Estimators in Detail

Andreas Dominik Cullmann

March 6, 2020

1 Introduction

Superscripts For partially exhaustive auxiliary information, Mandallaz ([1, p. 1023], [2, p. 383f] defines $Z^t(x) = Z^{(1)t}(x) + Z^{(2)t}(x)$ whereas Hill [3, 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 uses 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 have change all set indices to $s_{[012]}$.

Hill uses $\bar{Z}^{(1)}_{0,G}$ (and $\bar{Z}^{(1)}_0$ which ([3, p. 18]) is the exact mean). So I do drop the index, which is misleadingly referring to some set (and I do so for $\bar{\mathcal{Z}}^{(1)}_{0,G}$).

Mandallaz uses $\bar{\hat{R}}_{2,G}$ when calculating the variance of the residuals in G, for example in a2.26, where $\bar{\hat{R}}_{2,G}$ is clearly $\bar{\hat{R}}(x)$ while summing over s_2 and G. I use the latter form.

References I reference [4] as a1, [5] as a2, [6] as b1, [1] as b2, [7] as c1, [2] as c2 and [3] as h.

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 two-phase sampling only, three-phase sampling with full exhaustive auxiliary information is just two-phase sampling with full exhaustive auxiliary information with more observations), non-exhaustive and partially exhaustive. In each block the (pseudo) synthetic the (pseudo) small and the (pseudo) extended estimator and their variances are given.

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

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

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

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 [3] implemented the partially exhaustive auxiliary information using a full and a reduced model. So they see it as three-phase sampling where [1] clearly see it as two-phase sampling with partially exhaustive auxiliary information.

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

exh	type	ref	formula
yes	synthetic	a2.18	$\hat{Y}_{G,synth} = \bar{Z}^t_G \hat{\beta}_{s_2}$
-	-	a2.19	$\hat{V}\,(\times) = \bar{Z}^t_G \hat{\Sigma}_{\hat{\beta}_{s_2}} \bar{Z}_G$
yes	small	a2.20	$\hat{Y}_{G,small} = \hat{Y}_{G,synth} + \bar{\hat{R}}_{s_2,G}(x)$
-	-	a2.21	$\hat{V}\,(\times) \approx \hat{V}\left(\hat{Y}_{G,synth}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
yes	extended	a2.31	$\hat{\bar{Y}}_{G,synth} = \bar{\mathcal{Z}}^t_G \hat{\theta}_{s_2}$
-	-	a2.33	$\hat{V}\,(\times) = \bar{\mathcal{Z}}^t_G \hat{\Sigma}_{\hat{\theta}_{s_2}} \bar{\mathcal{Z}}_G$
no	synthetic	a2.22	$\hat{Y}_{G,psynth} = \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\beta}_{s_2}$
-	-	a2.23	$\hat{V}\,(\times) = \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{\bar{Z}}}_{s_1,G} + \hat{\beta}^t_{s_2} \hat{\Sigma}_{\hat{\bar{\bar{Z}}}_{s_1,G}} \hat{\beta}_{s_2}$
no	small	a2.25	$\hat{Y}_{G,psmall} = \hat{Y}_{G,psynth} + \bar{\hat{R}}_{s_2,G}(x)$
-	-	a2.26	$\hat{V}\,(\times) \approx \hat{V}\left(\hat{Y}_{G,synth}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
no	extended	a2.35	$\hat{\bar{Y}}_{G,psynth} = \hat{\bar{\bar{\mathcal{Z}}}}^t_{s_1,G} \hat{\theta}_{s_2}$
-	-	a2.36	$\hat{V}\,(\times) = \hat{\bar{\bar{\mathcal{Z}}}}^t_{s_1,G} \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{\bar{\mathcal{Z}}}}_{s_1,G} + \hat{\theta}^t_{s_2} \hat{\Sigma}_{\hat{\bar{\bar{\mathcal{Z}}}}_{s_1,G}} \hat{\theta}_{s_2}$
part	synthetic	b2.34	$\hat{Y}_{psynth,G,greg} = \left(\bar{Z}^{(1)}_G - \hat{\bar{\bar{Z}}}^{(1)}_{s_1,G}\right) \hat{\alpha}_{s_2} + \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\beta}_{s_2}$
-	-	b2.35	$\hat{V}\,(\times) = \frac{n_{s_2}}{n_{s_1}} \bar{Z}^{(1)t}_G \hat{\Sigma}_{\hat{\alpha}_{s_2}} \bar{Z}^{(1)}_G + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{\bar{Z}}}_{s_1,G}$
part	small	b2.24	$\hat{Y}_{G,greg} = \hat{Y}_{psynth,G,greg} + \bar{\hat{R}}_{s_2,G}(x)$
-	-	b2.23	$\hat{V}\,(\times) \approx \hat{V}\left(\hat{Y}_{psynth,G,greg}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
part	extended	b2.30	$\hat{\bar{Y}}_{G,greg} = \left(\bar{\mathcal{Z}}^{(1)}_G - \hat{\bar{\bar{\mathcal{Z}}}}^{(1)}_{s_1,G}\right) \hat{\gamma}_{s_2} + \hat{\bar{\bar{\mathcal{Z}}}}^t_{s_1,G} \hat{\theta}_{s_2}$
-	-	b2.31	$\hat{V}\,(\times) = \frac{n_{s_2}}{n_{s_1}} \bar{\mathcal{Z}}^{(1)t}_G \hat{\Sigma}_{\hat{\gamma}_{s_2}} \bar{\mathcal{Z}}^{(1)}_G + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\bar{\mathcal{Z}}}}^t_{s_1,G} \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{\bar{\mathcal{Z}}}}_{s_1,G}$

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), *small* denotes the area estimator.

exh	type	ref	formula
part	synthetic	h.26a	$\hat{Y}_{G,synth,3p} = \left(\bar{Z}^{(1)}_G - \hat{\bar{\bar{Z}}}^{(1)}_{s_1,G}\right) \hat{\alpha}_{s_2} + \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\beta}_{s_2}$
-	-	h.26c	$\hat{V}\,(\times) = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{\bar{Z}}}^{(1)t}_{s_0,G} \hat{\Sigma}_{\hat{\alpha}_{s_2}} \hat{\bar{\bar{Z}}}^{(1)}_{s_0,G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\bar{Z}}}^t_{s_1,G} \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{\bar{Z}}}_{s_1,G}$
part	small	h.22a	$\hat{Y}_{G,small,3p} = \hat{Y}_{G,synth,3p} + \bar{\hat{R}}_{s_2,G}(x)$
-	-	h.23a	$\hat{V}\,(\times) \approx \hat{V}\left(\hat{Y}_{G,synth,3p}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
part	extended	extending h.26a	$\hat{\bar{Y}}_{G,extsynth,3p} = \left(\bar{\mathcal{Z}}^{(1)}_G - \hat{\bar{\bar{\mathcal{Z}}}}^{(1)}_{s_1,G}\right) \hat{\gamma}_{s_2} + \hat{\bar{\bar{\mathcal{Z}}}}^t_{s_1,G} \hat{\theta}_{s_2}$

exh	type	ref	formula
-	-	extending h.26c	$\hat{V}\left(\times\right) = \frac{n_{s_2}}{n_{s_1}} \bar{\mathcal{Z}}^{(1)t}_{s_0,G} \hat{\Sigma}_{\hat{\gamma}_{s_2}} \bar{\mathcal{Z}}^{(1)}_{s_0,G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{s_1,G} \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{\hat{\mathcal{Z}}}}_{s_1,G}$
no	synthetic	h.26b	$\hat{Y}_{G,psynth,3p} = \left(\hat{\bar{\hat{Z}}}^{(1)}_{s_0,G} - \hat{\bar{\hat{Z}}}^{(1)}_{s_1,G} \right) \hat{\alpha}_{s_2} + \hat{\bar{\hat{Z}}}^{t}_{s_1,G} \hat{\beta}_{s_2}$
-	-	h.26d	$\hat{V}\left(\times\right) = \hat{\alpha}^{t}_{s_2} \hat{\Sigma}_{\hat{\bar{\hat{Z}}}^{(1)}_{s_0,G}} \hat{\alpha}_{s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{\hat{Z}}}^{(1)t}_{s_0,G} \hat{\Sigma}_{\hat{\alpha}_{s_2}} \hat{\bar{\hat{Z}}}^{(1)}_{s_0,G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\hat{Z}}}^{t}_{s_1,G} \hat{\Sigma}_{\hat{\beta}_{s_2}} \hat{\bar{\hat{Z}}}_{s_1,G}$
no	small	h.22b	$\hat{Y}_{G,psmall,3p} = \hat{Y}_{G,psynth,G,3p} + \bar{\hat{R}}_{s_2,G}(x)$
-	-	h.23b	$\hat{V}\left(\times\right) \approx \hat{V}\left(\hat{\hat{Y}}_{G,psynth,3p} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{s_2,G}(x))$
no	extended	c2.23	$\hat{\hat{Y}}_{G,g3reg} = \left(\hat{\bar{\hat{\mathcal{Z}}}}^{(1)}_{s_0,G} - \hat{\bar{\hat{\mathcal{Z}}}}^{(1)}_{s_1,G} \right) \hat{\gamma}_{s_2} + \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{s_1,G} \hat{\theta}_{s_2}$
-	-	c2.24	$\hat{V}\left(\times\right) = \hat{\gamma}^{t}_{s_2} \hat{\Sigma}_{\hat{\bar{\hat{\mathcal{Z}}}}^{(1)}_{s_0,G}} \hat{\gamma}_{s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{\hat{\mathcal{Z}}}}^{(1)t}_{s_0,G} \hat{\Sigma}_{\hat{\gamma}_{s_2}} \hat{\bar{\hat{\mathcal{Z}}}}^{(1)}_{s_0,G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{s_1,G} \hat{\Sigma}_{\hat{\theta}_{s_2}} \hat{\bar{\hat{\mathcal{Z}}}}_{s_1,G}$

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), *small* denotes the area estimator.

exh	small	ref	formula
yes	no	analogy	$\hat{Y}_{c,G,synth} = \bar{Z}^{t}_{G} \hat{\beta}_{c,s_2}$
-	-	analogy	$\hat{V}\left(\times\right) = \bar{Z}^{t}_{G} \hat{\Sigma}_{\hat{\beta}_{s_2}} \bar{Z}_{G}$
yes	yes	analogy	$\hat{Y}_{c,G,small} = \hat{Y}_{c,G,synth} + \bar{\hat{R}}_{c,s_2,G}(x)$
-	-	analogy	$\hat{V}(\times) = \hat{V}\left(\hat{Y}_{c,G,synth} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{c,s_2,G}(x))$
yes	no	a2.48	$\hat{\hat{Y}}_{c,G,synth} = \bar{\mathcal{Z}}^{t}_{G} \hat{\theta}_{c,s_2}$
-	-	a2.49	$\hat{V}\left(\times\right) = \bar{\mathcal{Z}}^{t}_{G} \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \bar{\mathcal{Z}}_{G}$
no	no	a2.42	$\hat{Y}_{c,G,psynth} = \hat{\bar{\hat{Z}}}^{1}_{c,s_1,G} \hat{\beta}_{c,s_2}$
-	-	a2.43	$\hat{V}(\times) = \hat{\bar{\hat{Z}}}^{t}_{c,s_1,G} \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{\hat{Z}}}_{c,s_1,G} + \hat{\beta}^{t}_{c,s_2} \hat{\Sigma}_{\hat{\bar{\hat{Z}}}_{c,s_1,G}} \hat{\beta}_{c,s_2}$
no	yes	a2.44	$\hat{Y}_{c,G,psmall} = \hat{Y}_{c,G,psynth} + \bar{\hat{R}}_{c,s_2,G}(x)$
-	-	a2.45	$\hat{V}(\times) = \hat{V}\left(\hat{Y}_{c,G,psynth} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{c,s_2,G}(x))$
no	no	a2.46	$\hat{\hat{Y}}_{c,G,psynth} = \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{c,s_1,G} \hat{\theta}_{c,s_2}$
-	-	a2.47	$\hat{V}\left(\times\right) = \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{c,s_1,G} \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{\bar{\hat{\mathcal{Z}}}}_{c,s_1,G} + \hat{\theta}^{t}_{c,s_2} \hat{\Sigma}_{\hat{\bar{\hat{\mathcal{Z}}}}_{c,s_1,G}} \hat{\theta}_{c,s_2}$
part	no	analogy	$\hat{Y}_{c,psynth,G,greg} = \left(\bar{Z}^{(1)}_{G} - \hat{\bar{\hat{Z}}}^{(1)}_{c,s_1,G} \right) \hat{\alpha}_{c,2} + \hat{\bar{\hat{Z}}}^{t}_{c,s_1,G} \hat{\beta}_{c,s_2}$
-	-	analogy	$\hat{V}\left(\times\right) = \frac{n_{s_2}}{n_{s_1}} \bar{Z}^{(1)t}_{G} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \bar{Z}^{(1)}_{G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\hat{Z}}}^{t}_{c,s_1,G} \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{\hat{Z}}}_{c,s_1,G}$
part	yes	analogy	$\hat{Y}_{c,G,greg} = \hat{Y}_{c,G,psynth} + \bar{\hat{R}}_{c,s_2,G}(x)$
-	-	analogy	$\hat{V}\left(\times\right) = \hat{V}\left(\hat{Y}_{c,psynth,G,greg} \right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{c,s_2,G}(x))$
part	no	b1.50	$\hat{\hat{Y}}_{c,G,greg} = \left(\bar{\mathcal{Z}}^{(1)}_{G} - \hat{\bar{\hat{\mathcal{Z}}}}^{(1)}_{c,s_1,G} \right) \hat{\gamma}_{c,2} + \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{c,s_1,G} \hat{\theta}_{c,2}$
-	-	b1.52	$\hat{V}\left(\times\right) = \frac{n_{s_2}}{n_{s_1}} \bar{\mathcal{Z}}^{(1)t}_{G} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \bar{\mathcal{Z}}^{(1)}_{G} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\hat{\mathcal{Z}}}}^{t}_{c,s_1,G} \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{\bar{\hat{\mathcal{Z}}}}_{c,s_1,G}$

Table 3: Predictors for clustered two-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *small* denotes the area estimator.

exh	type	ref	formula
part	no	analogy	$\hat{Y}_{c,G,synth,3p} = \left(\bar{Z}_G^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\alpha}_c, 2 + \hat{\bar{Z}}_{c,s_1,G}^{t} \hat{\beta}_{c,s_2}$
-	-	analogy	$\hat{V}\left(\times\right) = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{Z}}_{c,s_1,G}^{t} \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{Z}}_{c,s_1,G}$
part	yes	analogy	$\hat{Y}_{c,G,small,3p} = \hat{Y}_{c,G,synth,3p} + \hat{\bar{R}}_{c,s_2,G}(x)$
-	-	analogy	$\hat{V}\left(\times\right) \approx \hat{V}\left(\hat{Y}_{c,G,synth,3p}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{c,s_2,G}(x))$
part	no	analogy	$\hat{\tilde{Y}}_{c,G,extsynth,3p} = \left(\bar{\mathcal{Z}}_G^{(1)} - \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{(1)} \right) \hat{\gamma}_{c,2} + \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{t} \hat{\theta}_{c,2}$
-	-	analogy	$\hat{V}\left(\times\right) = \frac{n_{s_2}}{n_{s_1}} \hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{t} \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{\bar{\mathcal{Z}}}_{c,s_1,G}$
no	no	analogy	$\hat{Y}_{G,psynth,3p} = \left(\hat{\bar{Z}}_{c,s_0,G}^{(1)} - \hat{\bar{Z}}_{c,s_1,G}^{(1)} \right) \hat{\alpha}_{c,2} + \hat{\bar{Z}}_{c,s_1,G}^{t} \hat{\beta}_{c,s_2}$
-	-	analogy	$\hat{V}\left(\times\right) = \hat{\alpha}_{c,s_2}^{t} \hat{\Sigma}_{\hat{\bar{Z}}_{c,s_0,G}^{(1)}} \hat{\alpha}_{c,s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{Z}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\alpha}_{c,s_2}} \hat{\bar{Z}}_{c,s_0,G}^{(1)} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{Z}}_{c,s_1,G}^{t} \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\bar{Z}}_{c,s_1,G}$
no	yes	analogy	$\hat{Y}_{c,G,psmall,3p} = \hat{Y}_{G,psynth,G,3p} + \hat{\bar{R}}_{c,s_2,G}(x)$
-	-	analogy	$\hat{V}\left(\times\right) \approx \hat{V}\left(\hat{\tilde{Y}}_{c,G,psynth,3p}\right) + \frac{1}{n_{s_2,G}} \hat{V}(\hat{R}_{c,s_2,G}(x))$
no	no	c1.53	$\hat{\tilde{Y}}_{c,G,g3reg} = \left(\hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)} - \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{(1)} \right) \hat{\gamma}_{c,2} + \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{t} \hat{\theta}_{c,2}$
-	-	c1.55	$\hat{V}\left(\times\right) = \hat{\gamma}_{c,s_2}^{t} \hat{\Sigma}_{\hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)}} \hat{\gamma}_{c,s_2} + \frac{n_{s_2}}{n_{s_1}} \hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)t} \hat{\Sigma}_{\hat{\gamma}_{c,s_2}} \hat{\bar{\mathcal{Z}}}_{c,s_0,G}^{(1)} + (1 - \frac{n_{s_2}}{n_{s_1}}) \hat{\bar{\mathcal{Z}}}_{c,s_1,G}^{t} \hat{\Sigma}_{\hat{\theta}_{c,s_2}} \hat{\bar{\mathcal{Z}}}_{c,s_1,G}$

Table 4: Predictors for clustered three-phase sampling, *exh* denotes exhaustiveness of auxiliary information (yes/no/part, the latter meaning partially exhaustive auxiliary information), *small* denotes the area estimator.

```
> fake_weights <- function(df) {
+     df[["weights"]] <- 1
+     df[["weights"]][df[["x2"]] == 0] <- 0.12
+     return(df)
+ }
> suppressWarnings(rm(s1, s2, s0))
> data("s1", "s2", "s0", package = "maSAE")
> s0$x1 <- s0$x3 <- NULL
> s0 <- fake_weights(s0)
> s1 <- fake_weights(s1)
> s2 <- fake_weights(s2)
> s12 <- maSAE::bind_data(s1, s2)
> s012 <- maSAE::bind_data(s1, s2, s0)
> tm <- data.frame(x1 = c(150, 200), x2 = c(23, 23), x3 = c(7, 7.5), g = c("a", "b"))
> tm_p <- data.frame(x2 = c(23, 23), g = c("a", "b"))
> #% unclustered
>
> ##% un-weighted
> ###% two-phase
> ####% partially exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance  psynth var_psynth   psmall var_psmall
1         a   374.4659 288.8545 362.889   62.83338 374.3667   293.4668
```

```
2         b   384.7822 250.1123 378.087   61.99708 384.7325   202.7989

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   365.0818 221.8780 353.5113   40.29034 364.9890   270.9238
2         b   380.5571 144.4836 373.8780   40.03398 380.5234   180.8358

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% non-exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2")
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   378.8590 487.3680 367.2796   303.1129 378.7573   533.7463
2         b   391.8262 417.3442 385.1562   314.9198 391.8016   455.7217

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ###% three-phase
> object <- maSAE::saObj(data = s012, f = y ~ x1 + x2 + x3 | g, s1 = "phase1", s2 = "phase
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   397.1866 311.8078 385.5963   82.87085 397.0740   313.5043
2         b   404.2197 271.9036 397.5939   82.71266 404.2394   223.5145

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ##% weighted
> ###% two-phase
> ####% partially exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g,
+                          s2 = "phase2", smallAreaMeans = tm_p,
+                          auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))
```

```
  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   376.9497 291.8285 365.3586   61.26629 376.8363   291.8997
2         b   389.2490 243.3468 382.4815   60.04105 389.1269   200.8429
```

```
> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)
```

```
[1] TRUE
```

```
> ####% exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g,
+                         s2 = "phase2", smallAreaMeans = tm,
+                         auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))
```

```
  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   365.0818 221.8780 353.5113   40.29034 364.9890   270.9238
2         b   380.5571 144.4836 373.8780   40.03398 380.5234   180.8358
```

```
> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)
```

```
[1] TRUE
```

```
> ####% non-exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g,
+                         s2 = "phase2",
+                         auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))
```

```
  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   406.4824 494.2358 394.8737   304.2430 406.3514   534.8764
2         b   430.0021 413.9014 423.3800   317.5597 430.0255   458.3616
```

```
> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)
```

```
[1] TRUE
```

```
> ###% three-phase
> object <- maSAE::saObj(data = s012, f = y ~ x1 + x2 + x3 | g,
+                         s1 = "phase1", s2 = "phase2",
+                         auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))
```

```
  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   424.8187 322.9752 413.1992   86.27596 424.6769   316.9094
2         b   437.5184 272.6481 430.9232   86.47232 437.5686   227.2741
```

```
> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)
```

```
[1] TRUE
```

```
> #% clustered
> ##% un-weighted
> ###% two-phase
> ####% partially exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   377.0824 549.1471 363.3825   100.3859 376.8559   556.3546
2         b   385.1621 458.4486 380.3530   107.9309 385.0648   396.8564

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   368.1216 428.7513 354.4524   74.40279 367.9259   530.3716
2         b   381.1313 328.3773 376.3492   84.26918 381.0611   373.1947

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% non-exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", cluster = "c
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance  psynth var_psynth   psmall var_psmall
1         a   381.5729 950.4594 367.868   580.6438 381.3414  1036.6126
2         b   392.3395 807.3224 387.575   575.5981 392.2869   864.5236

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ###% three-phase
> object <- maSAE::saObj(data = s012, f = y ~ x1 + x2 + x3 | g, s1 = "phase1", s2 = "phase
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   400.3065 600.9477 386.5807   135.3258 400.0542   591.2946
2         b   404.9677 493.6005 400.2816   143.9249 404.9935   432.8504

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)
```

```
[1] TRUE

> ##% weighted
> ###% two-phase
> ####% partially exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
+                        auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   376.6663 550.5192 363.4477   100.4981 376.4431   558.1007
2         b   384.8009 457.7284 380.2073   107.8563 384.7049   396.4425

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", smallAreaMea
+                        auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   368.2336 430.3318 355.0431   74.32785 368.0385   531.9304
2         b   381.3361 327.0091 376.7705   83.95160 381.2681   372.5378

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ####% non-exhaustive
> object <- maSAE::saObj(data = s12, f = y ~ x1 + x2 + x3 | g, s2 = "phase2", cluster = "c
+                        auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   381.1568 950.9773 367.9332   579.8245 380.9286  1037.4271
2         b   391.9784 805.4768 387.4293   574.4109 391.9270   862.9971

> outlm <- maSAE::predict(object, use_lm = TRUE)
> RUnit::checkEquals(out, outlm)

[1] TRUE

> ###% three-phase
> object <- maSAE::saObj(data = s012, f = y ~ x1 + x2 + x3 | g, s1 = "phase1", s2 = "phase
+                        auxiliaryWeights = "weights")
> (out <- maSAE::predict(object, use_lm = FALSE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   399.8904 602.3198 386.6459   135.4381 399.6414   593.0406
2         b   404.6066 492.8803 400.1359   143.8503 404.6336   432.4365
```

```
> (outlm <- maSAE::predict(object, use_lm = TRUE))

  smallArea prediction variance   psynth var_psynth   psmall var_psmall
1         a   399.8904 602.3198 386.6459   135.4381 399.6414   593.0406
2         b   404.6066 492.8803 400.1359   143.8503 404.6336   432.4365

> RUnit::checkEquals(out, outlm)

[1] TRUE

>
>
```

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.