

A Taxonomy of Estimators

Andreas Dominik Cullmann

November 2, 2020

1 Introduction

There is a multitude of estimators given in [1], [2], [3], [4], [5], [6] and, finally, [7].

The notation varies, for partially exhaustive auxiliary information, the classification given in [7] even deviates from canon (see 2).

So this is an effort to systematically describe the various small area estimators.

Superscripts For partially exhaustive auxiliary information, Mandallaz ([4, p. 1023], [6, p. 383f] defines $Z^t(x) = Z^{(1)t}(x) + Z^{(2)t}(x)$ whereas Hill [7, p. 4 and p. 18] defines $Z^t(x) = Z^{(0)t}(x) + Z^{(1)t}(x)$. I will stick with Mandallaz' notation, changing $Z^{(0)t}(x)$ to $Z^{(1)t}(x)$ in Hill's formulae!

Indices Mandallaz and Hill inconsistently use the indices $_2$ and $_{s_2}$, they really both denote the same: the set s_2 . For the sets s_0 and s_1 they consistently use $_0$ and $_1$. I change all set indices to $s_{[012]}$.

Hill uses $\bar{Z}_{0,G}^{(1)}$ (and $\bar{Z}_0^{(1)}$ which ([7, p. 18]) is the exact mean). So I do drop the index, which is misleadingly referring to some set (and I do so for $\bar{Z}_{0,G}^{(1)}$).

Mandallaz uses $\bar{R}_{2,G}$ when calculating the variance of the residuals in G, for example in [2, eq 26], where $\bar{R}_{2,G}$ is clearly $\bar{R}(x)$ while summing over s_2 and G . I use the latter form.

Residuals I have replaced the empirical mean and variance of the Residuals in G for clustered sampling,

$$\frac{\sum_{x \in s_2, G} M(x) \hat{R}_c(x)}{\sum_{x \in s_2, G} M(x)}$$

and

$$\frac{1}{n_{s_2, G} - 1} \sum_{x \in s_2, G} \left(\frac{M(x)}{\bar{M}(x)} \right)^2 (\hat{R}_c(x) - \bar{R}_c(x))^2,$$

by their shorter notations $\bar{R}_{c, s_2, G}(x)$ and $\hat{V}(\hat{R}_{c, s_2, G}(x))$ and likewise for non-clustered sampling.

2 Estimators

In tables 1 and 2, we see the estimators for the two- and three-phase non-clustered sampling designs. The estimators are grouped by the type of auxiliary information: exhaustive (for three-phase sampling with full exhaustive auxiliary information is just two-phase sampling with full exhaustive auxiliary information with more observations, so there are no estimators), non-exhaustive and partially exhaustive. In each block the (pseudo) synthetic the (pseudo) small and the (pseudo) extended estimator and their variances are given.

Looking at the estimators for partially exhaustive auxiliary information we see that the estimators and variances are identical for two- and three-phase sampling. This is due to the fact that [7] implement the partially exhaustive auxiliary information using a full and a reduced model. So they see it as three-phase sampling where [4] clearly see it as two-phase sampling with partially exhaustive auxiliary information.

Tables 3 and 4 give the same information for clustered sampling designs.

References

- [1] Daniel Mandallaz. Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Technical report, Eidgenössische Technische Hochschule Zürich, Departement Umweltsystemwissenschaften, 2012.
- [2] Daniel Mandallaz. Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research*, 43(5):441–449, 2013.
- [3] Daniel Mandallaz. Regression estimators in forest inventories with two-phase sampling and partially exhaustive information with applications to small-area estimation. Technical report, Eidgenössische Technische Hochschule Zürich, Departement Umweltsystemwissenschaften, 2013.
- [4] Daniel Mandallaz, Jochen Breschan, and Andreas Hill. New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 43(11):1023–1031, 2013.
- [5] Daniel Mandallaz. Regression estimators in forest inventories with three-phase sampling and two multivariate components of auxiliary information. Technical report, Eidgenössische Technische Hochschule Zürich, Departement Umweltsystemwissenschaften, 2013.
- [6] Daniel Mandallaz. A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. *Canadian Journal of Forest Research*, early(online):22, 2013.
- [7] Andreas Hill and Alexander Massey. The r package forestinventory: Design-based global and small area estimations for multi-phase forest inventories. Technical report, 2017. Vignette of R package ‘forestinventory’ version 0.3.1.

exh	Type	Reference	Formula
yes	<i>sy</i>	[2, eq. 18] [2, eq. 19]	$\hat{Y}_{G, synth} = \bar{Z}_G^t \hat{\beta}_{s_2}$ $\hat{V} = \bar{Z}_G^t \hat{\Sigma}_{\hat{\beta}_{s_2}} \bar{Z}_G$
yes	<i>sm</i>	[2, eq. 20] [2, eq. 21]	$\hat{Y}_{G, small} = \hat{Y}_{G, synth} + \bar{R}_{s_2, G}(x)$ $\hat{V} \approx \hat{V}(\hat{Y}_{G, synth}) + \frac{1}{n_{s_2, G}} \hat{V}(\bar{R}_{s_2, G}(x))$
yes	<i>ex</i>	[2, eq. 31] [2, eq. 33]	$\hat{Y}_{G, synth} = \bar{Z}_G^t \hat{\theta}_{s_2}$ $\hat{V} = \bar{Z}_G^t \hat{\Sigma}_{\hat{\theta}_{s_2}} \bar{Z}_G$
no	<i>sy</i>	[2, eq. 22] [2, eq. 23]	$\hat{Y}_{G, psynth} = \hat{Z}_{s_1, G}^t \hat{\beta}_{s_2}$ $\hat{V} = \hat{Z}_{s_1, G}^t \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{Z}_{s_1, G} + \hat{\beta}_{s_2}^t \hat{\Sigma}_{\hat{Z}_{s_1, G}} \hat{\beta}_{s_2}$
no	<i>sm</i>	[2, eq. 25] [2, eq. 26]	$\hat{Y}_{G, psmall} = \hat{Y}_{G, psynth} + \bar{R}_{s_2, G}(x)$ $\hat{V} \approx \hat{V}(\hat{Y}_{G, psynth}) + \frac{1}{n_{s_2, G}} \hat{V}(\bar{R}_{s_2, G}(x))$
no	<i>ex</i>	[2, eq. 35] [2, eq. 36]	$\hat{Y}_{G, psynth} = \hat{Z}_{s_1, G}^t \hat{\theta}_{s_2}$ $\hat{V} = \hat{Z}_{s_1, G}^t \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{Z}_{s_1, G} + \hat{\theta}_{s_2}^t \hat{\Sigma}_{\hat{Z}_{s_1, G}} \hat{\theta}_{s_2}$
part	<i>sy</i>	[4, eq. 34] [4, eq. 35]	$\hat{Y}_{psynth, G, greg} = \left(\bar{Z}_G^{(1)} - \hat{Z}_{s_1, G}^{(1)} \right) \hat{\alpha}_{s_2} + \hat{Z}_{s_1, G}^t \hat{\beta}_{s_2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \bar{Z}_G^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{s_2}} \bar{Z}_G^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}} \right) \hat{Z}_{s_1, G}^t \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{Z}_{s_1, G}$
part	<i>sm</i>	[4, eq. 24] [4, eq. 23]	$\hat{Y}_{G, greg} = \hat{Y}_{psynth, G, greg} + \bar{R}_{s_2, G}(x)$ $\hat{V} \approx \hat{V}(\hat{Y}_{psynth, G, greg}) + \frac{1}{n_{s_2, G}} \hat{V}(\bar{R}_{s_2, G}(x))$
part	<i>ex</i>	[4, eq. 30] [4, eq. 31]	$\hat{Y}_{G, greg} = \left(\bar{Z}_G^{(1)} - \hat{Z}_{s_1, G}^{(1)} \right) \hat{\gamma}_{s_2} + \hat{Z}_{s_1, G}^t \hat{\theta}_{s_2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \bar{Z}_G^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{s_2}} \bar{Z}_G^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}} \right) \hat{Z}_{s_1, G}^t \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{Z}_{s_1, G}$

Table 1: Predictors for non-clustered two-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *Type* denotes the area estimator (*sy* for synthetic, *sm* for small and *ex* for extended synthetic). *Reference* gives the reference where found, else I derived them by analogy.

exh	Type	Reference	Formula
yes	<i>sy</i>	—	—
yes	<i>sm</i>	—	—
yes	<i>ex</i>	—	—
no	<i>sy</i>	[7, eq. 26b]	$\hat{Y}_{G,psynth,3p} = \left(\hat{\bar{Z}}_{s_0,G}^{(1)} - \hat{\bar{Z}}_{s_1,G}^{(1)} \right) \hat{\alpha}_{s_2} + \hat{\bar{Z}}_{s_1,G}^t \hat{\beta}_{s_2}$
		[7, eq. 26d]	$\hat{V} = \hat{\alpha}_{s_2}^t \hat{\Sigma}_{\hat{\bar{Z}}_{s_0,G}^{(1)}} \hat{\alpha}_{s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{s_2}} \hat{\bar{Z}}_{s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{Z}}_{s_1,G}$
no	<i>sm</i>	[7, eq. 22b]	$\hat{Y}_{G,psmall,3p} = \hat{Y}_{G,psynth,G,3p} + \hat{R}_{s_2,G}(x)$
		[7, eq. 23b]	$\hat{V} \approx \hat{V} \left(\hat{Y}_{G,psynth,3p} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
no	<i>ex</i>	[6, eq. 23]	$\hat{Y}_{G,g3reg} = \left(\hat{\bar{Z}}_{s_0,G}^{(1)} - \hat{\bar{Z}}_{s_1,G}^{(1)} \right) \hat{\gamma}_{s_2} + \hat{\bar{Z}}_{s_1,G}^t \hat{\theta}_{s_2}$
		[6, eq. 24]	$\hat{V} = \hat{\gamma}_{s_2}^t \hat{\Sigma}_{\hat{\bar{Z}}_{s_0,G}^{(1)}} \hat{\gamma}_{s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{s_2}} \hat{\bar{Z}}_{s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{Z}}_{s_1,G}$
part	<i>sy</i>	[7, eq. 26a]	$\hat{Y}_{G,ynth,3p} = \left(\bar{Z}_G^{(1)} - \hat{\bar{Z}}_{s_1,G}^{(1)} \right) \hat{\alpha}_{s_2} + \hat{\bar{Z}}_{s_1,G}^t \hat{\beta}_{s_2}$
		[7, eq. 26c]	$\hat{V} = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{s_2}} \hat{\bar{Z}}_{s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{Z}}_{s_1,G}$
part	<i>sm</i>	[7, eq. 22a]	$\hat{Y}_{G,small,3p} = \hat{Y}_{G,ynth,3p} + \hat{R}_{s_2,G}(x)$
		[7, eq. 23a]	$\hat{V} \approx \hat{V} \left(\hat{Y}_{G,ynth,3p} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
part	<i>ex</i>	—	$\hat{Y}_{G,extsynth,3p} = \left(\bar{Z}_G^{(1)} - \hat{\bar{Z}}_{s_1,G}^{(1)} \right) \hat{\gamma}_{s_2} + \hat{\bar{Z}}_{s_1,G}^t \hat{\theta}_{s_2}$
		—	$\hat{V} = \frac{n_{s_2}}{n_{s_1}} \bar{Z}_{s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{s_2}} \bar{Z}_{s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{Z}}_{s_1,G}$

Table 2: Predictors for non-clustered three-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *Type* denotes the area estimator (*sy* for synthetic, *sm* for small and *ex* for extended synthetic). *Reference* gives the reference where found, else I derived them by analogy.

exh	Type	Reference	Formula
yes	<i>sy</i>	– –	$\hat{Y}_{c,G,synth} = \bar{Z}_G^t \hat{\beta}_{c,s_2}$ $\hat{V} = \bar{Z}_G^t \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \bar{Z}_G$
yes	<i>sm</i>	– –	$\hat{Y}_{c,G,small} = \hat{Y}_{c,G,synth} + \bar{R}_{c,s_2,G}(x)$ $\hat{V} = \hat{V}(\hat{Y}_{c,G,synth}) + \frac{1}{n_{s_2,G}} \hat{V}(\bar{R}_{c,s_2,G}(x))$
yes	<i>ex</i>	[2, eq. 48] [2, eq. 49]	$\hat{Y}_{c,G,synth} = \bar{Z}_G^t \hat{\theta}_{c,s_2}$ $\hat{V} = \bar{Z}_G^t \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \bar{Z}_G$
no	<i>sy</i>	[2, eq. 42] [2, eq. 43]	$\hat{Y}_{c,G,psynth} = \hat{Z}_{c,s_1,G}^1 \hat{\beta}_{c,s_2}$ $\hat{V}(\cdot) = \hat{Z}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{Z}_{c,s_1,G} + \hat{\beta}_{c,s_2}^t \hat{\Sigma}_{\hat{Z}_{c,s_1,G}} \hat{\beta}_{c,s_2}$
no	<i>sm</i>	[2, eq. 44] [2, eq. 45]	$\hat{Y}_{c,G,psmall} = \hat{Y}_{c,G,psynth} + \bar{R}_{c,s_2,G}(x)$ $\hat{V} = \hat{V}(\hat{Y}_{c,G,psynth}) + \frac{1}{n_{s_2,G}} \hat{V}(\bar{R}_{c,s_2,G}(x))$
no	<i>ex</i>	[2, eq. 46] [2, eq. 47]	$\hat{Y}_{c,G,psynth} = \hat{Z}_{c,s_1,G}^t \hat{\theta}_{c,s_2}$ $\hat{V} = \hat{Z}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{Z}_{c,s_1,G} + \hat{\theta}_{c,s_2}^t \hat{\Sigma}_{\hat{Z}_{c,s_1,G}} \hat{\theta}_{c,s_2}$
part	<i>sy</i>	– –	$\hat{Y}_{c,psynth,G,greg} = \left(\bar{Z}_G^{(1)} - \hat{Z}_{c,s_1,G}^{(1)} \right) \hat{\alpha}_{c,2} + \hat{Z}_{c,s_1,G}^t \hat{\beta}_{c,s_2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \bar{Z}_G^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \bar{Z}_G^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}} \right) \hat{Z}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{Z}_{c,s_1,G}$
part	<i>sm</i>	– –	$\hat{Y}_{c,G,greg} = \hat{Y}_{c,G,psynth} + \bar{R}_{c,s_2,G}(x)$ $\hat{V} = \hat{V}(\hat{Y}_{c,psynth,G,greg}) + \frac{1}{n_{s_2,G}} \hat{V}(\bar{R}_{c,s_2,G}(x))$
part	<i>ex</i>	[3, eq. 50] [3, eq. 52]	$\hat{Y}_{c,G,greg} = \left(\bar{Z}_G^{(1)} - \hat{Z}_{c,s_1,G}^{(1)} \right) \hat{\gamma}_{c,2} + \hat{Z}_{c,s_1,G}^t \hat{\theta}_{c,2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \bar{Z}_G^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \bar{Z}_G^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}} \right) \hat{Z}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{Z}_{c,s_1,G}$

Table 3: Predictors for clustered two-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *Type* denotes the area estimator (*sy* for synthetic, *sm* for small and *ex* for extended synthetic). *Reference* gives the reference where found, else I derived them by analogy.

exh	Type	Reference	Formula
yes	<i>sy</i>	–	–
yes	<i>sm</i>	–	–
yes	<i>ex</i>	–	–
no	<i>sy</i>	–	$\hat{Y}_{G,psynth,3p} = \left(\hat{\bar{Z}}_{c,s_0,G}^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\alpha}_{c,2} + \hat{\bar{Z}}_{c,s_1,G}^t \hat{\beta}_{c,s_2}$ $\hat{V} = \hat{\alpha}_{c,s_2}^t \hat{\Sigma}_{\hat{\bar{Z}}_{c,s_0,G}^{(1)}} \hat{\alpha}_{c,s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{Z}}_{c,s_1,G}$
no	<i>sm</i>	–	$\hat{Y}_{c,G,psmall,3p} = \hat{Y}_{G,psynth,G,3p} + \bar{\hat{R}}_{c,s_2,G}(x)$ $\hat{V} \approx \hat{V} \left(\hat{Y}_{c,G,psynth,3p} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\bar{\hat{R}}_{c,s_2,G}(x))$
no	<i>ex</i>	[5, eq. 53] [5, eq. 55]	$\hat{Y}_{c,G,g3reg} = \left(\hat{\bar{Z}}_{c,s_0,G}^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\gamma}_{c,2} + \hat{\bar{Z}}_{c,s_1,G}^t \hat{\theta}_{c,2}$ $\hat{V} = \hat{\gamma}_{c,s_2}^t \hat{\Sigma}_{\hat{\bar{Z}}_{c,s_0,G}^{(1)}} \hat{\gamma}_{c,s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{c,2}} \hat{\bar{Z}}_{c,s_1,G}$
part	<i>sy</i>	–	$\hat{Y}_{c,G,ynth,3p} = \left(\bar{\hat{Z}}_G^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\alpha}_{c,2} + \hat{\bar{Z}}_{c,s_1,G}^t \hat{\beta}_{c,s_2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{Z}}_{c,s_1,G}$
part	<i>sm</i>	–	$\hat{Y}_{c,G,small,3p} = \hat{Y}_{c,G,ynth,3p} + \bar{\hat{R}}_{c,s_2,G}(x)$ $\hat{V} \approx \hat{V} \left(\hat{Y}_{c,G,ynth,3p} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\bar{\hat{R}}_{c,s_2,G}(x))$
part	<i>ex</i>	–	$\hat{Y}_{c,G,extsynth,3p} = \left(\bar{\hat{Z}}_G^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\gamma}_{c,2} + \hat{\bar{Z}}_{c,s_1,G}^t \hat{\theta}_{c,2}$ $\hat{V} = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + \left(1 - \frac{n_{s_2}}{n_{s_1}}\right) \hat{\bar{Z}}_{c,s_1,G}^t \hat{\Sigma}_{\hat{\theta}_{c,2}} \hat{\bar{Z}}_{c,s_1,G}$

Table 4: Predictors for clustered three-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *Type* denotes the area estimator (*sy* for synthetic, *sm* for small and *ex* for extended synthetic). *Reference* gives the reference where found, else I derived them by analogy.