0} \mathbb{V}_x(Y(x)) + (1 - \frac{n_1}{n_0}) \frac{1}{n_1} \mathbb{V}_x(R_1(x)) + (1 - \frac{n_2}{n_1}) \frac{1}{n_2} \mathbb{V}(R(x))$$

If the predictions $\hat{Y}_1(x)$ and the residuals $R_1(x)$ are orthogonal in the design-based sense (this is the case for their theoretical versions) then, because $\mathbb{V}_x(Y(x)) = \mathbb{V}_x(\hat{Y}_1(x)) +$

$\mathbb{V}_x(R_1(x))$, one obtains

[22] $$\mathbb{V}_{0,1,2}(\hat{Y}_{F,g3reg}) = \frac{1}{n_0}\mathbb{V}_x(\hat{Y}_1(x)) + \frac{1}{n_1}\mathbb{V}_x(R_1(x)) + (1 - \frac{n_2}{n_1})\frac{1}{n_2}\mathbb{V}(R(x))$$

Let us now define

[23] $$\hat{\bar{Z}}_0^{(1)} = \frac{1}{n_0}\sum_{x\in s_0}Z^{(1)}(x)dx, \quad \hat{\bar{Z}}_1^{(1)} = \frac{1}{n_1}\sum_{x\in s_1}Z^{(1)}(x), \quad \hat{\bar{Z}}_k = \frac{1}{n_k}\sum_{x\in s_k}Z(x), \ k = 1,2$$

and rewrite $\hat{Y}_{F,g3reg}$ as

$$\hat{Y}_{F,g3reg} = (\hat{\bar{Z}}_0^{(1)} - \hat{\bar{Z}}_1^{(1)})^t\hat{\alpha}_2 + (\hat{\bar{Z}}_1 - \hat{\bar{Z}}_2)^t\hat{\beta}_2 + \frac{1}{n_2}\sum_{x\in s_2}Y(x)$$

[24]
$$= (\hat{\bar{Z}}_0^{(1)} - \hat{\bar{Z}}_1^{(1)})^t\hat{\alpha}_2 + \hat{\bar{Z}}_1^t\hat{\beta}_2$$

The last equation follows from the fact that the sum of the residuals is zero by construction.

We now calculate the variance of [24] by using [12,13,20]. We begin with the easiest part, namely the last term T_3 in [20]. First, we obtain the asymptotic equivalence

[25] $$\mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) \approx (\hat{\bar{Z}}_0^{(1)} - \hat{\bar{Z}}_1^{(1)})^t\hat{\alpha}_1 + \hat{\bar{Z}}_1^t\hat{\beta}_1 \approx \hat{\bar{Z}}_0^{(1)t}\hat{\alpha}_1$$

The last equation results from the fact that both $\hat{\bar{Z}}_1^{(1)t}\hat{\alpha}_1$ and $\hat{\bar{Z}}_1^t\hat{\beta}_1$ tend to $\bar{Y}$.

We then get $\mathbb{E}_{1|0}\mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) \approx \hat{\bar{Z}}_0^{(1)t}\hat{\alpha}_0$. From [61] in the Appendix (section A) we get the asymptotically consistent estimate of $T_3 \approx \mathbb{V}_0(\hat{\bar{Z}}_0^{(1)t}\hat{\alpha}_0)$

[26] $$\hat{T}_3 = \frac{n_2}{n_0}\hat{\bar{Z}}_0^{(1)t}\hat{\Sigma}_{\hat{\alpha}_2}\hat{\bar{Z}}_0^{(1)} + \hat{\alpha}_2^t\hat{\Sigma}_{\hat{\bar{Z}}_0^{(1)}}\hat{\alpha}_2$$

9

where

$$[27] \qquad \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}} = \frac{1}{n_0} \frac{\sum_{x \in s_0} (\boldsymbol{Z}^{(1)}(x) - \hat{\bar{\boldsymbol{Z}}}_0^{(1)})(\boldsymbol{Z}(x) - \hat{\bar{\boldsymbol{Z}}}_0^{(1)})^t}{n_0 - 1}$$

The calculations for T_2 and T_1 are more complicated and rely on Taylor expansions for the regression coefficients. Details are given in section A of the Appendix.

According to [64] (section A of the Appendix) we have the following asymptotically consistent estimate of T_2

$$[28] \qquad \hat{T}_2 = \frac{n_2}{n_1}(1 - \frac{n_1}{n_0}) \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} \hat{\bar{\boldsymbol{Z}}}_0^{(1)}$$

and according to [69] the following asymptotically consistent estimate of T_1

$$[29] \qquad \hat{T}_1 \approx (1 - \frac{n_2}{n_1}) \hat{\bar{\boldsymbol{Z}}}_1^{t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2} \hat{\bar{\boldsymbol{Z}}}_1^{t}$$

and consequently the following asymptotically consistent estimate of the overall three-phase variance ([70], section A of the Appendix)

$$[30] \qquad \hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{F,g3reg}) = \hat{\boldsymbol{\alpha}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}} \hat{\boldsymbol{\alpha}}_2 + \frac{n_2}{n_1} \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} \hat{\bar{\boldsymbol{Z}}}_0^{(1)} + (1 - \frac{n_2}{n_1}) \hat{\bar{\boldsymbol{Z}}}_1^{t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2} \hat{\bar{\boldsymbol{Z}}}_1^{t}$$

By comparing with [14] we see that in large samples we have

$$[31] \qquad \hat{\mathbb{V}}(\hat{Y}_{F,g3reg}) - \hat{\mathbb{V}}(\hat{Y}_{F,greg}) = \hat{\boldsymbol{\alpha}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}} \hat{\boldsymbol{\alpha}}_2 = \frac{1}{n_0} \frac{\sum_{x \in s_0} (\hat{Y}_1(x) - \hat{\bar{Y}}_1)^2}{n_0 - 1}$$

where we have set $\hat{Y}_1(x) = \boldsymbol{Z}^{(1)t}(x) \hat{\boldsymbol{\alpha}}_2$ and $\hat{\bar{Y}}_1 = \frac{1}{n_0} \sum_{x \in s_0} \hat{Y}_1(x)$. Hence, the extra term in the variance due to three-phase sampling is given by the variance of the predictions of the reduced model in the very large null-phase sample, which can be made very small while

cutting down significantly the computing time. The g-weights are defined as

$$g(x) = \hat{\bar{\boldsymbol{Z}}}_1^t \boldsymbol{A}_2^{-1} \boldsymbol{Z}(x)$$

[32]
$$g^{(1)}(x) = \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} (\boldsymbol{A}_1^{(1)})^{-1} \boldsymbol{Z}^{(1)}(x)$$

They satisfy the calibration properties

$$\frac{1}{n_2} \sum_{x \in s_2} g(x) \boldsymbol{Z}(x) = \hat{\bar{\boldsymbol{Z}}}_1$$

[33]
$$\frac{1}{n_1} \sum_{x \in s_1} g^{(1)}(x) \boldsymbol{Z}^{(1)}(x) = \hat{\bar{\boldsymbol{Z}}}_0^{(1)}$$

Note the slights differences between [32], [33] and [15],[16].

It can be shown (see Mandallaz (2012, 2013b)) that $g^{(1)}(x)$ and $g(x)$ tend to 1 for $n_1, n_2 \to \infty$ and that

$$\frac{1}{n_1} \sum_{x_{s_1}} g^{(1)}(x) = 1$$

$$\frac{1}{n_2} \sum_{x \in s_2} g(x) = 1$$

One can rewrite [30] as

[34]
$$\hat{\mathbb{V}}(\hat{Y}_{F,g3reg}) = \hat{\boldsymbol{\alpha}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}} \hat{\boldsymbol{\alpha}}_2$$
$$+ \frac{1}{n_1} \frac{1}{n_2} \sum_{x \in s_2} (g^{(1)}(x))^2 \hat{R}_1^2(x) + \frac{1}{n_2} (1 - \frac{n_2}{n_1}) \frac{1}{n_2} \sum_{x \in s_2} g^2(x) \hat{R}^2(x)$$

11

Under the external model assumptions one obtains from [22] the following variance estimate

$$
\begin{aligned}
\hat{\mathbb{V}}_{ext}(\hat{Y}_{F,g3reg}) \;=\; & \frac{1}{n_0}\frac{\sum_{x\in s_0}(\hat{Y}_1(x)-\hat{\bar{Y}}_1)^2}{n_0-1} \\
[35]\qquad +\; & \frac{1}{n_1}\frac{1}{n_2-1}\sum_{x\in s_2}(\hat{R}_1(x)-\hat{\bar{R}}_1)^2 + (1-\frac{n_2}{n_1})\frac{1}{n_2(n_2-1)}\sum_{x\in s_2}(\hat{R}(x)-\hat{\bar{R}})^2
\end{aligned}
$$

where $\hat{\bar{Y}}_1 = \frac{1}{n_0}\sum_{x\in s_0}\hat{Y}_1(x)$, $\hat{\bar{R}}_1 = \frac{1}{n_2}\sum_{x\in s_2}\hat{R}_1(x) = 0$ and $\hat{\bar{R}} = \frac{1}{n_2}\sum_{x\in s_2}\hat{R}(x) = 0$. Since the g-weights $g^{(1)}(x)$ and $g(x)$ tend to 1 for $n_1, n_2 \to \infty$ we see that both variance estimates [34] and [35] are asymptotically equivalent. However, the g-weight version offers several statistical advantages from a theoretical point of view.

6   Application to small-area estimation

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:
$s_{0,G} = s_o \cap G$, $s_{1,G} = s_1 \cap G$, $s_{2,G} = s_2 \cap G$, $n_{k,G} = \sum_{s\in s_k} I_G(x)$, $k = 0,1,2$.
The simplest procedure is to restrict the samples to G, i.e. to consider **the generalized small-area estimator**

$$
[36]\quad \hat{Y}_{G,g3reg} = \frac{1}{n_{0,G}}\sum_{x\in s_{0,G}}\hat{Y}_1(x) + \frac{1}{n_{1,G}}\sum_{x\in s_{1,G}}(\hat{Y}(x)-\hat{Y}_1(x)) + \frac{1}{n_{2,G}}\sum_{x\in s_{2,G}}(Y(x)-\hat{Y}(x))
$$

12

and treat the internal model as an external one to obtain from [21] the variance estimate

$$
\begin{aligned}
\hat{\mathbb{V}}_{ext1}(\hat{Y}_{G,g3reg}) \;=\; & \frac{1}{n_{0,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (Y(x) - \hat{\bar{Y}}_G)^2 \\
& +\; (1 - \frac{n_{1,G}}{n_{0,G}})\frac{1}{n_{1,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}_1(x) - \hat{\bar{R}}_{1,G})^2 \\
& +\; (1 - \frac{n_{2,G}}{n_{1,G}})\frac{1}{n_{2,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}(x) - \hat{\bar{R}}_G)^2
\end{aligned}
$$

[37]

where $\hat{\bar{Y}}_G = \frac{1}{n_{0,G}}\sum_{x\in s_{0,G}} Y(x)$, $\hat{\bar{R}}_{1,G} = \frac{1}{n_{2,G}}\sum_{x\in s_{2,G}} \hat{R}_1(x)$ and $\hat{\bar{R}}_G = \frac{1}{n_{2,G}}\sum_{x\in s_{2,G}} \hat{R}(x)$. This variance estimate neglects the uncertainty of the regression coefficients but there is empirical evidence that this is acceptable in large samples (see Mandallaz (2013a) for examples with $\hat{Y}_{greg}$).

If the empirical covariance between predictions and residuals can be assumed to be zero over G (it is over F) then one could use [22] instead to obtain

$$
\begin{aligned}
\hat{\mathbb{V}}_{ext2}(\hat{Y}_{G,g3reg}) \;=\; & \frac{1}{n_{0,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2 \\
& +\; \frac{1}{n_{1,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}_1(x) - \hat{\bar{R}}_{1,G})^2 \\
& +\; (1 - \frac{n_{2,G}}{n_{1,G}})\frac{1}{n_{2,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}(x) - \hat{\bar{R}}_G)^2
\end{aligned}
$$

[38]

In the first term $\frac{1}{(n_{2,G}-1)}\sum_{x\in s_{2,G}}(\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2$ is the empirical variance of the predictions in $s_{2,G}$ and can be replaced by the variance of the prediction in $s_{0,G}$, i.e. by $\frac{1}{n_{0,G}-1}\sum_{x\in s_{0,G}}(\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2$ (with $\hat{\bar{Y}}_{1,G} = \frac{1}{n_{0,G}}\sum_{x\in s_{0,G}} \hat{Y}_1(x)$) to obtain the following

variance estimate under the external model assumption.

$$
\begin{aligned}
\hat{\mathbb{V}}_{ext}(\hat{Y}_{G,g3reg}) \;=\; & \frac{1}{n_{0,G}(n_{0,G}-1)} \sum_{x\in s_{0,G}} (\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2 \\
+ \; & \frac{1}{n_{1,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}_1(x) - \hat{\bar{R}}_{1,G})^2 \\
[39] \qquad\qquad + \; & (1 - \frac{n_{2,G}}{n_{1,G}}) \frac{1}{n_{2,G}(n_{2,G}-1)} \sum_{x\in s_{2,G}} (\hat{R}(x) - \hat{\bar{R}}_G)^2
\end{aligned}
$$

with

$$
\begin{aligned}
\hat{\bar{Y}}_{1,G} \;&=\; \frac{1}{n_{0,G}} \sum_{x\in s_{0,G}} \hat{Y}_1(x) \\
\hat{\bar{R}}_{1,G} \;&=\; \frac{1}{n_{2,G}} \sum_{x\in s_{2,G}} \hat{R}_1(x) \\
\hat{\bar{R}}_G \;&=\; \frac{1}{n_{2,G}} \sum_{x\in s_{2,G}} \hat{R}(x)
\end{aligned}
$$

Note that the first term in [39] can be expected to be more stable than the first terms in [37] and [38]. Also, as we shall see, $\hat{\mathbb{V}}_{ext3}(\hat{Y}_{G,0,greg})$ is asymptotically equivalent to a g-weight version of the variance, and should therefore be preferred to [37] and [38].

The main difficulty to obtain better variance estimate is that the residual third term in [36] is no longer zero in general. The most elegant way to bypass this difficulty is by extending the model with the indicator variable $I_G(x)$ of the small area G. Since the procedure has been discussed in details in Mandallaz (2012, 2013b,a); Mandallaz et al. (2013) we can here briefly sketch the procedure.

We have the 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 $\boldsymbol{\mathcal{Z}}^{(2)t}(x) = \boldsymbol{Z}^{(2)t}(x)$.

1. The large extended model M

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

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

2. The reduced extended model M_1

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

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

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

[40]

and

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

The estimated covariance matrices are

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

[42]

15

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 $\hat{\hat{Y}}_{G,g3reg}$ as in [11]

[43]
$$\hat{\hat{Y}}_{G,g3reg} = (\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G} - \hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{1,G})^t\hat{\boldsymbol{\gamma}}_2 + \hat{\bar{\boldsymbol{\mathcal{Z}}}}^t_{1,G}\hat{\boldsymbol{\theta}}_2$$

where we have set

$$\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G} = \frac{1}{n_{0,G}}\sum_{x \in s_{0,G}} \boldsymbol{\mathcal{Z}}^{(1)}(x)$$

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

$$\hat{\bar{\boldsymbol{\mathcal{Z}}}}_{1,G} = \frac{1}{n_{1,G}}\sum_{x \in s_{1,G}} \boldsymbol{\mathcal{Z}}(x)$$

To get an estimate of the design-based variance we use mutatis mutandis [30] and we obtain

[44]
$$\begin{aligned}\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{G,g3reg}) &= \hat{\boldsymbol{\gamma}}_2^t\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G}}\hat{\boldsymbol{\gamma}}_2 + \frac{n_2}{n_1}\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)t}_{0,G}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_2}\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G}\\ &+ (1 - \frac{n_2}{n_1})\hat{\bar{\boldsymbol{\mathcal{Z}}}}^t_{1,G}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_2}\hat{\bar{\boldsymbol{\mathcal{Z}}}}^t_{1,G}\end{aligned}$$

where

[45]
$$\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G}} = \frac{1}{n_{0,G}}\frac{1}{n_{0,G}-1}\sum_{x \in s_{0,G}} (\boldsymbol{\mathcal{Z}}^{(1)}(x) - \hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G})(\boldsymbol{\mathcal{Z}}(x) - \hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G})^t$$

is the empirical covariance matrix of $\hat{\bar{\boldsymbol{\mathcal{Z}}}}^{(1)}_{0,G}$ over G.

The first term in [44] can be rewritten as the empirical variance of the predictions $\hat{\hat{Y}}_1(x) =$

16

$\boldsymbol{\mathcal{Z}}^{(1)t}\hat{\boldsymbol{\gamma}}_2$, that is

$$\hat{\boldsymbol{\gamma}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mathcal{Z}}}_{0,G}^{(1)}} \hat{\boldsymbol{\gamma}}_2 = \frac{1}{n_{0,G}-1} \frac{1}{n_{0,G}} \sum_{x \in s_{0,G}} (\hat{\tilde{Y}}_1(x) - \hat{\bar{\tilde{Y}}}_{1,G})^2$$

with $\hat{\bar{\tilde{Y}}}_{1,G} = \frac{1}{n_{0,G}} \sum_{x \in s_{0,G}} \hat{\tilde{Y}}_1(x)$.

It is shown in section B that the external variance estimate [39] and the g-weight variance estimate [44] are asymptotically equivalent, which is not obvious at first glance.

For a very small area H the number of points $n_{2,H}$ may be to small or even zero. In such a case one could imbed H in a somewhat larger area $G \supset H$ with $n_{2,G}$ sufficiently large and consider the synthetic estimator in the extended model with respect to G: i.e. one replaces $\hat{\bar{\boldsymbol{\mathcal{Z}}}}_{0,G}^{(1)}$ by $\hat{\bar{\boldsymbol{\mathcal{Z}}}}_{0,H}^{(1)}$, $\hat{\bar{\boldsymbol{\mathcal{Z}}}}_{1,G}^t$ by $\hat{\bar{\boldsymbol{\mathcal{Z}}}}_{1,H}^t$ in equations [43], [44] and $\hat{\boldsymbol{\gamma}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mathcal{Z}}}_{0,G}^{(1)}} \hat{\boldsymbol{\gamma}}_2$ by $\hat{\boldsymbol{\gamma}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mathcal{Z}}}_{0,H}^{(1)}} \hat{\boldsymbol{\gamma}}_2 = \frac{1}{n_{0,H}-1} \frac{1}{n_{0,H}} \sum_{x \in s_{0,H}} (\hat{\tilde{Y}}_1(x) - \hat{\bar{\tilde{Y}}}_{1,H})^2$ with $\hat{\bar{\tilde{Y}}}_{1,H} = \frac{1}{n_{0,H}} \sum_{x \in s_{0,H}} \hat{\tilde{Y}}_1(x)$.

The variance $\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{H,g3reg})$ will be small but at the cost of a potential bias, which can be expected to be smaller than the bias of the synthetic estimator with respect to $G = F$. 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.

7 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 $e_1, \ldots e_M$ 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

null phase we have n_0 clusters identified by $x \in s_0$, in the first phase 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 in s_0, or s_1 respectively. As the estimators $\hat{Y}_{reg}, \hat{Y}_{F,greg}$ and $\hat{\tilde{Y}}_{G,greg}$ have been generalized to cluster sampling in Mandallaz (2013b,a); Mandallaz et al. (2013) we will only sketch the main features.

The design-unbiased estimate $\hat{\boldsymbol{\beta}}_{c,2}$ for $\boldsymbol{\beta}_c$ (the equivalent of $\boldsymbol{\beta}$ in simple random sampling) is

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

For the reduced model with $\boldsymbol{Z}^{(1)}(x)$ alone we obtain similarly

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

The residuals for the reduced model are $\hat{R}_{1,c}(x) = Y_c(x) - \hat{Y}_{1,c}$ with the predictions $\hat{Y}_{1,c} = \boldsymbol{Z}_c^{(1)t}(x) \hat{\boldsymbol{\alpha}}_{c,2}$.

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

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

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 [11]

[49]
$$\hat{Y}_{c,g3reg} = (\hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)} - \hat{\bar{\boldsymbol{Z}}}_{c,1}^{(1)})^t \hat{\boldsymbol{\alpha}}_{c,2} + \hat{\bar{\boldsymbol{Z}}}_{c,1}^t \hat{\boldsymbol{\beta}}_{c,2}$$

where

$$
\begin{aligned}
\hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)} &= \frac{\sum_{x \in s_0} M(x) \boldsymbol{Z}_c^{(1)}(x)}{\sum_{x \in s_0} M(x)} \\
\hat{\bar{\boldsymbol{Z}}}_{c,1}^{(1)} &= \frac{\sum_{x \in s_1} M(x) \boldsymbol{Z}_c^{(1)}(x)}{\sum_{x \in s_1} M(x)} \\
\hat{\bar{\boldsymbol{Z}}}_{c,1} &= \frac{\sum_{x \in s_1} M(x) \boldsymbol{Z}_c(x)}{\sum_{x \in s_1} M(x)}
\end{aligned}
$$

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

[50]
$$
\begin{aligned}
\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{c,g3reg}) &= \hat{\boldsymbol{\alpha}}_{c,2}^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)}} \hat{\boldsymbol{\alpha}}_{c,2} \\
&+ \frac{n_2}{n_1} \hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_{c,2}} \hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)} + (1 - \frac{n_2}{n_1}) \hat{\bar{\boldsymbol{Z}}}_{c,1}^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,2}} \hat{\bar{\boldsymbol{Z}}}_{c,1}
\end{aligned}
$$

where
$$\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)}} = \frac{1}{n_0(n_0 - 1)} \sum_{x \in s_0} \left(\frac{M(x)}{\bar{M}_0}\right)^2 (\boldsymbol{Z}_c^{(1)}(x) - \hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)})(\boldsymbol{Z}_c^{(1)}(x) - \hat{\bar{\boldsymbol{Z}}}_{c,0}^{(1)})^t$$

is the empirical covariance matrix of the $\boldsymbol{Z}_c^{(1)}(x)$ in s_0.

The first term in [50] can be rewritten as the empirical variance over s_0 of the predictions

of the reduced model, i.e.

$$\frac{1}{n_0(n_0-1)} \sum_{x \in s_0} \left(\frac{M(x)}{\bar{M}_0}\right)^2 (\hat{Y}_{1,c}(x) - \hat{\bar{Y}}_{1,0})^2$$

with $\hat{\bar{Y}}_{1,0} = \frac{\sum_{x \in s_0} M(x)\hat{Y}_{1,c}(x)}{\sum_{x \in s_0} M(x)}$ and $\bar{M}_0 = \frac{1}{n_0} \sum_{x \in s_0} M(x)$.

The external version of the variance estimate can be easily adapted from [35] after introducing the weights $\left(\frac{M(x)}{\bar{M}_k}\right)^2$.

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

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

with $\boldsymbol{Z}^{(1)t}(x) = (\boldsymbol{Z}^{(1)t}(x), I_G^t(x))$, $\boldsymbol{Z}^{(2)t}(x) = \boldsymbol{Z}^{(2)t}(x)$ and $I_G(x)$ is the indicator of the small area G. At the cluster level we have $\boldsymbol{Z}_c^{(1)t}(x) = (\boldsymbol{Z}_c^{(1)t}(x), I_{c,G}^t(x))$ 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

[51]
$$\hat{\boldsymbol{\gamma}}_{c,2} = (\boldsymbol{A}_{c,2}^{(1)})^{-1} \left(\frac{1}{n_2} \sum_{x \in s_2} M(x)\boldsymbol{Z}_c^{(1)}(x)Y(x)\right)$$

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

[52]
$$\hat{\boldsymbol{\theta}}_{c,2} = \boldsymbol{A}_{c,2}^{-1} \left(\frac{1}{n_2} \sum_{x \in s_2} M(x)\boldsymbol{Z}_c(x)Y(x)\right)$$

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

We define as in [49] the generalized small-area estimator by

$$[53] \qquad \hat{\tilde{Y}}_{c,G,g3reg} = (\hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,0,G}^{(1)} - \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G}^{(1)})^t \hat{\boldsymbol{\gamma}}_{c,2} + \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G}^t \hat{\boldsymbol{\theta}}_{c,2}$$

where we have set

$$
\begin{aligned}
\hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,0,G}^{(1)} &= \frac{\sum_{x \in s_{0,G}} M(x) \mathcal{Z}_c^{(1)}(x)}{\sum_{x \in s_{0,G}} M(x)} \\
\hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G}^{(1)} &= \frac{\sum_{x \in s_{1,G}} M(x) \boldsymbol{\mathcal{Z}}_c^{(1)}(x)}{\sum_{x \in s_{1,G}} M(x)} \\
\hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G} &= \frac{\sum_{x \in s_{1,G}} M(x) \boldsymbol{\mathcal{Z}}_c(x)}{\sum_{x \in s_{1,G}} M(x)}
\end{aligned}
$$

The estimated design-based covariance matrix are now

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

where

$$
\begin{aligned}
\hat{\mathcal{R}}_{1,c}(x) &= Y_c(x) - \hat{\tilde{Y}}_{1,c}(x) := Y_c(x) - \boldsymbol{\mathcal{Z}}_c^{(1)t}(x) \hat{\boldsymbol{\gamma}}_2 \\
\hat{\mathcal{R}}_c(x) &= Y_c(x) - \hat{\tilde{Y}}_c(x) := Y_c(x) - \boldsymbol{\mathcal{Z}}_c^t(x) \hat{\boldsymbol{\theta}}_2
\end{aligned}
$$

We obtain as in [50] the estimated design-based variance

$$
\begin{aligned}
\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{c,G,g3reg}) &= \hat{\boldsymbol{\gamma}}_{c,2}^t \hat{\boldsymbol{\Sigma}}_{\hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,0,G}^{(1)}} \hat{\boldsymbol{\gamma}}_{c,2}^t \\
[55] \qquad &+ \frac{n_2}{n_1} \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,0,G}^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\gamma}}_{c,2}} \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,0,G}^{(1)} + (1 - \frac{n_2}{n_1}) \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G}^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_{c,2}} \hat{\tilde{\boldsymbol{\mathcal{Z}}}}_{c,1,G}
\end{aligned}
$$

21

where

$$\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{z}}}_{c,0,G}^{(1)}} = \frac{1}{n_{0,G}(n_{0,G}-1)} \sum_{x \in s_{0,G}} \left(\frac{M(x)}{\bar{M}_0}\right)^2 (\boldsymbol{\mathcal{Z}}_c^{(1)}(x) - \hat{\bar{\boldsymbol{z}}}_{c,0,G}^{(1)})(\boldsymbol{\mathcal{Z}}_c^{(1)}(x) - \hat{\bar{\boldsymbol{z}}}_{c,0,G}^{(1)})^t$$

is the empirical covariance of $\hat{\bar{\boldsymbol{z}}}_{c,0,G}^{(1)}$ over G.

The first term in [55] can be rewritten as the variance of the predictions, that is

$$\hat{\boldsymbol{\gamma}}_{c,2}^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{z}}}_{c,0,G}^{(1)}} \hat{\boldsymbol{\gamma}}_{c,2}^t = \frac{1}{n_{0,G}(n_{0,G}-1)} \sum_{x \in s_{0,G}} \left(\frac{M(x)}{\bar{M}_0}\right)^2 (\hat{\bar{Y}}_{1,c}(x) - \bar{\bar{\hat{Y}}}_{1,c})^2$$

with $\bar{\bar{\hat{Y}}}_{1,c} = \frac{\sum_{x \in s_{0,G}} M(x) \hat{Y}_{1,c}(x)}{\sum_{x \in s_{0,G}} M(x)}$.

The external version of the variance estimate can be easily adapted from [39]: replace all occurring expressions $U(x)$ by the cluster means $U_c(x)$, calculate the mean values according to $\bar{U} = \frac{\sum_x M(x) U_{c}(x)}{\sum_x M(x)}$ and introduce the weights $\left(\frac{M(x)}{\bar{M}_k}\right)^2$.

8 Generalization to two-stage sampling at the plot level

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

[56]
$$J_i(x) = \begin{cases} 1 \text{ if } i \in s_3(x) \\ \\ 0 \text{ if } i \notin s_3(x) \end{cases}$$

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

$$
\begin{aligned}
Y^*(x) &= \frac{1}{\lambda(F)} \left(\sum_{i=1}^{N} \frac{I_i(x)Y_i^*}{\pi_i} + \sum_{i=1}^{N} \frac{I_i(x)J_i(x)R_i}{\pi_i p_i} \right) \\
[57] \quad &= \frac{1}{\lambda(F)} \left(\sum_{i \in s_2(x)} \frac{Y_i^*}{\pi_i} + \sum_{i \in s_3(x)} \frac{R_i}{\pi_i p_i} \right)
\end{aligned}
$$

where the p_i are the conditional inclusion probabilities for the the second stage sampling, i.e. $p_i = \mathbb{P}(J_i(x) = 1 \mid I_i(x) = 1)$.

One can use all the previous results by making the following changes:

- $Y(x) \to Y^*(x)$

- $\hat{\boldsymbol{\beta}} \to \hat{\boldsymbol{\beta}}_2^* = \boldsymbol{A}_2^{-1} \left(\frac{1}{n_2} \sum_{x \in s_2} Y^*(x)\boldsymbol{Z}(x) \right)$, $\hat{\boldsymbol{\alpha}}_2^* = (\boldsymbol{A}_2^{(1)})^{-1} \left(\frac{1}{n_2} \sum_{x \in s_2} Y^*(x)\boldsymbol{Z}^{(1)}(x) \right)$

- $\hat{R}_1(x) \to \hat{R}_1^*(x) = Y^*(x) - \boldsymbol{Z}^{(1)t}(x)\hat{\boldsymbol{\alpha}}_2^*$, $\hat{R}(x) \to \hat{R}^*(x) = Y^*(x) - \boldsymbol{Z}^t(x)\hat{\boldsymbol{\beta}}_2^*$

- Similarly in cluster sampling with $Y(x_l) \to Y^*(x_l)$, $Y_c(x) \to Y_c^*(x)$ etc.

One obtains for the resulting three-phase two-stage estimator $\hat{Y}_{F,g3reg}^*$ the theoretical vari-

ance

$$[58] \qquad \mathbb{V}_{0,1,2,3}(\hat{Y}_{F,g3reg}^*) = \mathbb{V}_{0,1,2,3}(\hat{Y}_{F,g3reg}) + \frac{1}{n_2}\mathbb{E}_{x\in F}V(x)$$

where $\mathbb{E}_{x\in F}V(x) = \frac{1}{\lambda^2(F)}\sum_{i=1}^N \frac{R_i^2(1-p_i)}{\pi_i p_i}$ is the second stage variance. The subscript 3 in $\mathbb{V}_{0,1,2,3}$ refers to the second-stage random selection at points $x \in s_2$.

The proof of [58] is straightforward under the external model assumption (including the small-area version), it suffices to use the techniques presented in (Mandallaz (2008), section 4.4 pp. 71-73, section 4.5 pp. 82-83), which also show that one can use the variance estimates [38] and [39] (valid also for $G = F$), after making the afore mentioned changes $Y(x) \to Y^*(x)$ etc. The second-stage variance is automatically taken into account. The same applies to cluster sampling (by using Mandallaz (2008), section 5.3 pp. 86-83). Because of the algebraic asymptotic equivalence established in section B of the Appendix this will also hold true for all the g-weights variance estimates.

References

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, http://e-collection.library.ethz.ch.

Mandallaz, D. (2013a). Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, **43**:pp. 441–449.

Mandallaz, D. (2013b). Regression estimators in forest inventories with two-phase

sampling and partially exhaustive information with application to small-area estimation. Technical report, ETH Zurich, Department of Environmental Systems Science, http://e-collection.library.ethz.ch.

Mandallaz, D., Breschan, J., and Hill, A. (2013). New regression estimators in forest inventory with two-phase sampling and partially exhaustive: a design-based Monte Carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, **43**:pp. 1023–1031.

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

A   Variance of $\hat{Y}_{F,g3reg}$

According to Mandallaz (2008) (pp 124-125) the asymptotic covariance matrices of the regression coefficients are given by

[59]
$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_k} = \boldsymbol{A}^{-1}\Big(\frac{1}{n_k}\mathbb{E}R^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\Big)\boldsymbol{A}^{-1}$$

[60]
$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_k} = (\boldsymbol{A}^{(1)})^{-1}\Big(\frac{1}{n_k}\mathbb{E}R_1^2(x)\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)\Big)(\boldsymbol{A}^{(1)})^{-1}$$

where $\boldsymbol{A} = \mathbb{E}_x\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)$ and $\boldsymbol{A}^{(1)} = \mathbb{E}_x\boldsymbol{Z}^{(1)}(x)\boldsymbol{Z}^{(1)t}(x)$, which can be consistently estimated by $\boldsymbol{A}_k$ ($k = 1, 2$) and $\boldsymbol{A}_k^{(1)}$ ($k = 0, 1, 2$). To ensure the calibration properties we will use $\boldsymbol{A}_2$ and $\boldsymbol{A}_1^{(1)}$.

We start with the calculation of $T_3 = \mathbb{V}_0(\mathbb{E}_{1|0}\mathbb{E}_{1,2|0}(\hat{Y}_{F,g3reg})) \approx \mathbb{V}_0(\hat{\bar{\boldsymbol{Z}}}_0^{(1t)}\hat{\boldsymbol{\alpha}}_0)$. Since $\mathbb{E}_0(\hat{\bar{\boldsymbol{Z}}}_0^{(1t)}\hat{\boldsymbol{\alpha}}_0) \approx \bar{\boldsymbol{Z}}^{(1t)}\boldsymbol{\alpha} = \bar{Y}$ we consider

$$\Delta = \hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_0 - \bar{\boldsymbol{Z}}^{(1)t}\boldsymbol{\alpha} = \hat{\bar{\boldsymbol{Z}}}_0^{(1)t}(\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}) + (\hat{\bar{\boldsymbol{Z}}}_0^{(1)t} - \bar{\boldsymbol{Z}}^{(1)t})\boldsymbol{\alpha}$$

Asymptotically we get

$$\mathbb{E}\Delta^2 = \bar{\boldsymbol{Z}}^{(1)t}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_0}\bar{\boldsymbol{Z}}^{(1)} + \boldsymbol{\alpha}^t\boldsymbol{\Sigma}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}}\boldsymbol{\alpha} + 2\mathbb{E}\Big((\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha})^t\hat{\bar{\boldsymbol{Z}}}_0^{(1)}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)} - \bar{\boldsymbol{Z}}^{(1)})^t\boldsymbol{\alpha}\Big)$$

where $\boldsymbol{\Sigma}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}}$ is the covariance matrix of $\hat{\bar{\boldsymbol{Z}}}_0^{(1)}$. The first two terms are of order n_0^{-1}. For the third term we note that $\mathbb{E}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)} - \bar{\boldsymbol{Z}}^{(1)})^t)$ is equal to the covariance matrix $\boldsymbol{\Sigma}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}}$ and therefore of order n_0^{-1} and that $\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}$ is of order $n_0^{-\frac{1}{2}}$. The last term is therefore of smaller order than the first two. Since $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_0} = \frac{n_2}{n_0}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}_2}$ we get the following asymptotically

consistent estimate

[61]
$$\hat{T}_3 = \frac{n_2}{n_0}\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2}\hat{\bar{\boldsymbol{Z}}}_0^{(1)} + \hat{\boldsymbol{\alpha}}_2^t\hat{\Sigma}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}}\hat{\boldsymbol{\alpha}}_2$$

We have to calculate

$$T_2 = \mathbb{E}_0\mathbb{V}_{1|0}\mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) \approx \mathbb{E}_0\mathbb{V}_{1|0}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1)$$

because, as we have seen, $\mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) \approx \hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1$. First, we note that

$$
\begin{aligned}
\mathbb{V}_{1|0}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1) &= \mathbb{E}_{1|0}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1 - \mathbb{E}_{1|0}\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1)^2 \\
&\approx \mathbb{E}_{1|0}(\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_1 - \hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\hat{\boldsymbol{\alpha}}_0)^2 \\
&= \hat{\bar{\boldsymbol{Z}}}_0^{(1)t}\mathbb{E}_{1|0}(\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0)(\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0)^t\hat{\bar{\boldsymbol{Z}}}_0^{(1)t}
\end{aligned}
$$

Using the Taylor expansion given in Mandallaz (2008) (p. 124, but at point $(\boldsymbol{A}_0^{(1)}, \hat{\boldsymbol{\alpha}}_0)$ instead of $(\boldsymbol{A}^{(1)}, \boldsymbol{\alpha})$) we get (with the notation defined in [5])

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0 &\approx -(\boldsymbol{A}_0^{(1)})^{-1}(\boldsymbol{A}_1^{(1)} - \boldsymbol{A}_0^{(1)})(\boldsymbol{A}_0^{(1)})^{-1}\boldsymbol{U}_0 + (\boldsymbol{A}_0^{(1)})^{-1}(\boldsymbol{U}_1 - \boldsymbol{U}_0) \\
&= (\boldsymbol{A}_0^{(1)})^{-1}(\boldsymbol{U}_1 - \boldsymbol{A}_1^{(1)}\hat{\boldsymbol{\alpha}}_0)
\end{aligned}
$$

Since $\hat{\boldsymbol{\alpha}}_0 = \boldsymbol{\alpha} + O(n_0^{-\frac{1}{2}})$ in probability this is asymptotically equivalent to

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0 &\approx (\boldsymbol{A}_0^{(1)})^{-1}\Big(\frac{1}{n_1}\sum_{x\in s_1}(Y(x) - \boldsymbol{Z}^{(1)t}(x)\boldsymbol{\alpha})\boldsymbol{Z}^{(1)}(x)\Big)
\end{aligned}
$$

[62]
$$= (\boldsymbol{A}_0^{(1)})^{-1}\Big(\frac{1}{n_1}\sum_{x\in s_1} R_1(x)\boldsymbol{Z}^{(1)}(x)\Big)$$

27

With the indicator variables $J(x) = 1$ if $x \in s_0$ and $J(x) = 0$ if $x \notin s_0$ we obtain

$$
\begin{aligned}
\mathbb{E}_{1|0}(\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0)(\hat{\boldsymbol{\alpha}}_1 - \hat{\boldsymbol{\alpha}}_0)^t \approx\ & (\boldsymbol{A}_0^{(1)})^{-1} \frac{1}{n_1^2} \mathbb{E}_{1|0} \sum_{x \in s_0} J(x) R_1^2(x) \boldsymbol{Z}^{(1)}(x) \boldsymbol{Z}^{(1)t}(x) \\
+\ & (\boldsymbol{A}_0^{(1)})^{-1} \frac{1}{n_1^2} \mathbb{E}_{1|0} \sum_{x \neq y \in s_0} J(x) J(y) R_1(x) \boldsymbol{Z}^{(1)}(x) R_1(y) \boldsymbol{Z}^{(1)t}(y)
\end{aligned}
$$

Using simple random sampling within the set s_0 we have

$$
\begin{aligned}
\mathbb{E}_{1|0} J(x) &= \frac{n_1}{n_0} \\
\mathbb{E}_{1|0} J(x) J(y) &= \frac{n_1(n_1 - 1)}{n_0(n_0 - 1)} \\
\sum_{x \neq y \in s_0} J(x) J(y) R_1(x) \boldsymbol{Z}^{(1)}(x) R_1(y) \boldsymbol{Z}^{(1)t}(y) &= \Big(\sum_{x \in s_0} J(x) R_1(x) \boldsymbol{Z}^{(1)}(x) \Big)^2 \\
&\quad - \sum_{x \in s_0} R_1^2(x) \boldsymbol{Z}^{(1)}(x) \boldsymbol{Z}^{(1)t}(x) \\
\mathbb{E}_0 \mathbb{E}_{1|0} \sum_{x \in s_0} J(x) R_1(x) \boldsymbol{Z}^{(1)}(x) &= 0 \quad \text{(orthogonality relationship)}
\end{aligned}
$$

With the approximation $\frac{n_1 - 1}{n_0 - 1} \approx \frac{n_1}{n_0}$ we obtain after tedious but simple algebra

$$[63] \qquad T_2 = (1 - \frac{n_1}{n_0}) \bar{\boldsymbol{Z}}^{(1)t} (\boldsymbol{A}^{(1)})^{-1} \Big(\frac{1}{n_1} \mathbb{E}_x (R_1^2(x) \boldsymbol{Z}^{(1)}(x) \boldsymbol{Z}^{(1)t}(x)) \Big) (\boldsymbol{A}^{(1)})^{-1} \bar{\boldsymbol{Z}}^{(1)}$$

and the 'plug-in' asymptotically consistent estimate

$$
\begin{aligned}
\hat{T}_2 &\approx (1 - \frac{n_1}{n_0}) \frac{n_2}{n_1} \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} \hat{\bar{\boldsymbol{Z}}}_0^{(1)} \\
[64] \qquad &\approx (1 - \frac{n_1}{n_0}) \frac{n_2}{n_1} \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} (\boldsymbol{A}_1^{(1)})^{-1} \Big(\frac{1}{n_2^2} \sum_{x \in s_2} \hat{R}_1^2(x) \boldsymbol{Z}^{(1)}(x) \boldsymbol{Z}^{(1)t}(x) \Big) (\boldsymbol{A}_1^{(1)})^{-1} \hat{\bar{\boldsymbol{Z}}}_0^{(1)}
\end{aligned}
$$

We now proceed to calculate the first term T_1. First we note that

$$[65] \quad \hat{Y}_{F,g3reg} - \mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg}) \approx (\hat{\bar{\boldsymbol{Z}}}_0^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t (\hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_1) + \hat{\bar{\boldsymbol{Z}}}_1^t (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1) \approx \hat{\bar{\boldsymbol{Z}}}_1^t (\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)$$

The last asymptotic equivalence in [65] holds because $(\hat{\bar{\boldsymbol{Z}}}_0^{(1)} - \hat{\bar{\boldsymbol{Z}}}_1^{(1)})^t(\hat{\boldsymbol{\alpha}}_2 - \hat{\boldsymbol{\alpha}}_1)$ is the product of two terms of order $O(n_2^{-\frac{1}{2}})$ in probability whereas the second is of order $O(n_2^{-\frac{1}{2}})$. One gets therefore

$$\mathbb{V}_{2|0,1}(\hat{Y}_{F,g3reg} - \mathbb{E}_{2|0,1}(\hat{Y}_{F,g3reg})) \approx \hat{\bar{\boldsymbol{Z}}}_1^t \mathbb{E}_{2|0,1}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1)^t \hat{\bar{\boldsymbol{Z}}}_1$$

Using mutatis mutandis [62] (i.e. Taylor expansion at $\hat{\boldsymbol{\beta}}_1$) we have

[66]
$$\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1 \approx \boldsymbol{A}_1^{-1}\Big(\frac{1}{n_2}\sum_{x \in s_2} R(x)\boldsymbol{Z}(x)\Big)$$

Using the same techniques as for T_2 (simple random sampling within the set s_1 instead of s_0) we have

[67]
$$T_1 = (1 - \frac{n_2}{n_1})\bar{\boldsymbol{Z}}^t \boldsymbol{A}^{-1}\Big(\frac{1}{n_2}\mathbb{E}_x(R^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x))\Big)\boldsymbol{A}^{-1}\bar{\boldsymbol{Z}}$$

and

[68]
$$T_1 \approx (1 - \frac{n_2}{n_1})\bar{\boldsymbol{Z}}^t \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_2}\bar{\boldsymbol{Z}}$$

which leads to the 'plug-in' asymptotically consistent estimate

[69]
$$\begin{aligned}\hat{T}_1 &\approx (1 - \frac{n_2}{n_1})\hat{\bar{\boldsymbol{Z}}}_1^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_2}\hat{\bar{\boldsymbol{Z}}}_1 \\ &= (1 - \frac{n_2}{n_1})\hat{\bar{\boldsymbol{Z}}}_1^t \boldsymbol{A}_2^{-1}\Big(\frac{1}{n_2^2}\sum_{x \in s_2} \hat{R}^2(x)\boldsymbol{Z}(x)\boldsymbol{Z}^t(x)\Big)\boldsymbol{A}_2^{-1}\hat{\bar{\boldsymbol{Z}}}_1\end{aligned}$$

Hence, one obtains the asymptotically consistent estimate of the variance

$$\hat{\mathbb{V}}(\hat{Y}_{F,g3reg}) \approx \hat{T}_3 + \hat{T}_2 + \hat{T}_1$$

29

Because two terms cancel out we finally obtain

$$[70] \qquad \hat{\mathbb{V}}(\hat{Y}_{F,g3reg}) \approx \hat{\boldsymbol{\alpha}}_2^t \hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0} \hat{\boldsymbol{\alpha}}_2 + \frac{n_2}{n_1} \hat{\bar{\boldsymbol{Z}}}_0^{(1)t} \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} \hat{\bar{\boldsymbol{Z}}}_0^{(1)} + (1 - \frac{n_2}{n_1}) \hat{\bar{\boldsymbol{Z}}}_1^t \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_2} \hat{\bar{\boldsymbol{Z}}}_1$$

Since the variance-covariance matrix of the $\boldsymbol{Z}^{(1)}(x)$ is given by

$$\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\boldsymbol{Z}}}_0^{(1)}} = \frac{1}{n_0} \frac{\sum_{x \in s_0} (\boldsymbol{Z}^{(1)}(x) - \hat{\bar{\boldsymbol{Z}}}_0^{(1)})(\boldsymbol{Z}(x) - \hat{\bar{\boldsymbol{Z}}}_0^{(1)})^t}{n_0 - 1}$$

it is easily checked that the first term in [70] is equal to the variance of the predictions under the reduced model, i.e.

$$\frac{1}{n_0} \frac{\sum_{x \in s_0} (\hat{Y}_1(x) - \hat{\bar{Y}}_1)^2}{n_0 - 1}$$

with $\hat{Y}_1(x) = \boldsymbol{Z}^{(1)t}(x) \hat{\boldsymbol{\alpha}}_2$ and $\hat{\bar{Y}}_1 = \frac{1}{n_0} \sum_{x \in s_0} \hat{Y}_1(x)$.

B   Asymptotic equivalence of variance estimates for small areas

We define the g-weights

$$\begin{aligned}
\tilde{g}(x) &= \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{1,G}^t \boldsymbol{A}_2^{-1} \boldsymbol{\mathcal{Z}}(x) \\
[71] \qquad \tilde{g}^{(1)}(x) &= \hat{\bar{\boldsymbol{\mathcal{Z}}}}_{0,G}^{(1)t} (\boldsymbol{A}_1^{(1)})^{-1} \boldsymbol{\mathcal{Z}}^{(1)}(x)
\end{aligned}$$

First, we note that [44] can be rewritten as

$$\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{G,g3reg}) = \frac{1}{n_{0,G}(n_{0,G}-1)} \sum_{x \in s_{0,G}} (\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2$$

$$+ \frac{1}{n_1 n_2} \sum_{x \in s_2} (\tilde{g}^{(1)}(x))^2 \hat{\mathcal{R}}_1^2(x) + (1 - \frac{n_2}{n_1}) \frac{1}{n_2^2} \sum_{x \in s_2} \tilde{g}^2(x) \hat{\mathcal{R}}^2(x)$$

Second, we consider

$$\hat{\bar{\mathbf{Z}}}_{0,G}^{(1)t} \hat{\boldsymbol{\gamma}}_1 = \frac{1}{n_1} \sum_{x \in s_1} \tilde{g}^{(1)}(x) Y(x) \approx \frac{1}{n_{1,G}} \sum_{x \in s_{1,G}} Y(x)$$

because all terms converge towards $\bar{Y}_G$ (the fact that they are not observable is irrelevant in this respect). Hence, we see that $\tilde{g}^{(1)}(x) \to 0$ for $x \notin G$ and $\tilde{g}^{(1)}(x) \to \frac{n_1}{n_{1,G}}$ for $x \in G$. Likewise we have

$$\hat{\bar{\mathbf{Z}}}_{1,G}^{t} \hat{\boldsymbol{\theta}}_2 = \frac{1}{n_2} \sum_{x \in s_2} \tilde{g}(x) Y(x) \approx \frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} Y(x)$$

so that $g(x) \to 0$ for $x \notin G$ and $g(x) \to \frac{n_2}{n_{2,G}}$ for $x \in G$. Note that the behavior of the g-weights for G and for F is completely different.

Since obviously $\frac{n_{2,G}}{n_{1,G}} \approx \frac{n_2}{n_1}$ we get

[72] $\qquad \hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{G,g3reg}) \approx \frac{1}{n_{0,G}(n_{0,G}-1)} \sum_{x \in s_{0,G}} (\hat{Y}_1(x) - \hat{\bar{Y}}_{1,G})^2$

[73] $\qquad\qquad + \frac{1}{n_{1,G} n_{2,G}} \sum_{x \in s_{2,G}} \hat{\mathcal{R}}_1^2(x) + (1 - \frac{n_{2,G}}{n_{1,G}}) \frac{1}{n_{2,G}^2} \sum_{x \in s_{2,G}} \hat{\mathcal{R}}^2(x)$

Finally, using the relation between $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\theta}}_2$ given in equation [42] of Mandallaz (2012), we see that $\hat{\mathcal{R}}(x) = \hat{R}(x) - \phi(x)\hat{\bar{R}}_G$ with $\frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} \phi(x) = 1$, and likewise $\hat{\mathcal{R}}_1(x) = \hat{R}_1(x) - \phi_1(x)\hat{\bar{R}}_{1,G}$ with $\frac{1}{n_{2,G}} \sum_{x \in s_{2,G}} \phi_1(x) = 1$. Asymptotically we can replace $\hat{\mathcal{R}}(x)$ by $\hat{R}(x) - \hat{\bar{R}}_G$ and $\hat{\mathcal{R}}_1(x)$ by $\hat{R}_1(x) - \hat{\bar{R}}_{1,G}$, which leads to $\hat{\mathbb{V}}_{0,1,2}(\hat{Y}_{G,g3reg}) \approx \hat{\mathbb{V}}_{ext3}(\hat{Y}_{G,g3reg})$ and the asymptotic equivalence of external and g-weights variance estimates.