Master's thesis:
Combining probability and non-probability samples for estimation

Study program:
Methodology and Statistics for the Behavioural, Biomedical, and Social Sciences

Student:
Sofía Villalobos Aliste (6060714)

Program coordinator:
Rens van de Schoot

Supervisors:
Ton de Waal
Sander Scholtus

Examiner:
Daniel Oberski

Date:
May 9th, 2022

Preferred journal of publication:
Journal of Survey Statistics and Methodology

Word count:
6970

MASTER'S THESIS

COMBINING PROBABILITY AND NON-PROBABILITY SAMPLES FOR ESTIMATION

SOFÍA F. VILLALOBOS ALISTE*

Probability surveys are experiencing important drawbacks nowadays: participation rates are decreasing, and the respondent burden is increasing, which could yield less accurate estimates, needing a big investment of resources in order to benefit from the feature of unbiasedness that makes them so desirable. Alternatively, non-probability samples like administrative records are having a rise in popularity due to their convenience and low costs, and research is ongoing on how to deal with their unknown bias. This study seeks to find a method that combines a probability and a non-probability sample to overcome the drawbacks of both. We do this by computing a weight for each sample that results in a reduction of the mean squared error (MSE) for an estimator, in comparison to the MSEs of the separate estimators from the probability and the non-probability sample. We performed simulation studies where two different methods of modelling the bias of an estimator based on the non-probability sample were tested, and the MSE was evaluated. We then applied these methods to a real dataset from Statistics Netherlands (CBS) to show that the MSEs can indeed be reduced. For both methods we find that (i) when the size of the probability sample is large, using the estimator based on the probability sample is best, (ii) when the non-probability sample is highly selective, using the estimator based on the probability sample is best, (iii) when the non-probability sample is hardly selective, using the estimator based on the nonprobability sample is best, (iv) in all other cases the combined estimator is the best.

KEYWORDS: Administrative records; Combining samples; Bias model; Small area estimation; Weighting samples.

* Methodology & Statistics for the Behavioural, Biomedical and Social Sciences, Utrecht University, the Netherlands. E-mail: s.f.villalobosaliste@students.uu.nl

1. INTRODUCTION

The sample survey, and the probability sampling framework with it, emerged as a field in the first half of the twentieth century intending to satisfy the need of keeping a general record of the nationwide population. This need has evolved into more complex ones, increasingly broken down to the point of requiring more and more precise statistics for different subpopulations which we can also call domains. The needs can be reflected by policymakers and the importance of resource allocation, social program implementation, and environmental planning and by the private market, especially from small businesses that rely on such estimations (Pfeffermann 2002, Rao 2003).

Implementing a probability survey is demanding, but it carries several benefits, the most important being that it allows for making unbiased inferences about a target population (Lohr, 2019). Nevertheless, it has encountered many drawbacks in the last few years. And due to its high cost, increasing non-response, and debate about the coverage error that this implies, their validity has started to be questioned (Brick, 2014).

Hand in hand with that, in the latest years, there has been an increasing interest in the statistical use of data that can be obtained from different data sources far from the probability sampling framework: the so-called non-probability samples. This type of data lacks the virtuous properties of data that have been drawn from a probability design, the framework that undoubtedly prevailed until the end of the 20th century. However, they have earned their relevance, especially from one of their most kind characteristics: cost reduction.

Statistical Offices are particularly interested in non-probability samples since they are constantly motivated to improve the quality of estimators and reduce data collection costs (Van den Brakel, 2019). Examples of non-probability data at national Statistical institutes such as Statistics Netherlands (CBS) are many administrative datasets, such as administrative data from the Tax Office, administrative data from energy providers, and administrative data from education registers.

On one hand, a non-probability sample can be counterproductive for various reasons: there is no clear framework to work with and such data can easily lead to a large bias in estimators because of selectivity. Selectivity is a major issue to deal with when using non-probability samples, and it refers to the situation in which the part of the target population included in the sample is substantially different from the part of the population that is not included in the sample, making it difficult to do inference about the target population. These samples are likely to be selective regarding the population and the selection probability of units remains unknown (Elliott & Valliant, 2017).

But then how do non-probability samples help the matter of improving the quality of estimators? On the other hand, this type of sample can be considered very inexpensive in comparison with a probability one, and methods that try to account for selectivity have been developed. Moreover, the use of non-probability samples could eventually help reduce the variance of estimators for parameters of interest. In the case of CBS, efforts have been made to go from obtaining their estimates from surveys to obtaining them from administrative records, which could be a valuable source of information to overcome the shortcomings of surveys (Linder et al., 2014).

The increased interest in non-probability samples can be observed, first, in Valliant (2020), who examines different methods to obtain estimates from a non-probability sample such as quasi randomization, superpopulation, and doubly robust estimation, obtaining mixed results; doubly robust estimation was the least biased of all methods, but still showed a large bias that could not be corrected with currently used methods. Also, Baker et al. (2013) review sample matching, network sampling, and weighting methods, pointing out their difficulties and providing guidelines for future research. In these cases, we can observe that trying to compute estimators just from non-probability samples requires a lot of research, and results are not always so promising. Literature has addressed these issues by investigating some methods to integrate probability and non-probability samples, even suggesting avoiding non-probability samples in some scenarios when some type of probability sample is available (Elliott & Valliant, 2017).

Elliott & Haviland (2007) combined a probability survey design and a web convenience sample, concluding that this could be useful under very specific circumstances regarding selectivity and bias due to that, size of the probability sample, and inexpensiveness of the convenience sample. Wiśniowski et al. (2020) use a Bayesian approach to combine a probability sample and a non-probability sample by including the non-probability sample as prior information. They conclude that this is effective in reducing variance and mean squared error (MSE), assuming that the probability sample is unbiased.

In general, the combination of samples has shown more encouraging results than using a non-probability sample only, and in that direction, the method this project investigates is utilizing a combination of data from a large non-probability sample and a relatively small probability sample, using separate estimators for each sample, and weighting them, similar to Elliott & Haviland (2007). Elliott & Haviland (2007) assume that the bias in the estimate based on the non-probability sample is known. In our situation, we relax this assumption by using ideas from Pannekoek & De Waal (1998) who use a simple model for the bias of the estimator based on the non-probability sample. Based on this model they construct a combined estimator that is a weighted sum of the estimator for the probability sample and the estimator for the non-probability sample.

This study seeks to find a method that allows computing the weights that a probability sample and a non-probability sample should have in a combined estimator to result in a reduction of the MSE of that estimator in comparison to the MSEs of separate estimators for the probability and the non-probability sample, limiting attention to categorical data and focusing on estimating proportions.

The rest of the manuscript is organized as follows: In section 2, we specify the methodology and two modelling approaches A and B. In section 3, we describe the simulation conditions for the assessment of the methods and assess their performances. In section 4 we applied model B to real data from CBS. In section 5 we point out our main conclusions and propose suggestions for future research.

2. METHODOLOGY

The method this project investigates is inspired by a small area estimation approach. In this approach, we seek to estimate a proportion for different classes in a target population as observed in Table 1, where each row represents a domain (k), each column represents different categories (c), and each cell (Z_{kc}) represents the proportion of units within domain k that belongs to category c.

Table 1. Categories per domain

Domains	Categories					Total
	c = 1	c = 2	c = 3	. . .	c = C	
k = 1	Z_{11}	Z_{12}	Z_{13}	. . .	Z_{1C}	1
k = 2	Z_{21}					1
k = 3	Z_{31}					1
.	.					.
.	.					.
.	.					.
k = K	Z_{K1}				Z_{KC}	1

In small area estimation, we usually seek a reliable estimate per domain for which there have been only a few units sampled, and sometimes no units at all. In this situation, two estimates for the same proportion are computed. First, a direct estimator of Z ($\hat{Z}$) is computed per domain using the data available from a probability sample.

Then, an estimator $\hat{S}$ is also computed as an estimator of Z, which is an indirect estimator that assumes that domains can be similar regarding the parameter of interest. Therefore, it borrows this information to compute estimates from a model for each domain.

Estimators $\hat{Z}$ and $\hat{S}$ are combined into a new estimator ($\widehat{D}$) by taking a weighted average determined by the MSE of both estimators. When the MSE of both $\hat{Z}$ and $\hat{S}$ can be computed, estimator $\widehat{D}$ will have a lower MSE than both separate estimators (Pfeffermann, 2013).

In our case, the direct estimator $\hat{Z}$ is still based on the probability survey and the non-probability survey will take the place of a synthetic estimator $\hat{S}$. Estimator $\hat{Z}$ is generally not biased but will have a large variance, whereas estimator $\hat{S}$ is likely to be biased but generally has a smaller variance since it is based on a lot of units. Also, the bias of $\hat{S}$ is unknown, and the MSE of $\hat{S}$ cannot be computed. We, therefore, must rely on a model to compute an *expected* MSE.

2.1 Modelling approach A

Suppose there is a target population of N units ($i = 1, \dots, N$) with a categorical target variable y with $C \geq 2$ categories – i.e., $y_i \in \{1, \dots, C\}$ for each unit i – and that observations on y are available from two samples, regardless of whether the same unit is present in one or both samples. Furthermore, suppose units in the target population are divided into K domains ($k = 1, \dots, K$), where domain k contains N_k units.

We are interested in estimating Z_{kc}, this is the proportion of units in domain k of the target population that belongs to a category c, as illustrated in Table 1.

First, consider an example with two independent probability samples P_1 and P_2. We then would have two unbiased estimators $\hat{Z}_{kc,P_1}$ and $\hat{Z}_{kc,P_2}$. For many sampling designs, and certainly, for simple random sampling, we can estimate the sampling variance of $\hat{Z}_{kc,P_1}$ and $\hat{Z}_{kc,P_2}$, and therefore we can estimate MSEs of $\hat{Z}_{kc,P_1}$ and $\hat{Z}_{kc,P_2}$ (in this example, MSEs would be equal to the corresponding variances, since the estimators are unbiased).

We could construct a combined estimator of the form

$$\widehat{D}_{kc} = W_{kc}\, \hat{Z}_{kc,P_1} + (1 - W_{kc})\, \hat{Z}_{kc,P_2} \tag{2.1}$$

We would find optimal weights W_{kc} by minimizing the MSE of $\widehat{D}_{kc}$. These optimal weights are given by (Särndal et al., 1992, Section 9.9.1)

$$W_{kc} = \frac{MSE\left(\hat{Z}_{kc,P_2}\right)}{MSE\left(\hat{Z}_{kc,P_1}\right) + MSE\left(\hat{Z}_{kc,P_2}\right)} \tag{2.2}$$

In our situation, we have a probability sample $\mathcal{S}_P$ and a non-probability sample $\mathcal{S}_{NP}$. The primary interest in this research is to estimate the proportions of units belonging to certain categories across different domains. We want to estimate a proportion from a relatively small probability sample ($\hat{Z}_{kc,P}$), and from a non-probability one ($\hat{Z}_{kc,NP}$). We assume that the probability sample, $\mathcal{S}_P$, is a simple random sample and that $n_{Pk} \ll N_k$ where n_{Pk} stands for the sample size of domain k in the probability sample. We are especially interested in situations where $n_{NPk} \gg n_{Pk}$, where n_{NPk} stands for the sample size of domain k in the non-probability sample $\mathcal{S}_{NP}$.

The main problem that we now have is that the bias b_{kc} of $\hat{Z}_{kc,NP}$ is unknown and therefore $MSE(\hat{Z}_{kc,NP})$ cannot be computed. But, since we have that every unit in domain k has to fall into one category c, the over-estimation of a category will be compensated for by the under-estimation of one or more other categories. More precisely, this implies that, for each k,

$$\sum_{c=1}^{C} b_{kc} = 0 \tag{2.3}$$

This property suggests the following model (Model A) for b_{kc}: $E_b(b_{kc}) = 0$ and variance $E_b(b_{kc}^2) = \sigma_k^2$, where the subscript b indicates that this is a model for the bias b_{kc}. That is, Model A assumes that the expected value of the bias in the non-probability sample is 0 for each category c. Using this model, we can compute *expected* MSEs (EMSEs), and we will base our combined estimator on these EMSEs.

Under the model for the bias b_{kc} we can derive the following expressions for the model-based EMSE of $\hat{Z}_{kc,NP}$ and $\hat{Z}_{kc,P}$ (see Appendix A):

$$EMSE(\hat{Z}_{kc,P}) = \frac{E_b[\tilde{Z}_{kc}(1-\tilde{Z}_{kc})] - \sigma_k^2}{n_{Pk}} = \frac{1}{n_{Pk}}\left(\frac{n_{NPk}}{n_{NPk}-1} v_{kc} - \sigma_k^2\right) \tag{2.4}$$

$$EMSE(\hat{Z}_{kc,NP}) = \sigma_k^2 + \frac{v_{kc}}{n_{NPk}-1} \tag{2.5}$$

where $\tilde{Z}_{kc} = E_d(\hat{Z}_{kc,NP})$ and $v_{kc} = E_b E_d\left[\hat{Z}_{kc,NP}(1-\hat{Z}_{kc,NP})\right]$ and the subscript d refers to the known sampling design of the probability sample and the unknown sampling "design" of the non-probability sample.

To construct estimator $\widehat{D}_{kc}$, it is necessary to assign complementary weights to estimators $\hat{Z}_{kc,P}$ and $\hat{Z}_{kc,NP}$ (W_{kc}, respectively $1 - W_{kc}$). This weight is constructed by minimizing the EMSE of the combined estimator and is given analogously to (2.2) by

$$W_{kc} = \frac{EMSE\left(\hat{Z}_{kc,NP}\right)}{EMSE\left(\hat{Z}_{kc,NP}\right) + EMSE\left(\hat{Z}_{kc,P}\right)} \tag{2.6}$$

Here we assume that $\hat{Z}_{kc,NP}$ and $\hat{Z}_{kc,P}$ are independent estimators, which is true if the two samples were drawn independently from each other.

Substituting formulas (2.4) and (2.5) into (2.6) we obtain

$$\begin{aligned} W_{kc} &= \frac{\sigma_k^2 + \frac{v_{kc}}{n_{NPk} - 1}}{\left(1 - \frac{1}{n_{Pk}}\right)\sigma_k^2 + \frac{1 + n_{NPk}/n_{Pk}}{n_{NPk} - 1} v_{kc}} \\ &= \frac{(n_{NPk} - 1)\sigma_k^2 + v_{kc}}{(n_{NPk} - 1)\left(1 - \frac{1}{n_{Pk}}\right)\sigma_k^2 + \left(1 + \frac{n_{NPk}}{n_{Pk}}\right) v_{kc}} \end{aligned} \tag{2.7}$$

We also need to estimate the variance per domain and category (v_{kc}) and variance per domain (σ_k^2). We can use the following unbiased estimator for v_{kc}:

$$\hat{v}_{kc} = \hat{Z}_{kc,NP}\left(1 - \hat{Z}_{kc,NP}\right) \tag{2.8}$$

A consistent estimator for σ_k^2 is given by (see Appendix A):

$$\hat{\sigma}_k^2 = \frac{n_{Pk}}{C(n_{Pk} - 1)} \sum_{c=1}^{C} \left[\left(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}\right)^2 - \frac{1 + n_{NPk}/n_{Pk}}{n_{NPk} - 1} \hat{v}_{kc}\right] \tag{2.9}$$

Finally, we compute the combined estimator $\widehat{D}$, where A indicates the model used to compute the weights**:**

$$\widehat{D}_{kc}^{A} = W_{kc}\,\hat{Z}_{kc,P} + (1 - W_{kc})\,\hat{Z}_{kc,NP} \tag{2.10}$$

When a category inside a domain does not have any sampled units, $\hat{v}_{kc}$ would be 0. If added to that $\hat{\sigma}_k^2$ is also 0, then the calculation of (2.6) is undefined, and therefore it is not possible to obtain the EMSE of the combined estimator. In that case, we assume that $v_{kc} \approx \sigma_k^2$ and we compute W_{kc} in the following way instead:

$$W_{kc} = \frac{(n_{NPk}-1)+1}{(n_{NPk}-1)\left(1-\frac{1}{n_{Pk}}\right)+\left(1+\frac{n_{NPk}}{n_{Pk}}\right)} \tag{2.11}$$

2.2 Modelling approach B

As already mentioned, Model A assumes that the expected value of the bias in the non-probability sample is 0 for each category c. This assumption can be replaced instead by assuming that the expected value of the bias is constant across domains but potentially different for different categories.

We could therefore assume a model (Model B) such that b_{kc} is distributed as a random variable with mean $E_b(b_{kc}) = \beta_c$ (with $\sum_{c=1}^{C}\beta_c = 0$) and variance $E_b[(b_{kc}-\beta_c)^2] = \sigma_k^2$. Again, the subscript b indicates that this is a model for the bias b_{kc}

Under this model for the bias b_{kc} we can derive the following expressions for the model-based EMSE of $\hat{Z}_{kc,NP}$ and $\hat{Z}_{kc,P}$

$$EMSE\left(\hat{Z}_{kc,P}\right) = \frac{1}{n_{Pk}}\left(\frac{n_{NPk}}{n_{NPk}-1}v_{kc} + \beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right] - \beta_c^2 - \sigma_k^2\right) \tag{2.12}$$

$$EMSE\left(\hat{Z}_{kc,NP}\right) = \beta_c^2 + \sigma_k^2 + \frac{v_{kc}}{n_{NPk}-1} \tag{2.13}$$

where v_{kc} is the same as previously.

Substituting the above expressions (2.12) and (2.13) into (2.6) we obtain:

$$W_{kc} = \frac{\beta_c^2 + \sigma_k^2 + \frac{v_{kc}}{n_{NPk}-1}}{\left(1-\frac{1}{n_{Pk}}\right)(\beta_c^2+\sigma_k^2) + \frac{1+\frac{n_{NPk}}{n_{Pk}}}{n_{NPk}-1}v_{kc} + \frac{1}{n_{Pk}}\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]}$$

$$= \frac{(n_{NPk}-1)\left(\beta_c^2+\sigma_k^2\right)+v_{kc}}{(n_{NPk}-1)\left(1-\frac{1}{n_{Pk}}\right)(\beta_c^2+\sigma_k^2) + \left(1+\frac{n_{NPk}}{n_{Pk}}\right)v_{kc} + \frac{n_{NPk}-1}{n_{Pk}}\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]} \tag{2.14}$$

To apply this in practice we need to estimate σ_k^2, $\beta_c, E_b(\tilde{Z}_{kc})$ and $v_{kc} = E_b E_d\{\hat{Z}_{kc,NP}(1 - \hat{Z}_{kc,NP})\}$ from the two given samples. The latter two parts are easy: by definition an unbiased estimator of $E_b(\tilde{Z}_{kc}) = E_b[E_d(\hat{Z}_{kc,NP})]$ is given by $\hat{Z}_{kc,NP}$ itself, and an unbiased estimator of v_{kc} by

$$\hat{v}_{kc} = \hat{Z}_{kc,NP}(1 - \hat{Z}_{kc,NP}) \tag{2.15}$$

Thus, it only remains to find estimators for β_c and σ_k^2. First, to estimate β_c we can use

$$\hat{\beta}_c = \frac{1}{K}\sum_{k=1}^{K}(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}) \tag{2.16}$$

i.e., the average difference between the non-probability and the probability sample for category c across all domains, since we have under the assumed model that

$$E_b[E_d(\hat{\beta}_c)] = \frac{1}{K}\sum_{k=1}^{K} E_b[E_d(\hat{Z}_{kc,NP}) - E_d(\hat{Z}_{kc,P})] = \frac{1}{K}\sum_{k=1}^{K} E_b(\tilde{Z}_{kc} - Z_{kc}) = \frac{1}{K}\sum_{k=1}^{K} E_b(b_{kc}) = \beta_c$$

We also have (see Appendix A):

$$\hat{\sigma}_k^2 = \frac{n_{Pk}}{C(n_{Pk}-1)}\sum_{c=1}^{C}\left[(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P})^2 - \left(1 - \frac{1}{n_{Pk}}\right)\hat{\beta}_c^2 - \frac{2}{n_{Pk}}\hat{\beta}_c\hat{Z}_{kc,NP} - \frac{1 + n_{NPk}/n_{Pk}}{n_{NPk}-1}\hat{v}_{kc}\right] \tag{2.17}$$

Finally, we compute the combined estimator $\widehat{D}_{kc}^B$ as in (2.10).

3. MODEL ASSESSMENT WITH SIMULATED DATA

3.1 Simulated conditions

To assess the proposed models, we simulate a population and repeatedly draw two datasets from that population: a probability sample and a non-probability one. Then, we estimate the EMSE for Z and S of both samples through both models to obtain the weight that the estimator for each sample should have, and we combine these estimators.

We generate a population consisting of $N = 100{,}000$ units from which we will generate an outcome variable that will define the membership category. We also generate a w variable indicating selectivity, which will be used to manipulate the levels of selectivity in the non-probability sample. We consider 1, 4, 10, and 15 domains, 3, 5, 8, and 15 categories, and a first scenario where all these categories are of equal size in each domain, and a second scenario where categories have unequal sizes. The sizes of unequal size categories with 3, 5, and 8 categories were based on the real data distribution of data of CBS for the variable educational attainment. The distribution of the categories is shown here as the cumulative proportion of the data belonging to that category. For 3 categories, the distribution is {0.29, 0.65}, i.e., the first category ranges from 0 to 0.29, the second category ranges from 0.29 to 0.65, and the third from 0.65 to 1. Then the distribution for 5 categories is {0.09,0.29,0.65,0.87}, and the distribution for 8 categories is {0.09,0.20,0.28, 0.42,0.55,0.65,0.87}. In the case of 15 categories, the data was not available with this number of categories, but it was split so it would follow a similar distribution to the previous ones, with a distribution of {0.03,0.06,0.11,0.18,0.27,0.37,0.49,0.58,0.68,0.77,0.85,0.93,0.95, 0.97}.

We also simulate scenarios with sample size per domain for a probability sample of $n_P \in \{10, 100, 400, 900\}$. For the non-probability samples, sample sizes per domain are $n_{NP} \in \{100, 1000, 2000, 6000\}$, and we introduce two levels of selectivity.

For each unit, a continuous outcome variable y is generated that follows a linear relationship with the independent variables, with the residuals drawn from a normal distribution. We use two independent variables. The total variance of y is set to be 80, of which two-thirds are explained by the independent variables and the remaining third is unexplained variance due to the residuals. The continuous variable y is later transformed into a categorical outcome variable with a different number of categories according to the above-mentioned simulation scenarios. The independent variables are drawn independently from each other from a multivariate normal distribution with mean 0 and standard deviation 1, but they might be correlated with the selectivity variable w. This correlation of w with the independent variables specifies the two levels of selectivity, weak and severe: the correlation is 0.228 when the selectivity is weak, and 0.632 when it is severe.

The probability sample will be drawn with equal inclusion probabilities using simple random sampling, whereas the non-probability sample will be drawn with unequal probabilities using randomized systematic sampling, following the approach in Smit (2021), where the probability of inclusion is proportional to λ:

$$\lambda_i = \frac{1}{1 + \exp\left(-\left(2 * (w_i - 0.75)\right)\right)} \tag{3.1}$$

We consider a full factorial design, giving rise to 1024 scenarios of different simulation conditions, and draw 1000 simulations for each of them.

After drawing the samples, we obtain a cross-table with categories in the columns and domains on the rows as in Table 1 for each sample, and we compute weights with equation (2.6) to obtain the final table of the combined estimators. We carry out these steps for model A and model B separately.

We also considered another simulation approach in which 10% of the variance was shared within the domain and therefore domain-specific, and the other 90% was due to a random error contributed by each observation independently. This consideration was simulated for modelling approaches A and B, but its results did not show any significant differences compared to the results reported below. Therefore, the results are not reported in this thesis. The scripts for this alternative approach are included in the research repository.

3.2 Evaluation

We evaluate the following performance measures: the root mean squared error and the bias.

First, the root mean squared error will be used where R is the number of simulations and r denotes a specific simulation round ($r = 1, \dots, R$)

$$RMSE_{kc} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(Z_{kc} - \widehat{D}_{kc,r})^2} \tag{3.2}$$

Here $\widehat{D}_{kc,r}$ may also be an estimator based on one of the samples only instead of a combined estimator.

We combine the $RMSE_{kc}$ into one performance measure per domain $ARMSE_k$ defined by

$$ARMSE_k = \frac{1}{C}\sum_{c=1}^{C} RMSE_{kc} \tag{3.3}$$

And we combine the $ARMSE_k$ into one overall performance measure by taking the mean of the values by domain

$$MARMSE = \frac{1}{K}\sum_{k=1}^{K} ARMSE_k \tag{3.4}$$

Second, the bias is assessed as the mean of absolute bias (MAB)

$$MAB = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{KC}\sum_{k=1}^{K}\sum_{c=1}^{C}\left|Z_{kc} - \widehat{D}_{kc,r}\right|\right) \tag{3.5}$$

3.3 MARMSE

The first interest here is to answer the question of whether it is possible to decrease the mean squared error of an estimator by combining two types of samples. In Table 2, this is presented as the proportion of times, out of a total of 256 simulation conditions different in the number of domains, the number of categories, and sample sizes, that the combined estimator shows a lower MARMSE than a probability sample (PS), a non-probability sample (NPS), and both samples (Both), for the cases of different selectivity levels, different category sizes, and different models.

Table 2. Proportion of combined estimators with lower MARMSE

Selectivity	Size of categories	Model	PS	NPS	Both
Weak	Equal	A	0.75	0.67	0.42
		B	0.84	0.66	0.50
	Unequal	A	0.8	0.67	0.47
		B	0.83	0.66	0.50
Severe	Equal	A	0.36	0.83	0.19
		B	0.54	0.86	0.40
	Unequal	A	0.38	0.84	0.22
		B	0.48	0.86	0.34

In the first place, we observe the level of selectivity. The proportion of times that a combined estimator performs better than an estimator from a non-probability sample is about 0.66 in the case of a weak selectivity and about 0.85 in the case of a severe selectivity. As expected, when a high selectivity in the non-probability sample has been introduced this results in a large bias, and the combined estimator is pulled toward the probability sample by assigning it more weight and having a less biased estimate as result (See section 3.4). We also observe that the combined estimator performs better than the estimator from the probability sample in a very high proportion of cases when there is a weak selectivity in the non-probability sample. The explanation is that if the non-probability sample is large enough and not very selective, the weight of the estimator based on the non-probability sample in the combined estimator will increase. The opposite can be observed in the case of severe selectivity where the combined estimator performs better than the probability sample in a low proportion of cases (on average 0.44 of the time) because the selective non-probability sample is not contributing to decreasing the MARMSE.

In the second place, category distribution can be equal or unequal. For every situation in which the selectivity and model factors remain constant, the unequal category distribution only seems to lead to an improvement in the case of model A, for which with weak selectivity the proportion of cases in which MARMSE of the combined estimator is lower than both of the samples increases with 0.05, and with 0.03 in the case of severe selectivity.

In the third place, we have models A and B. Model A assumes that on average the bias per category will be 0 for all categories, and Model B assumes that the bias per category will be a constant number across domains, calculated as the average difference between the non-probability and the probability samples for that category. Comparing the proportion of cases in which the combined estimator has lower MARMSE than both samples, model B performs better than model A in all scenarios, indicating that considering that the bias is a constant number instead of 0 is a more accurate assumption.

The reduction in MARSME that the combined estimator implies in each scenario can be observed in Table 3, which contains the average result of subtracting the MARMSE of the probability sample from that of the combined estimator in column four, and from subtracting the MARMSE of the non-probability sample in column five.

Table 3. Average difference between MARMSE of the combined estimator (C) and the probability (PS) and non-probability (NPS) samples

Selectivity	Size of categories	Model	Difference C-PS	Difference C-NPS
Weak	Equal	A	-0.016	0.036
		B	-0.015	-0.006
	Unequal	A	-0.015	-0.006
		B	-0.014	-0.005
Severe	Equal	A	-0.006	-0.054
		B	-0.007	-0.055
	Unequal	A	-0.005	-0.054
		B	-0.006	-0.055

We observe for scenarios with a weak selectivity, that the reduction is larger for the probability sample than for the non-probability sample; the MARMSE of the combined estimator is on average 0.015 lower than the MARMSE of the probability sample. On the contrary, in the scenarios with a severe selectivity, the reduction of the MARMSE of the combined estimator compared to the non-probability sample is on average 0.055, substantially larger than the reduction compared to the probability sample.

Note that for independent samples where at least one of the estimators is unbiased, which is the case here, the MSE (and therefore also the MARMSE) of the combined estimator is by construction always lower than at least one of its constituent estimators[1].

All tables with the results of the estimated MARMSE for different combinations of selectivity, size of categories, and model are displayed in Appendix B. Each table contains 256 different scenarios where the estimated MARMSE is shown. Here we examine 4 of these tables. The first row on the tables indicates the number of categories, the first column indicates the number of domains, the last row indicates the sample size per domain of the non-probability sample, and the last column indicates the sample size per domain of the probability sample. Each cell is the value of the MARMSE of the combined estimator for such conditions, a cell colored with the lightest grey represents a situation where the MARMSE of the combined estimator is only lower than the one from the probability sample. A cell colored with medium grey represents a situation where the MARMSE of the combined estimator is only lower than the one from the non-probability sample. And a cell colored with dark grey represents a situation where the MARMSE of the combined estimator is lower than the MARMSE from both the probability and the non-probability sample. As noted in the previous paragraph, there was no situation in which the MARMSE from the combined estimator was higher than both estimators from the two samples.

As mentioned previously, the four situations where selectivity is weak are where models A and B perform better than the estimators based on the separate samples regardless of whether the size of the categories was equal or unequal (see tables B1-B4 in Appendix B). In the case of model A, the improvement of the combined estimator is concentrated where the number of categories is 5, 8, or 15, whereas for model B this is equally distributed across the different number of categories. We explain this because of the assumption we made for model A that the average bias per category is 0 (see formula (2.3)). If the distance of each bias value to 0 is asymmetric, which is more likely to happen when we have fewer categories, this model is not capturing the bias correctly and the combined estimator would not benefit as much from the estimator for the non-probability sample. On the other hand, this is not the case for model B because it does not rely on formula (2.3).

For the four situations of weak selectivity, it holds that when the probability sample size is small, here 10, the combined estimator has a lower MARMSE than only the probability sample, and it never has a lower MARMSE than both of the samples. This is because if the non-probability sample is not too biased, this is already good enough and the contribution of the probability sample is not improving the estimation.

[1] We do not prove this result here. It can be easily proved with simple algebra.

Differences regarding the equal and unequal distribution of categories have already been mentioned and there is no clear pattern of where having an unequal distribution would lead to a -small- improvement.

Differences between domains can be observed by comparing the situation of only one domain ($K = 1$) to situations where the number of domains is bigger than one. For model A the pattern of color down the table stays the same whereas for model B the pattern is different from having one domain, compared to having more than one. This can be explained because model B shares information across domains, as can be observed in formula (2.16) that an average between domains is computed as an estimate for β_c. Consequently, working with one domain only would not provide enough information to significantly improve the estimators. Tables 4 and 5 are shown here to illustrate this.

Table 4. MARMSE of equal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.08	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	10
	0.04	0.05	0.05	0.05	0.03	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	0.03	0.03	0.02	0.02	10
	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.1	0.1	0.09	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.02	0.02	10
	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.1	0.1	0.1	0.08	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	10
	0.05	0.05	0.05	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table 5. MARMSE of equal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.1	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.06	0.05	0.05	0.05	10
	0.04	0.05	0.05	0.05	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03	10
	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.1	0.1	0.09	0.07	0.07	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	10
	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.1	0.1	0.1	0.08	0.07	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	10
	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

From the situations with a weak selectivity in the non-probability sample, we can say that model A would be more suitable if we have: a probability sample of a sample size of at least 100 observations, five or more categories, and one domain. In the case that we have fewer categories than five and two or more domains, model B would be more suitable.

In the situation of a severe level of selectivity in the non-probability sample, the tables with results are shown in Appendix B. We use tables 6 and 7 to illustrate the pattern of the equal size distribution of models A and B.

Table 6 shows the lowest proportion of cases (0.19) with better performance of the combined estimator with respect to the MARMSE. We can observe that on the few occasions that this is the case, it occurs mostly with a large non-probability sample size, and the improvements are concentrated on the number of categories 5 and 15 for all number of domains. Table 7 shows model B in equal conditions, the improvements are distributed more similarly when the number of domains is more than 1 and improves especially when the size of categories is 8 or 15.

Table 6. MARMSE of equal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.11	0.15	0.15	0.15	0.1	0.11	0.11	0.11	0.08	0.08	0.08	0.08	0.06	0.05	0.05	0.05	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.11	0.11	0.11	0.11	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.13	0.11	0.11	0.11	0.09	0.08	0.08	0.08	0.06	0.05	0.05	0.05	0.03	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.08	0.11	0.11	0.11	0.05	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.02	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table 7. MARMSE of equal size categories and severe selectivity. Model B.

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.12	0.15	0.15	0.15	0.11	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.06	0.06	0.06	0.06	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.14	0.14	0.14	0.14	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.14	0.14	0.14	0.12	0.1	0.1	0.1	0.08	0.07	0.07	0.07	0.06	0.05	0.04	0.04	0.03	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.14	0.14	0.14	0.08	0.1	0.1	0.1	0.06	0.07	0.07	0.07	0.04	0.05	0.04	0.04	0.02	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Contrary to the four situations with a weak selectivity, here it is possible to find a situation where the MARMSE of the combined estimator is lower than both of the samples when the sample size of the probability sample is 10. This is due to the poor samples shown in this situation, for the non-probability sample, this is explained due to the severe selectivity introduced, and the probability sample having a too-small sample size. Nevertheless, if both samples are not contributing with relatively decent estimators, the combined estimator would hardly be a good estimator, even with a lower MARMSE.

This pattern is also similar to the situations of unequal size categories and severe selectivity for models A and B, which also have a low proportion of improvement in both samples (See Appendices B7 and B8).

We can see from these tables (and from the ones shown in Appendices B, C, and D), first, that overall, when the size of the probability sample is large enough, using only the estimate from the probability sample is best. Second, when the non-probability sample has a severe selectivity level using the estimator from the probability sample only is also best. Third, when the non-probability sample is not too biased using the estimator from the non-probability sample only is the best. And fourth, in any other cases using the combined estimator will be the best.

3.4 Bias

In table 8 we can observe the proportion of times in which the mean absolute bias of the combined estimator was equal to or lower than the mean absolute bias of the estimator from the non-probability sample only. These proportions are high for the situation of weak selectivity and very high for the case of severe selectivity.

Table 8. Proportion of combined estimator with lower bias

Selectivity	Size of categories	Model	Proportion
Weak	Equal	A	0.78
		B	0.73
	Unequal	A	0.71
		B	0.89
Severe	Equal	A	0.96
		B	0.94
	Unequal	A	0.94
		B	0.94

Regarding the bias of the estimators, the bias of the combined estimator is mostly higher than the bias of the estimator for the non-probability sample in the case when the sample size of the probability sample is 10, this occurs in a 0.77 proportion of the times where the scenario includes this sample size (see Appendix E).

Table 9 is shown here to illustrate this pattern, which is followed also by all the other tables in Appendix E. This table should be read as previous tables on the margins, and, in this case, each cell corresponds to the difference of the mean absolute bias of the non-probability sample minus the mean absolute bias of the combined estimator, where a negative number represents that the bias of the combined estimator is higher than the bias from the non-probability sample, in which case the number is highlighted in bold. The numbers have been rounded to the second digit, so, when the difference was 0 until the second decimal and the third one was lower than 0,005, this is shown here just like 0.

The four situations of weak selectivity show a lower proportion of times in which the bias of the combined estimator is lower than the one from the non-probability sample, this makes sense because if there is a large enough sample, not highly biased, then a probability sample with a small sample size as here is 10, would not contribute to reducing the estimation error, assigning it less weight and keeping instead the estimator of the non-probability sample very close to how it was originally. On the contrary for the four opposite situations, this is, of severe selectivity, in a very high proportion of times (almost 1 for all 4) the combined estimator shows less bias, this is because if the non-probability sample is highly biased, any bit of unbiasedness even with small sample size, could be a useful contribution to help to improve the estimator.

Table 9 . Difference of mean absolute bias between the non-probability and the combined estimator of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	**-0.22**	**-0.2**	**-0.2**	**-0.2**	**-0.04**	**-0.08**	**-0.08**	**-0.08**	0.01	0	0	0	0	0	**-0.01**	**-0.01**	10
	0.03	0.06	0.06	0.05	0.02	0.02	0	0.01	0	0	0	0	0.01	0	0	0	100
	0.01	0.07	0.05	0.05	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.05	0.05	0.06	0.06	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.02	0.01	0.01	0.01	0.01	900
K = 4	**-0.07**	**-0.06**	**-0.06**	**-0.06**	0	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	10
	0.05	0.03	0.02	0.02	0.02	0	0	0	0.01	0.01	0	0.01	0	0	0	0	100
	0.07	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	0	400
	0.07	0.05	0.05	0.05	0.02	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.01	0.01	0.01	0.01	900
K = 10	**-0.04**	**-0.04**	**-0.04**	**-0.05**	**-0.01**	**-0.02**	**-0.02**	**-0.02**	**-0.01**	0	**-0.01**	**-0.01**	0	0	0	0	10
	0.03	0.02	0.02	0.01	0.01	0	0	**-0.01**	0.01	0	0	0	0.01	0	0	0	100
	0.05	0.05	0.04	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0	0.01	0	0	0	400
	0.04	0.05	0.04	0.03	0.05	0.03	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	900
K = 15	**-0.03**	**-0.03**	**-0.03**	**-0.04**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0	**-0.01**	**-0.01**	**-0.01**	0	0	0	**-0.01**	100
	0.05	0.03	0.04	0	0.01	0.01	0.01	**-0.01**	0.01	0	0	0	0.01	0	0	0	1000
	0.05	0.05	0.04	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	0.01	0.01	0	0	4000
	0.05	0.05	0.05	0	0.04	0.03	0.03	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

The mean absolute bias of the combined estimator itself is shown in table 10. We observe that the bias for situations when the number of categories is 3 is substantially higher than in the other scenarios. Most of the time, the bias decreases as the number of categories increases, becoming a stable pattern when the sample size of the probability sample is not 10.

If we compare tables 9 and 10, we can see for the first row and column, that the combined estimator has a bias of 0.27 while the bias from the non-probability sample only was 0.05. High biases (>0.10) always decrease as the sample size of both samples increases.

Table 10. Mean absolute bias of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.27	0.25	0.26	0.26	0.08	0.11	0.11	0.11	0.03	0.03	0.03	0.02	0.01	0.02	0.02	0.02	10
	0.04	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.12	0.12	0.12	0.12	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.03	0.02	0.02	0.02	10
	0.03	0.03	0.03	0.03	0.04	0.03	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.11	0.1	0.1	0.09	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.01	0.01	10
	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
K = 15	0.09	0.09	0.09	0.06	0.05	0.05	0.05	0.02	0.04	0.03	0.03	0.01	0.02	0.02	0.02	0.01	10
	0.02	0.03	0.03	0.01	0.03	0.03	0.03	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0	100
	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

4. APPLICATION TO REAL DATA

4.1 Data

The Educational Attainment File (EAF) provided by CBS combines data from the Labour Force Survey and various administrative records on the educational level of people. The target population consists of all the people who were registered in the Municipal Personal Records Database on the 1st of October 2019.

The Labour Force Survey (LFS) is a rotating panel survey with five waves, with a target population of people 15 years or older who live in the Netherlands. The EAF contains LFS data from several years, which have been integrated with other available information from administrative records (Linder et al., 2014).

The administrative records are non-probability samples including the following: Files with educational histories as submitted by job seekers at the UWV Work company, CRIHO (Central Register of Registrations in Higher Education), Education number files from secondary education, Education number files from primary education (WPO) and special education (WEC). These records also compile the measurement of educational attainment and contain 11,092,584 observations.

The variables used to create the domains and categories are Municipality, and Highest-level education achieved respectively. The variable Municipality has 355 categories. For our study, we select 11 municipalities of different sizes and use those as our domains. After the selection, we ended up with 17,193 observations from the probability sample, and 623,114 from the non-probability sample. Highest-level education has originally 18 categories of unequal size per domain, which we have recoded into 3 categories namely low, middle, and high education.

For its regular output based on the Educational Attainment File, CBS computes estimates by weighting the LFS data to represent the subset of the population that is not covered by the administrative records (Linder et al., 2014). Here, we compare these regular estimates to our proposed combined estimator. For this study, the administrative records in the EAF are treated as a non-probability sample and $\hat{Z}_{kc,NP}$ is the estimate based only on these records. As a probability sample for the estimate $\hat{Z}_{kc,P}$, we use a separate data set consisting of LFS data from 2016. We focus on the target population of persons 15 years or older.

4.2 Results

We have chosen to apply model B because we observe a variable with 3 categories, 11 domains, and large enough sample sizes, in which case the results in Section 3 have already indicated that model B performs better.

In table 11 we observe the regular CBS estimates, the combined estimator, estimators from the probability and non-probability sample, and the respective sample sizes for each sample.

Table 11. Estimates of educational attainment per city

	CBS estimates			Combined estimator			Probability sample(1)			Non-probability sample(2)			n(1)	n(2)
	1	2	3	1	2	3	1	2	3	1	2	3		
Amsterdam	0.24	0.29	0.46	0.23	0.29	0.49	0.23	0.29	0.49	0.23	0.3	0.46	11051	473324
Amstelveen	0.23	0.33	0.44	0.2	0.28	0.52	0.2	0.28	0.52	0.2	0.34	0.45	1365	37815
Krimpenerwaard	0.34	0.43	0.23	0.34	0.42	0.23	0.35	0.42	0.23	0.29	0.44	0.27	1128	23925
Medemblik	0.35	0.45	0.2	0.34	0.41	0.23	0.35	0.41	0.24	0.34	0.43	0.23	845	19961
Teylingen	0.26	0.42	0.33	0.23	0.39	0.37	0.22	0.41	0.37	0.25	0.39	0.37	684	16990
Boxtel	0.34	0.4	0.26	0.36	0.39	0.25	0.37	0.39	0.24	0.31	0.4	0.3	683	14174
Wassenaar	0.26	0.33	0.41	0.23	0.51	0.26	0.23	0.52	0.25	0.25	0.33	0.42	336	10177
Sluis	0.35	0.46	0.19	0.4	0.42	0.18	0.41	0.41	0.18	0.32	0.45	0.23	448	8822
West Maas en Waal	0.33	0.46	0.21	0.44	0.4	0.16	0.45	0.4	0.16	0.3	0.45	0.26	329	8900
Ouder-Amstel	0.23	0.34	0.43	0.26	0.31	0.43	0.28	0.3	0.42	0.21	0.33	0.45	264	6552
Terschelling	0.27	0.51	0.22	0.22	0.43	0.35	0.21	0.39	0.39	0.25	0.49	0.25	60	2474

We observe that estimates of the combined estimator are, overall, close to the regular CBS estimates, which we consider the true estimates. Important absolute differences (larger than 0.10) are observed in 5 out of the 33 estimates in total. This is the case for categories 2 and 3 in Wassenaar, category 1 in West Maas en Waal, and category 3 in Terschelling. The estimates are not particularly affected by the sample size, since of these 3 cities, only Terschelling has a sample size that we would consider small (n=60), and two other cities on the table have similar sample sizes and they do not seem to be affected by it. Nevertheless, if we observe how different are the probability and non-probability samples compared to the CBS estimates, we see that the biggest differences in the combined estimator occur when the estimates from the probability sample are also very different from the regular CBS estimates. This means that in these situations, the combined estimator is being computed with a very high weight for the probability sample because the method is estimating a large bias in the non-probability sample.

The differences between the estimates from the probability sample and the regular CBS estimates can be explained also by three other aspects: first by the different procedures of estimation, second, by differences in the data processing and definitions used, and third, because our target population is from 2019 and we are using a probability sample from 2016, which is still a reasonable approximation because the distribution of educational attainment changes slowly over the years.

Because these differences could have deviated the combined estimator from the estimates from CBS, we have also simulated two scenarios where we first draw a new simple random sample with the same sample size as the already mentioned probability sample, and one with a smaller sample size, directly from the observed distribution in the EAF. Then, we calculate the combined estimator again with each of the new samples as we can see in table 12 with its respective sample sizes.

Table 12. Combined estimator with simulated probability samples of educational attainment per city

	CBS estimates			Combined estimator with same sample size (1)			Combined estimator with smaller sample size (2)			n(1)	n(2)
	1	2	3	1	2	3	1	2	3		
Amsterdam	0.24	0.29	0.46	0.24	0.29	0.46	0.26	0.28	0.47	11051	1105
Amstelveen	0.23	0.33	0.44	0.24	0.33	0.43	0.20	0.34	0.45	1365	136
Krimpenerwaard	0.34	0.43	0.23	0.34	0.42	0.23	0.38	0.40	0.22	1128	112
Medemblik	0.35	0.45	0.2	0.34	0.42	0.23	0.37	0.44	0.18	845	84
Teylingen	0.26	0.42	0.33	0.26	0.44	0.30	0.25	0.39	0.37	684	68
Boxtel	0.34	0.4	0.26	0.31	0.45	0.24	0.31	0.40	0.30	683	68
Wassenaar	0.26	0.33	0.41	0.26	0.34	0.42	0.31	0.26	0.44	336	33
Sluis	0.35	0.46	0.19	0.34	0.49	0.17	0.32	0.50	0.16	448	44
West Maas en Waal	0.33	0.46	0.21	0.29	0.44	0.26	0.30	0.45	0.28	329	32
Ouder-Amstel	0.23	0.34	0.43	0.20	0.34	0.45	0.21	0.33	0.45	264	26
Terschelling	0.27	0.51	0.22	0.27	0.49	0.25	0.25	0.49	0.25	60	6

Here we observe that the difference between the regular estimates from CBS with the ones from the combined estimator computed with a probability sample of the same size is drastically reduced, with no absolute difference larger than 0.05. And, even for a smaller sample size, reduced to a tenth of the original, estimators are still very close with barely two estimators with an absolute difference larger than 0.05.

5. DISCUSSION

In this study, we have evaluated a way to combine probability and non-probability samples with the aim of reducing the mean squared error of an estimator of proportions. Using a small area estimation framework, we proposed two different methods to estimate the bias of the non-probability sample to obtain its expected mean squared error and combine it with the expected mean squared error from a relatively small probability sample, aiming to obtain a combined estimator with a lower mean squared error.

Through simulation studies, we have shown that the combined estimator can lead to a reduction of the MSE, which we evaluate through the MARMSE of each simulated scenario. This evaluation indicates that using model A or model B to compute the weights of the combined estimator is better than not doing it when we have a relatively small probability sample and not a too biased estimator for the non-probability sample, which is concordant with our initial expectations. It is also stated that either model A or B could be more useful depending on the situation; if we have only 1 domain, a large probability sample ($\geq$100), and a few categories model A would be best. In any other situation model B would be better than model A.

These findings are supported by an empirical application of model B to a real dataset and simulation based on that dataset using regular CBS estimates as the true values, where we have shown that with all the combined estimators calculated, we can obtain close estimates to those that have been obtained with methods already tested and validated and are currently in usage for this purpose. The application also underlined that our method relies on the assumption that the estimator based on the probability sample is unbiased.

Overall, the proposed methods have the advantages that it is not necessary to link the two samples at the level of individual observations. They are robust, meaning that they will never yield a worse result, i.e., higher MSE, than the highest MSE that you could obtain from one of the two samples separately. And it is very easy to implement in software like R since it requires very few lines of code.

Some considerations we should have regarding this study:

First, this is a method for categorical variables. The same approach could be extended to other types of variables, especially if it is intended to be used beyond the field of official statistics, and several continuous variables could be of interest, but this would require a different model for bias in the non-probability sample. Second, one important assumption to consider is that the probability sample is unbiased. We assume so during the study, but the same problems that probability samples have been facing lately already pointed out in the introduction could mean that this assumption might not be true. The proposed approaches would need to be adapted if this assumption does not hold. Third, the probability sample still needs to be weighted or adjusted, as it is usually done to reproduce the structure of the target population. In future research, the proposed approaches could be adapted to take the survey weights into account. Fourth, this study was done for a simple design. Note that, unless the probability sample happens to be stratified by domain, $\hat{Z}_{kc,P}$ is a ratio estimator and in general, its variance would have to be computed by a more complicated formula than the one provided here.

A last remark on the major challenge of this study and probably of many other studies is how to deal with the selectivity in the non-probability samples. The models proposed here could and should be considered when searching for other ways to model the bias of an estimator based on a non-probability sample, The same models could be re-examined under different simulation conditions to study the properties of these models in more detail.

Finally, we could think that we are currently in a transitional period before we can start seeing less and less of the classic survey methodology approach. Even for real data used here, it is expected to dispense the probability sample in the future. But for now, there will be a long period in which both will have to coexist before we completely tame non-probability samples enough for them to outweigh the benefits of probability samples, therefore paramount efforts must be done in that direction.

6. SUPPORTING INFORMATION

The research repository of this study can be found in the following Github repository: https://github.com/svillalobosaliste/master_thesis. The data are available in the secure environment at CBS. To get access to the data one needs to contact CBS. The simulation study and use of data are approved by the Ethical Review Board of the Faculty of Social and Behavioural Sciences of Utrecht University. The approval is based on the documents sent by the researchers as requested in the form of the Ethics committee and filed under numbers 21-2067 and 22-1257 respectively.

The used software was R version 4.1.2.

7. APPENDICES

7.1 Appendix A: Modelling approaches

7.1.1 Model A

For completeness's sake, we start this appendix by introducing the notation, even though the notation is also defined in the main report itself.

Suppose there is a target population of N units ($i = 1, \dots, N$) with a categorical target variable y with $C \geq 2$ categories – i.e., $y_i \in \{1, \dots, C\}$ for each unit i – and suppose that observations on y are available from two samples.

Furthermore, suppose that the units in the target population are divided into K domains ($k = 1, \dots, K$), where domain k contains N_k units.

The first sample $\mathcal{S}_P$ is a probability sample of size $n_P = \sum_{k=1}^{K} n_{Pk}$. To keep matters simple we assume that $\mathcal{S}_P$ is a simple random sample and that $n_{Pk} \ll N_k$ for all domains. The second sample $\mathcal{S}_{NP}$ is a non-probability sample of size $n_{NP} = \sum_{k=1}^{K} n_{NPk}$. We are especially interested in situations where $n_{NP} \gg n_P$.

Let δ_{kci} denote a 0-1-indicator that equals 1 if and only if unit i belongs to domain k and has $y_i = c$. Suppose that we are interested in estimating $Z_{kc} = \frac{1}{N_k} \sum_{i=1}^{N} \delta_{kci}$, the true proportion of units in domain k of the target population that belong to category c (for $k = 1, \dots, K$, and $c = 1, \dots, C$).

To estimate Z_{kc} from the probability sample, we can use the sample domain mean:

$$\hat{Z}_{kc,P} = \frac{1}{n_{Pk}} \sum_{i \in \mathcal{S}_P} \delta_{kci}. \tag{1}$$

This estimator is unbiased. Assuming that the number of categories C is large and each category contains only a limited proportion of units within each domain, the design-based variance (and mean squared error) of $\hat{Z}_{kc,P}$ should be approximated well by

$$MSE_d(\hat{Z}_{k,P}) = V_d(\hat{Z}_{k,P}) = \frac{1}{n_{Pk}} Z_{kc}(1 - Z_{kc}) \tag{2}$$

where the subscript d indicates that the bias, variance and mean squared error are calculated under the known sampling design for the probability sample and the unknown sampling "design" for the non-probability sample.

Similarly, to estimate Z_{kc} from the non-probability sample, we could use the sample domain mean (in the absence of more information about the sample design):

$$\hat{Z}_{kc,NP} = \frac{1}{n_{NPk}} \sum_{i \in \mathcal{S}_{NP}} \delta_{kci} \tag{3}$$

This estimator is likely to be biased, in the sense that $E_d(\hat{Z}_{kc,NP}) \neq Z_{kc}$. It will be useful below to introduce the following short-hand notation: $\tilde{Z}_{kc} = E_d(\hat{Z}_{kc,NP})$. Moreover, we will use

$$b_{kc} = E_d(\hat{Z}_{kc,NP}) - Z_{kc} = \tilde{Z}_{kc} - Z_{kc} \tag{4}$$

to denote the bias of $\hat{Z}_{kc,NP}$ as an estimator of Z_{kc} under the (unknown) sampling design of $\mathcal{S}_{NP}$.

Technically, the design-based variance of $\hat{Z}_{kc,NP}$ is unknown/undefined. Assuming that the sample is large, a reasonable variance approximation might be

$$V_d(\hat{Z}_{kc,NP}) = \frac{1}{n_{NPk}} \tilde{Z}_{kc}(1 - \tilde{Z}_{kc}) \tag{5}$$

The mean squared error of $\hat{Z}_{kc,NP}$ is then equal to

$$MSE_d(\hat{Z}_{kc,NP}) = b_{kc}^2 + \frac{\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})}{n_{NPk}} \tag{6}$$

The bias b_{kc} is unknown in practice. However, we observe that, within each domain,

$$\sum_{c=1}^{C} b_{kc} = \sum_{c=1}^{C} E_d(\hat{Z}_{kc,NP} - Z_{kc}) = E_d\left(\sum_{c=1}^{C} \hat{Z}_{kc,NP} - \sum_{c=1}^{C} Z_{kc}\right) = E_d(1-1) = 0$$

Thus, the average value of b_{kc} across all categories must be zero for each domain. Following Pannekoek & De Waal (1998), we could therefore assume a model such that b_{kc} is distributed as a random variable with mean $E_b(b_{kc}) = 0$ and variance $E_b(b_{kc}^2) = \sigma_k^2$.

Here the subscript b indicates that the bias, variance and mean squared error are calculated under the above model for the bias b_{kc}. We will also make the technical assumption that $E_b(\tilde{Z}_{kc} b_{kc}) = 0$, i.e., that b_{kc} is not correlated to $\tilde{Z}_{kc}$. Then, using expressions (2) and (6), the model-based expected mean squared errors of $\hat{Z}_{kc,P}$ and $\hat{Z}_{kc,NP}$ equal:

$$EMSE(\hat{Z}_{kc,P}) = \frac{E_b[Z_{kc}(1 - Z_{kc})]}{n_{Pk}} \quad (7)$$

and

$$EMSE(\hat{Z}_{kc,NP}) = E_b\left[b_{kc}^2 + \frac{\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})}{n_{NPk}}\right] = \sigma_k^2 + \frac{E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})]}{n_{NPk}} \quad (8)$$

To elaborate on these expressions, we first note that, by the definition of b_{kc} in (4),

$$E_b(Z_{kc}) = E_b(\tilde{Z}_{kc} - b_{kc}) = E_b(\tilde{Z}_{kc})$$

and

$$\begin{aligned} E_b(Z_{kc}^2) &= E_b\left[(\tilde{Z}_{kc} - b_{kc})^2\right] \\ &= E_b(\tilde{Z}_{kc}^2) + E_b(b_{kc}^2) - 2E_b(\tilde{Z}_{kc} b_{kc}) \\ &= E_b(\tilde{Z}_{kc}^2) + \sigma_k^2 \end{aligned}$$

where we used the assumption that b_{kc} is not correlated to $\tilde{Z}_{kc}$. It follows that

$$E_b[Z_{kc}(1 - Z_{kc})] = E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})] - \sigma_k^2 \quad (9)$$

Furthermore, recalling that $\tilde{Z}_{kc} = E_d(\hat{Z}_{kc,NP})$ and that the variance of any random variable X is defined as $V(X) = E(X^2) - [E(X)]^2$, we obtain:

$$\begin{aligned} E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})] &= E_b\left\{E_d(\hat{Z}_{kc,NP}) - [E_d(\hat{Z}_{kc,NP})]^2\right\} \\ &= E_b[E_d(\hat{Z}_{kc,NP}) - E_d(\hat{Z}_{kc,NP}^2) + V_d(\hat{Z}_{kc,NP})] \\ &= E_b\{E_d[\hat{Z}_{kc,NP}(1 - \hat{Z}_{kc,NP})] + V_d(\hat{Z}_{kc,NP})\} \\ &= v_{kc} + \frac{E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})]}{n_{NPk}} \end{aligned}$$

where $v_{kc} = E_b E_d[\hat{Z}_{kc,NP}(1 - \hat{Z}_{kc,NP})]$ and we used expression (5). We conclude that

$$\left(1 - \frac{1}{n_{NPk}}\right) E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})] = v_{kc}$$

i.e.,

$$\frac{E_b[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})]}{n_{NPk}} = \frac{v_{kc}}{n_{NPk} - 1} \quad (10)$$

Substituting this result into expression (8), we obtain:

$$EMSE(\hat{Z}_{kc,NP}) = \sigma_k^2 + \frac{v_{kc}}{n_{NPk} - 1} \tag{11}$$

Similarly, for the estimator based on the probability sample, we obtain from expressions (7), (9), and (10) that

$$EMSE(\hat{Z}_{kc,P}) = \frac{E_b\left[\tilde{Z}_{kc}(1 - \tilde{Z}_{kc})\right] - \sigma_k^2}{n_{Pk}} = \frac{1}{n_{Pk}}\left(\frac{n_{NPk}}{n_{NPk} - 1} v_{kc} - \sigma_k^2\right) \tag{12}$$

For large non-probability samples, in (12) the factor $n_{NPk}/(n_{NPk} - 1) \approx 1$ could be ignored. This completes the derivation of expressions (2.4) and (2.5) in the report.

Finally, we show how to derive expression (2.9) in the report as an estimator for σ_k^2. Under the assumptions made so far it holds that:

$$\begin{aligned}
E_b E_d\left[\sum_{c=1}^{C}\left(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}\right)^2\right] &= \sum_{c=1}^{C} E_b E_d\left[\left(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}\right)^2\right] \\
&= \sum_{c=1}^{C} E_b E_d\left[\left(\hat{Z}_{kc,NP} - Z_{kc} + Z_{kc} - \hat{Z}_{kc,P}\right)^2\right] \\
&= \sum_{c=1}^{C}\left\{E_b E_d\left[\left(\hat{Z}_{kc,NP} - Z_{kc}\right)^2\right] + E_b E_d\left[\left(\hat{Z}_{kc,P} - Z_{kc}\right)^2\right]\right. \\
&\quad \left. - 2E_b E_d\left[\left(\hat{Z}_{kc,NP} - Z_{kc}\right)\left(\hat{Z}_{kc,P} - Z_{kc}\right)\right]\right\} \\
&= \sum_{c=1}^{C}\left[EMSE(\hat{Z}_{kc,NP}) + EMSE(\hat{Z}_{kc,P})\right. \\
&\quad \left. - 2E_b(b_{kc}) E_b E_d\left(\hat{Z}_{kc,P} - Z_{kc}\right)\right] \\
&= C\left(1 - \frac{1}{n_{Pk}}\right)\sigma_k^2 + \frac{1 + n_{NPk}/n_{Pk}}{n_{NPk} - 1}\sum_{c=1}^{C} v_{kc}
\end{aligned}$$

For the penultimate equality, we used the assumption that $\hat{Z}_{kc,NP}$ and $\hat{Z}_{kc,P}$ are independent, which is true if the two samples were drawn independently of each other. In the last line, we used that $E_b(b_{kc}) = 0$ and we used expressions (11) and (12).

Hence, a consistent estimator of σ_k^2 is given by:

$$\hat{\sigma}_k^2 = \frac{n_{Pk}}{C(n_{Pk}-1)} \sum_{c=1}^{C} \left[\left(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}\right)^2 - \frac{1 + n_{NPk}/n_{Pk}}{n_{NPk} - 1} \hat{v}_{kc} \right] \quad (13)$$

with $\hat{v}_{kc}$ from expression (2.8) in the report.

7.1.2 Model B

For this case, and keeping the same notation, the design-based variance (and mean squared error) of $\hat{Z}_{kc,P}$ is approximated by

$$MSE_d\left(\hat{Z}_{kc,P}\right) = V_d\left(\hat{Z}_{kc,P}\right) = \frac{1}{n_{Pk}} Z_{kc}(1 - Z_{kc}) \quad (14)$$

Formulas (3), (4), (5), (6), and (7) remain the same.

As mentioned in Section 2.2, we now assume a model such that b_{kc} is distributed as a random variable with mean $E_b(b_{kc}) = \beta_c$ (with $\sum_{c=1}^{C} \beta_c = 0$) and variance $E_b[(b_{kc} - \beta_c)^2] = \sigma_k^2$. And we make the technical assumption that $E_b\left(\tilde{Z}_{kc} b_{kc}\right) = E_b\left(\tilde{Z}_{kc}\right) E_b(b_{kc})$, i.e., that b_{kc} is not correlated to $\tilde{Z}_{kc}$.

Then using expressions (14) and (6)

$$EMSE\left(\hat{Z}_{kc,NP}\right) = E_b\left[b_{kc}^2 + \frac{\tilde{Z}_{kc}\left(1 - \tilde{Z}_{kc}\right)}{n_{NPk}}\right] = \beta_c^2 + \sigma_k^2 + \frac{E_b\left[\tilde{Z}_{kc}\left(1 - \tilde{Z}_{kc}\right)\right]}{n_{NPk}} \quad (15)$$

To elaborate on these expressions, we first note that, by the definition of b_{kc} in (4),

$$E_b(Z_{kc}) = E_b\left(\tilde{Z}_{kc} - b_{kc}\right) = E_b\left(\tilde{Z}_{kc}\right) - \beta_c \quad (18)$$

and

$$\begin{aligned} E_b\left(Z_{kc}^2\right) &= E_b\left[\left(\tilde{Z}_{kc} - b_{kc}\right)^2\right] \\ &= E_b\left[\left(\tilde{Z}_{kc} - \beta_c + \beta_c - b_{kc}\right)^2\right] \\ &= E_b\left[\left(\tilde{Z}_{kc} - \beta_c\right)^2\right] + E_b[(\beta_c - b_{kc})^2] \\ &= E_b\left[\left(\tilde{Z}_{kc} - \beta_c\right)^2\right] + \sigma_k^2 \end{aligned} \quad (17)$$

where we used the assumption that b_{kc} is not correlated to $\tilde{Z}_{kc}$. It follows that

$$E_b[Z_{kc}(1-Z_{kc})] = E_b\left[\left(\tilde{Z}_{kc}-\beta_c\right)\left(1-\tilde{Z}_{kc}+\beta_c\right)\right]-\sigma_k^2 \tag{18}$$

Furthermore, recalling that $\tilde{Z}_{kc} = E_d\left(\hat{Z}_{kc,NP}\right)$ and that the variance of any random variable X is defined as $V(X) = E(X^2) - [E(X)]^2$, we obtain:

$$E_b\left[\left(\tilde{Z}_{kc}-\beta_c\right)\left(1-\tilde{Z}_{kc}+\beta_c\right)\right] = E_b\left[\tilde{Z}_{kc}\left(1-\tilde{Z}_{kc}\right)\right]+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2$$

$$= E_b\left\{E_d\left(\hat{Z}_{kc,NP}\right)-\left[E_d\left(\hat{Z}_{kc,NP}\right)\right]^2\right\}+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2$$

$$= E_b\left[E_d\left(\hat{Z}_{kc,NP}\right)-E_d\left(\hat{Z}_{kc,NP}^2\right)+V_d\left(\hat{Z}_{kc,NP}\right)\right]+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2$$

$$= E_b\left\{E_d\left[\hat{Z}_{kc,NP}\left(1-\hat{Z}_{kc,NP}\right)\right]+V_d\left(\hat{Z}_{kc,NP}\right)\right\}+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2$$

$$= v_{kc}+\frac{E_b\left[\tilde{Z}_{kc}\left(1-\tilde{Z}_{kc}\right)\right]}{n_{NPk}}+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2$$

where $v_{kc} = E_bE_d\left[\hat{Z}_{kc,NP}\left(1-\hat{Z}_{kc,NP}\right)\right]$ and we used expression (5). Comparing the right-hand-side of the first line with the last line, we conclude that

$$\left(1-\frac{1}{n_{NPk}}\right)E_b\left[\tilde{Z}_{kc}\left(1-\tilde{Z}_{kc}\right)\right] = v_{kc}$$

i.e.,

$$\frac{E_b\left[\tilde{Z}_{kc}\left(1-\tilde{Z}_{kc}\right)\right]}{n_{NPk}} = \frac{v_{kc}}{n_{NPk}-1} \tag{19}$$

Substituting this result into expression (8), we obtain:

$$EMSE\left(\hat{Z}_{kc,NP}\right) = \beta_c^2+\sigma_k^2+\frac{v_{kc}}{n_{NPk}-1} \tag{20}$$

Similarly, for the estimator based on the probability sample, we obtain from expressions (7), (9), and (10) that

$$EMSE\left(\hat{Z}_{kc,P}\right) = \frac{E_b\left[\left(\tilde{Z}_{kc}-\beta_c\right)\left(1-\tilde{Z}_{kc}+\beta_c\right)\right]-\sigma_k^2}{n_{Pk}}$$
$$= \frac{E_b\left[\tilde{Z}_{kc}\left(1-\tilde{Z}_{kc}\right)\right]+\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right]-\beta_c^2-\sigma_k^2}{n_{Pk}}$$

and hence

$$EMSE(\hat{Z}_{kc,P}) = \frac{1}{n_{Pk}}\left(\frac{n_{NPk}}{n_{NPk}-1}v_{kc} + \beta_c\left[2E_b(\tilde{Z}_{kc}) - 1\right] - \beta_c^2 - \sigma_k^2\right) \tag{21}$$

We now consider a combined estimator for Z_{kc} of the form

$$\hat{Z}_{kc,w} = W_{kc}\hat{Z}_{kc,P} + (1 - W_{kc})\hat{Z}_{kc,NP} \tag{22}$$

Again, following Pannekoek & De Waal (1998), the choice of a weight that leads to the optimal estimator of this form (in the sense that the expected MSE under the model is minimized) is given by:

$$W_{kc} = \frac{EMSE(\hat{Z}_{kc,NP})}{EMSE(\hat{Z}_{kc,NP}) + EMSE(\hat{Z}_{kc,P})}$$

Substituting the above results from expressions (11) and (12), we obtain:

$$\begin{aligned} W_{kc} &= \frac{\beta_c^2 + \sigma_k^2 + \dfrac{v_{kc}}{n_{NPk}-1}}{\left(1 - \dfrac{1}{n_{Pk}}\right)(\beta_c^2 + \sigma_k^2) + \dfrac{1 + n_{NPk}/n_{Pk}}{n_{NPk}-1}v_{kc} + \dfrac{1}{n_{Pk}}\beta_c\left[2E_b(\tilde{Z}_{kc}) - 1\right]} \\ &= \frac{(n_{NPk}-1)(\beta_c^2 + \sigma_k^2) + v_{kc}}{(n_{NPk}-1)\left(1 - \dfrac{1}{n_{Pk}}\right)(\beta_c^2 + \sigma_k^2) + \left(1 + \dfrac{n_{NPk}}{n_{Pk}}\right)v_{kc} + \dfrac{n_{NPk}-1}{n_{Pk}}\beta_c\left[2E_b(\tilde{Z}_{kc}) - 1\right]} \end{aligned} \tag{23}$$

with $\hat{v}_{kc}$ from expression (2.8) in the report. It only remains to find estimators for β_c and σ_k^2. First, to estimate β_c we can use

$$\hat{\beta}_c = \frac{1}{K}\sum_{k=1}^{K}(\hat{Z}_{kc,NP} - \hat{Z}_{kc,P}), \tag{24}$$

since it holds under the assumed model that

$$E_b\left[E_d(\hat{\beta}_c)\right] = \frac{1}{K}\sum_{k=1}^{K}E_b\left[E_d(\hat{Z}_{kc,NP}) - E_d(\hat{Z}_{kc,P})\right] = \frac{1}{K}\sum_{k=1}^{K}E_b(\tilde{Z}_{kc} - Z_{kc}) = \frac{1}{K}\sum_{k=1}^{K}E_b(b_{kc}) = \beta_c$$

Secondly, under the assumptions made so far it holds that:

$$
\begin{aligned}
E_b E_d\left[\sum_{c=1}^{C}\left(\hat{Z}_{kc,NP}-\hat{Z}_{kc,P}\right)^2\right] &= \sum_{c=1}^{C} E_b E_d\left[\left(\hat{Z}_{kc,NP}-\hat{Z}_{kc,P}\right)^2\right] \\
&= \sum_{c=1}^{C} E_b E_d\left[\left(\hat{Z}_{kc,NP}-Z_{kc}+Z_{kc}-\hat{Z}_{kc,P}\right)^2\right] \\
&= \sum_{c=1}^{C}\left\{E_b E_d\left[\left(\hat{Z}_{kc,NP}-Z_{kc}\right)^2\right]+E_b E_d\left[\left(\hat{Z}_{kc,P}-Z_{kc}\right)^2\right]\right. \\
&\quad \left. -2E_b E_d\left[\left(\hat{Z}_{kc,NP}-Z_{kc}\right)\left(\hat{Z}_{kc,P}-Z_{kc}\right)\right]\right\} \\
&= \sum_{c=1}^{C}\left[EMSE\left(\hat{Z}_{kc,NP}\right)+EMSE\left(\hat{Z}_{kc,P}\right)-2E_b(b_{kc})E_b E_d\left(\hat{Z}_{kc,P}-Z_{kc}\right)\right] \\
&= C\left(1-\frac{1}{n_{Pk}}\right)\sigma_k^2+\left(1-\frac{1}{n_{Pk}}\right)\sum_{c=1}^{C}\beta_c^2+\frac{1}{n_{Pk}}\sum_{c=1}^{C}\beta_c\left[2E_b\left(\tilde{Z}_{kc}\right)-1\right] \\
&\quad +\frac{1+\frac{n_{NPk}}{n_{Pk}}}{n_{NPk}-1}\sum_{c=1}^{C} v_{kc} \\
&= C\left(1-\frac{1}{n_{Pk}}\right)\sigma_k^2+\left(1-\frac{1}{n_{Pk}}\right)\sum_{c=1}^{C}\beta_c^2+\frac{2}{n_{Pk}}\sum_{c=1}^{C}\beta_c E_b\left(\tilde{Z}_{kc}\right) \\
&\quad +\frac{1+n_{NPk}/n_{Pk}}{n_{NPk}-1}\sum_{c=1}^{C} v_{kc}.
\end{aligned}
$$

For the second-to-last equality, we used the assumption that $\hat{Z}_{kc,NP}$ and $\hat{Z}_{kc,P}$ are independent, which is true if the two samples were drawn independently of each other. In the penultimate line, we used that $E_b E_d\left(\hat{Z}_{kc,P}-Z_{kc}\right)=0$ and we used expressions (11) and (12). In the final line, we used that, by assumption, $\sum_{c=1}^{C}\beta_c=0$.

Hence, a consistent estimator of σ_k^2 is given by

$$
\begin{aligned}
\hat{\sigma}_k^2 = \frac{n_{Pk}}{C(n_{Pk}-1)}\sum_{c=1}^{C}\Bigg[&\left(\hat{Z}_{kc,NP}-\hat{Z}_{kc,P}\right)^2-\left(1-\frac{1}{n_{Pk}}\right)\hat{\beta}_c^2-\frac{2}{n_{Pk}}\hat{\beta}_c\hat{Z}_{kc,NP} \\
&-\frac{1+n_{NPk}/n_{Pk}}{n_{NPk}-1}\hat{v}_{kc}\Bigg],
\end{aligned}
\qquad (2514)
$$

with $\hat{v}_{kc}$ from expression (2.8) in the report and $\hat{\beta}_c$ from expression (24).

7.2 Appendix B: MARMSE of combined estimator

Table B1. MARMSE of the combined estimator of equal size categories and weak selectivity. Model A

K	PS	c = 3, NPS 100	c = 3, NPS 1000	c = 3, NPS 2000	c = 3, NPS 6000	c = 5, NPS 100	c = 5, NPS 1000	c = 5, NPS 2000	c = 5, NPS 6000	c = 8, NPS 100	c = 8, NPS 1000	c = 8, NPS 2000	c = 8, NPS 6000	c = 15, NPS 100	c = 15, NPS 1000	c = 15, NPS 2000	c = 15, NPS 6000
K = 1	10	0.08	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
	100	0.04	0.05	0.05	0.05	0.03	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	900	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 4	10	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	0.03	0.03	0.02	0.02
	100	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	900	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 10	10	0.1	0.1	0.1	0.09	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.02	0.02
	100	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
	900	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 15	100	0.1	0.1	0.1	0.08	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02
	1000	0.05	0.05	0.05	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.01
	4000	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0
	9000	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0

Table B2 MARMSE of the combined estimator of equal size categories and weak selectivity. Model B

K	PS	c = 3, NPS 100	c = 3, NPS 1000	c = 3, NPS 2000	c = 3, NPS 6000	c = 5, NPS 100	c = 5, NPS 1000	c = 5, NPS 2000	c = 5, NPS 6000	c = 8, NPS 100	c = 8, NPS 1000	c = 8, NPS 2000	c = 8, NPS 6000	c = 15, NPS 100	c = 15, NPS 1000	c = 15, NPS 2000	c = 15, NPS 6000
K = 1	10	0.1	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.07	0.06	0.05	0.05	0.05
	100	0.04	0.05	0.05	0.05	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	900	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 4	10	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03
	100	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	900	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 10	10	0.1	0.1	0.1	0.09	0.07	0.07	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02
	100	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01
	400	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	900	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
K = 15	100	0.1	0.1	0.1	0.08	0.07	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02
	1000	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01
	4000	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0
	9000	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0

Table B3 MARMSE of the combined estimator of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.1	0.1	0.1	0.1	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.03	0.03	0.03	0.03	10
	0.05	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.1	0.1	0.1	0.1	0.07	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03	10
	0.05	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.1	0.1	0.09	0.07	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.02	10
	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.1	0.1	0.1	0.08	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.03	0.04	0.03	0.03	0.02	100
	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01	1000
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	4000
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table B4. MARMSE of the combined estimator of unequal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.1	0.11	0.11	0.11	0.08	0.09	0.09	0.09	0.08	0.07	0.07	0.07	0.05	0.05	0.05	0.05	10
	0.04	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.1	0.1	0.1	0.1	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.03	0.03	0.03	10
	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.1	0.1	0.09	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.03	0.02	10
	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.1	0.1	0.1	0.08	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.02	100
	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0.01	1000
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0	4000
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table B5. MARMSE of the combined estimator of equal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.11	0.15	0.15	0.15	0.1	0.11	0.11	0.11	0.08	0.08	0.08	0.08	0.06	0.05	0.05	0.05	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.11	0.11	0.11	0.11	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.05	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.13	0.11	0.11	0.11	0.09	0.08	0.08	0.08	0.06	0.05	0.05	0.05	0.03	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.08	0.11	0.11	0.11	0.05	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.02	100
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	1000
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	4000
	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table B6. MARMSE of the combined estimator of equal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.12	0.15	0.15	0.15	0.11	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.06	0.06	0.06	0.06	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.14	0.14	0.14	0.14	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.14	0.14	0.14	0.12	0.1	0.1	0.1	0.08	0.07	0.07	0.07	0.06	0.05	0.04	0.04	0.03	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.14	0.14	0.14	0.08	0.1	0.1	0.1	0.06	0.07	0.07	0.07	0.04	0.05	0.04	0.04	0.02	100
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	1000
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	4000
	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table B7. MARMSE of the combined estimator of unequal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.12	0.15	0.15	0.15	0.11	0.1	0.1	0.1	0.09	0.08	0.08	0.08	0.06	0.05	0.05	0.05	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	0.05	0.05	0.05	0.04	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.13	0.11	0.1	0.1	0.08	0.08	0.08	0.08	0.06	0.05	0.05	0.05	0.03	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.08	0.1	0.1	0.1	0.06	0.08	0.08	0.08	0.04	0.05	0.05	0.04	0.02	100
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	1000
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	4000
	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	9000
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table B8. MARMSE of the combined estimator of unequal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.13	0.15	0.15	0.15	0.11	0.11	0.11	0.11	0.09	0.09	0.09	0.09	0.06	0.06	0.06	0.06	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.14	0.14	0.14	0.14	0.1	0.1	0.1	0.1	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.14	0.14	0.14	0.12	0.1	0.09	0.09	0.08	0.07	0.07	0.07	0.06	0.04	0.04	0.04	0.03	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.14	0.14	0.14	0.08	0.1	0.09	0.09	0.06	0.07	0.07	0.07	0.04	0.04	0.04	0.04	0.02	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

7.3 Appendix C: MARMSE of probability sample

Table C1. MARMSE of the probability sample of equal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C2. MARMSE of the probability sample of equal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C3. MARMSE of the probability sample of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C4. MARMSE of the probability sample of unequal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C5. MARMSE of the probability sample of equal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C6. MARMSE of the probability sample of equal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.13	0.13	0.13	0.13	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C7. MARMSE of the probability sample of unequal size categories and severe selectivity. Model A

MARMSE - Unequal size categories and severe selectivity - Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table C8. MARMSE of the probability sample of unequal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.15	0.15	0.15	0.15	0.12	0.12	0.12	0.12	0.1	0.1	0.1	0.1	0.08	0.08	0.08	0.08	10
	0.05	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

7.4 Appendix D: MARMSE of non-probability sample

Table D1. MARMSE of the non-probability sample of equal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.04	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	10
	0.04	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	100
	0.04	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	400
	0.04	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	900
K = 4	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	100
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	400
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 10	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	100
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	400
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 15	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.02	0.01	0	10
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.02	0.01	0	100
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.02	0.01	0	400
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.02	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D2. MARMSE of the non-probability sample of equal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.05	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.05	0.06	0.06	0.06	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	100
	0.04	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	400
	0.05	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	900
K = 4	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.02	0.02	0.03	0.02	0.01	0.01	100
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	400
	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 10	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.03	0.04	0.03	0.02	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.02	0.01	0.01	100
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	400
	0.08	0.06	0.06	0.04	0.06	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 15	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0.01	0.03	0.01	0.01	0	10
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0.01	0.03	0.02	0.01	0	100
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0.01	0.03	0.02	0.01	0	400
	0.08	0.06	0.06	0.01	0.06	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.02	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D3. MARMSE of the non-probability sample of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.07	0.07	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	10
	0.07	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	100
	0.07	0.07	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	400
	0.07	0.07	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	900
K = 4	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.01	0.01	0.01	100
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.01	0.01	0.01	400
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 10	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	10
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	100
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	400
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	900
K = 15	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.01	0.01	0	10
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.01	0.01	0	100
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.01	0.01	0	400
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D4. MARMSE of the non-probability sample of unequal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.05	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.01	0.01	10
	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01	100
	0.06	0.06	0.06	0.06	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.01	0.01	0.01	400
	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	900
K = 4	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	100
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	400
	0.08	0.06	0.06	0.06	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	900
K = 10	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.02	0.01	0.01	10
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	100
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.03	0.02	0.03	0.01	0.01	0.01	400
	0.08	0.06	0.06	0.04	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.02	0.03	0.01	0.01	0.01	900
K = 15	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.03	0.01	0.03	0.01	0.01	0	10
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0	0.03	0.01	0.01	0	100
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0.01	0.03	0.01	0.01	0	400
	0.08	0.06	0.06	0.01	0.05	0.04	0.04	0.01	0.04	0.03	0.02	0.01	0.03	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D5. MARMSE of the non-probability sample of equal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.09	0.18	0.18	0.18	0.1	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.1	0.18	0.18	0.18	0.1	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.09	0.18	0.18	0.18	0.1	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.09	0.18	0.18	0.18	0.1	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	900
K = 4	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	900
K = 10	0.19	0.18	0.18	0.12	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	10
	0.19	0.18	0.18	0.12	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	100
	0.19	0.18	0.18	0.12	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	400
	0.19	0.18	0.18	0.12	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	900
K = 15	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	10
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	100
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	400
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D6. MARMSE of the non-probability sample of equal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.09	0.18	0.18	0.18	0.09	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.09	0.18	0.18	0.18	0.08	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.09	0.18	0.18	0.18	0.09	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.09	0.18	0.18	0.18	0.1	0.11	0.11	0.11	0.07	0.07	0.07	0.07	0.04	0.04	0.04	0.04	900
K = 4	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	10
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.18	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	900
K = 10	0.19	0.18	0.18	0.12	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.03	10
	0.19	0.18	0.18	0.11	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	100
	0.19	0.18	0.18	0.11	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.03	400
	0.18	0.18	0.18	0.11	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.05	0.04	0.04	0.02	900
K = 15	0.18	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	10
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	100
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	400
	0.19	0.18	0.18	0.04	0.12	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.05	0.04	0.04	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D7. MARMSE of the non-probability sample of unequal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.12	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.12	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.11	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.12	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	900
K = 4	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	10
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	100
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	400
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	900
K = 10	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	10
	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	100
	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	400
	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	900
K = 15	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	10
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	100
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	400
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table D8. MARMSE of the non-probability sample of unequal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.11	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	10
	0.1	0.18	0.18	0.18	0.13	0.11	0.11	0.11	0.09	0.07	0.07	0.07	0.05	0.04	0.04	0.04	100
	0.1	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	400
	0.11	0.18	0.18	0.18	0.12	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	900
K = 4	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	10
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	100
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	400
	0.18	0.18	0.18	0.18	0.11	0.11	0.11	0.11	0.08	0.07	0.07	0.07	0.04	0.04	0.04	0.04	900
K = 10	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	10
	0.18	0.18	0.18	0.11	0.12	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	100
	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	400
	0.19	0.18	0.18	0.11	0.11	0.11	0.11	0.07	0.08	0.07	0.07	0.05	0.04	0.04	0.04	0.02	900
K = 15	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	10
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	100
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	400
	0.19	0.18	0.18	0.04	0.11	0.11	0.11	0.02	0.08	0.07	0.07	0.01	0.04	0.04	0.04	0.01	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

7.5 Appendix E: Bias

Table E1. Difference of mean absolute bias between the non-probability and the combined estimator of equal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	**-0.08**	0	0	0	**-0.03**	**-0.04**	**-0.04**	**-0.04**	0	0	0	0	**-0.02**	**-0.03**	**-0.03**	**-0.03**	10
	-0.01	0.01	0.01	0	0	0.01	0	0	0	0.01	0.01	0.01	0	0	0	0	100
	0.04	0.02	0.03	0.02	0	0.03	0.01	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.03	0.06	0.05	0.05	0.01	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0	0	0	900
K = 4	0.02	0.01	0	0	**-0.02**	**-0.02**	**-0.02**	**-0.02**	0	0	0	0	0	0	0	0	10
	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0	0	0	0	0	0	0	100
	0.05	0.05	0.03	0.04	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	400
	0.08	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	900
K = 10	0	0	0	0	0	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	10
	0.04	0.03	0.03	0.01	0.02	0.01	0	0	0.01	0	0	0	0.01	0	0	0	100
	0.06	0.03	0.03	0.02	0.04	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0	400
	0.07	0.05	0.05	0.03	0.04	0.03	0.03	0.01	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0	900
K = 15	0.01	0.01	0.01	0	**-0.01**	**-0.01**	**-0.01**	**-0.03**	0	0	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	10
	0.03	0.02	0.02	**-0.01**	0.02	0.01	0.01	0	0.01	0	0	**-0.01**	0.01	0	0	0	100
	0.06	0.04	0.04	0	0.04	0.02	0.02	0	0.02	0.01	0.01	0	0.02	0.01	0.01	0	400
	0.06	0.05	0.05	0	0.05	0.03	0.03	0	0.02	0.02	0.02	0	0.02	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E2. Difference of mean absolute bias between the non-probability and the combined estimator of equal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0	0	0	0.01	0	0	0.01	0	**-0.04**	**-0.05**	**-0.05**	**-0.04**	**-0.02**	**-0.02**	**-0.02**	**-0.02**	10
	0.02	0.02	0.02	0.01	0	0	0	0	0	0.01	0.01	0.01	0	0	0	0	100
	0	0.05	0.05	0.04	0.02	0.04	0.03	0.03	0.01	0.01	0.01	0.01	0.01	0	0	0	400
	0.04	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.01	900
K = 4	0	0	0	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	0	0	0	0	10
	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0	0	0	100
	0.06	0.05	0.05	0.05	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	400
	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	900
K = 10	**-0.03**	**-0.03**	**-0.03**	**-0.04**	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0	0	0	**-0.01**	10
	0.03	0.02	0.02	0	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0	0.01	0	0	0	100
	0.05	0.05	0.05	0.02	0.04	0.02	0.02	0.01	0.03	0.01	0.01	0	0.02	0.01	0.01	0	400
	0.06	0.05	0.05	0.03	0.04	0.03	0.03	0.01	0.03	0.02	0.01	0.01	0.02	0.01	0.01	0	900
K = 15	0	0	**-0.01**	**-0.02**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0	**-0.01**	**-0.01**	**-0.01**	10
	0.03	0.03	0.02	**-0.01**	0.02	0.02	0.01	0	0.01	0.01	0.01	0	0.01	0	0	0	100
	0.03	0.04	0.05	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	0.01	0.01	0.01	0	400
	0.06	0.05	0.05	0	0.04	0.03	0.03	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E3. Difference of mean absolute bias between the non-probability and the combined estimator of unequal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	**-0.22**	**-0.2**	**-0.2**	**-0.2**	**-0.04**	**-0.08**	**-0.08**	**-0.08**	0.01	0	0	0	0	0	**-0.01**	**-0.01**	10
	0.03	0.06	0.06	0.05	0.02	0.02	0	0.01	0	0	0	0	0.01	0	0	0	100
	0.01	0.07	0.05	0.05	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.05	0.05	0.06	0.06	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.02	0.01	0.01	0.01	0.01	900
K = 4	**-0.07**	**-0.06**	**-0.06**	**-0.06**	0	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	10
	0.05	0.03	0.02	0.02	0.02	0	0	0	0.01	0.01	0	0.01	0	0	0	0	100
	0.07	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	0	400
	0.07	0.05	0.05	0.05	0.02	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.01	0.01	0.01	0.01	900
K = 10	**-0.04**	**-0.04**	**-0.04**	**-0.05**	**-0.01**	**-0.02**	**-0.02**	**-0.02**	**-0.01**	0	**-0.01**	**-0.01**	0	0	0	0	10
	0.03	0.02	0.02	0.01	0.01	0	0	**-0.01**	0.01	0	0	0	0.01	0	0	0	100
	0.05	0.05	0.04	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.01	0	0.01	0	0	0	400
	0.04	0.05	0.04	0.03	0.05	0.03	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0	900
K = 15	**-0.03**	**-0.03**	**-0.03**	**-0.04**	**-0.01**	**-0.01**	**-0.01**	**-0.02**	0	**-0.01**	**-0.01**	**-0.01**	0	0	0	**-0.01**	10
	0.05	0.03	0.04	0	0.01	0.01	0.01	**-0.01**	0.01	0	0	0	0.01	0	0	0	100
	0.05	0.05	0.04	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	0.01	0.01	0	0	400
	0.05	0.05	0.05	0	0.04	0.03	0.03	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E4. Difference of mean absolute bias between the non-probability and the combined estimator of unequal size categories and weak selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.04	0	0	0	0.02	0.01	0.01	0.01	**-0.01**	**-0.02**	**-0.02**	**-0.03**	**-0.01**	**-0.02**	**-0.02**	**-0.02**	10
	0.04	0.02	0.05	0.03	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0	0	0	0	100
	0.03	0.05	0.05	0.04	0.04	0.03	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0	0	0	400
	0.04	0.06	0.05	0.05	0.04	0.03	0.02	0.03	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.02	0.01	0.01	0.01	**-0.01**	0	**-0.01**	**-0.01**	0.01	0	0	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	10
	0.01	0.03	0.03	0.03	0.03	0.01	0.02	0.01	0.02	0	0	0	0.01	0	0	0	100
	0.06	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	400
	0.06	0.05	0.05	0.05	0.03	0.03	0.02	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	900
K = 10	0	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	10
	0.04	0.03	0.03	0.01	0.03	0.01	0.01	0	0.01	0.01	0	0	0	0	0	0	100
	0.05	0.04	0.04	0.02	0.04	0.03	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.06	0.05	0.05	0.03	0.04	0.03	0.02	0.01	0.02	0.02	0.02	0.01	0.02	0.01	0.01	0	900
K = 15	**-0.01**	**-0.01**	**-0.01**	**-0.04**	**-0.02**	**-0.03**	**-0.03**	**-0.04**	0	0	**-0.01**	**-0.01**	0	0	0	**-0.01**	10
	0.03	0.02	0.02	**-0.01**	0.02	0.01	0.01	0	0.01	0.01	0.01	0	0.01	0	0	0	100
	0.05	0.04	0.04	0	0.03	0.02	0.02	0	0.02	0.01	0.01	0	0.01	0.01	0.01	0	400
	0.06	0.05	0.05	0	0.04	0.03	0.03	0	0.02	0.02	0.02	0	0.02	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E5. Difference of mean absolute bias between the non-probability and the combined estimator of equal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	**-0.02**	0	0	0	0.04	0.05	0.07	0.06	0.03	0.03	0.02	0.02	**-0.02**	**-0.01**	**-0.01**	**-0.01**	10
	0.04	0.14	0.14	0.13	0.05	0.08	0.08	0.07	0.04	0.04	0.05	0.04	0.02	0.01	0.02	0.02	100
	0.06	0.15	0.16	0.16	0.08	0.09	0.09	0.09	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	400
	0.09	0.16	0.17	0.17	0.1	0.11	0.09	0.11	0.06	0.06	0.06	0.06	0.04	0.03	0.03	0.03	900
K = 4	0.08	0.07	0.07	0.07	0	0	0	0	0.02	0.01	0.01	0.01	0	**-0.01**	**-0.01**	**-0.01**	10
	0.12	0.11	0.12	0.12	0.07	0.08	0.08	0.08	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.02	100
	0.15	0.17	0.17	0.16	0.1	0.1	0.1	0.1	0.06	0.06	0.06	0.06	0.03	0.03	0.03	0.03	400
	0.16	0.16	0.17	0.16	0.1	0.1	0.1	0.1	0.06	0.06	0.06	0.06	0.03	0.03	0.03	0.03	900
K = 10	0.05	0.03	0.03	**-0.02**	0.02	0.03	0.02	0.02	0	0	0	**-0.01**	0	0	0	0	10
	0.15	0.15	0.15	0.08	0.08	0.08	0.08	0.04	0.05	0.04	0.04	0.02	0.02	0.02	0.02	0.01	100
	0.15	0.16	0.15	0.09	0.1	0.09	0.09	0.06	0.06	0.06	0.06	0.03	0.03	0.03	0.03	0.02	400
	0.17	0.17	0.17	0.11	0.1	0.1	0.1	0.06	0.06	0.06	0.06	0.04	0.03	0.03	0.03	0.02	900
K = 15	0.1	0.08	0.08	0	0.02	0.02	0.02	**-0.01**	0.01	0.01	0.01	**-0.01**	0	**-0.01**	**-0.01**	**-0.01**	10
	0.13	0.13	0.13	0	0.07	0.07	0.07	0	0.05	0.05	0.05	0	0.02	0.02	0.02	0	100
	0.17	0.15	0.16	0.02	0.09	0.09	0.1	0.01	0.06	0.06	0.06	0	0.03	0.03	0.03	0	400
	0.16	0.16	0.16	0.03	0.1	0.1	0.1	0.01	0.07	0.06	0.06	0.01	0.03	0.03	0.03	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E6. Difference of mean absolute bias between the non-probability and the combined estimator of equal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.01	0.14	0.14	0.14	0	0.01	0.01	0.01	**-0.02**	**-0.01**	**-0.01**	**-0.02**	0	**-0.01**	**-0.02**	**-0.01**	10
	0.03	0.16	0.17	0.16	0.04	0.09	0.09	0.09	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.02	100
	0.06	0.18	0.16	0.17	0.07	0.08	0.09	0.09	0.06	0.06	0.06	0.06	0.03	0.03	0.03	0.02	400
	0.07	0.16	0.17	0.17	0.08	0.1	0.1	0.1	0.05	0.06	0.06	0.06	0.04	0.03	0.03	0.03	900
K = 4	0.06	0.08	0.07	0.07	0.04	0.04	0.03	0.03	0	0	0	0	0	0	0	0.01	10
	0.15	0.14	0.15	0.15	0.07	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.02	0.02	0.02	0.02	100
	0.17	0.16	0.16	0.16	0.1	0.1	0.09	0.09	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	400
	0.16	0.17	0.17	0.17	0.1	0.1	0.1	0.1	0.07	0.06	0.06	0.06	0.04	0.03	0.03	0.03	900
K = 10	0.07	0.07	0.07	0.02	0.03	0.03	0.03	0.01	0.02	0.02	0.01	0.01	0	0	0	0	10
	0.15	0.15	0.15	0.09	0.08	0.08	0.08	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.01	100
	0.17	0.16	0.16	0.1	0.1	0.1	0.1	0.06	0.06	0.06	0.05	0.03	0.03	0.03	0.03	0.02	400
	0.16	0.16	0.17	0.1	0.11	0.1	0.1	0.06	0.07	0.06	0.06	0.04	0.04	0.03	0.03	0.02	900
K = 15	0.06	0.06	0.05	**-0.03**	0.03	0.03	0.03	**-0.01**	0.02	0.02	0.02	0	0.01	0.01	0	**-0.01**	10
	0.15	0.14	0.14	0	0.08	0.08	0.08	0	0.05	0.05	0.05	0	0.02	0.02	0.02	0	100
	0.16	0.16	0.16	0.02	0.1	0.1	0.1	0.01	0.06	0.06	0.05	0	0.03	0.03	0.03	0	400
	0.17	0.16	0.17	0.03	0.11	0.1	0.1	0.01	0.07	0.06	0.06	0.01	0.04	0.03	0.03	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E7. Difference of mean absolute bias between the non-probability and the combined estimator of unequal size categories and severe selectivity, Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.03	0.01	0.06	0.05	0.01	0.01	0.01	0.01	0.01	0	0	0	0.01	0.01	0	0	10
	0.04	0.14	0.15	0.14	0.1	0.09	0.09	0.09	0.06	0.06	0.06	0.06	0.03	0.02	0.02	0.02	100
	0.08	0.16	0.15	0.17	0.12	0.09	0.09	0.09	0.07	0.05	0.05	0.05	0.03	0.03	0.03	0.03	400
	0.1	0.16	0.17	0.17	0.09	0.1	0.09	0.1	0.06	0.06	0.06	0.06	0.04	0.03	0.03	0.03	900
K = 4	0.06	0.06	0.07	0.06	0.03	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0	0	0	0	10
	0.16	0.15	0.14	0.14	0.08	0.08	0.08	0.08	0.04	0.05	0.05	0.05	0.02	0.02	0.02	0.02	100
	0.15	0.16	0.15	0.16	0.09	0.08	0.08	0.08	0.06	0.05	0.05	0.05	0.03	0.03	0.03	0.03	400
	0.16	0.16	0.16	0.16	0.09	0.1	0.1	0.1	0.07	0.06	0.06	0.06	0.03	0.03	0.03	0.03	900
K = 10	0.03	0.02	0.02	**-0.02**	0.01	0.01	0.01	**-0.01**	0.01	0.01	0.01	0	0	0	0	0	10
	0.15	0.14	0.15	0.08	0.08	0.08	0.08	0.03	0.05	0.04	0.04	0.02	0.02	0.02	0.02	0.01	100
	0.15	0.16	0.16	0.09	0.09	0.09	0.09	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	0.02	400
	0.19	0.16	0.16	0.09	0.1	0.1	0.1	0.05	0.07	0.06	0.06	0.04	0.03	0.03	0.03	0.02	900
K = 15	0.04	0.04	0.04	**-0.04**	0.03	0.02	0.02	**-0.02**	0	0.01	0	**-0.01**	0	0	0	**-0.01**	10
	0.15	0.15	0.15	0.01	0.07	0.08	0.07	0	0.05	0.04	0.04	0	0.02	0.02	0.02	0	100
	0.17	0.16	0.16	0.02	0.09	0.09	0.09	0.01	0.06	0.05	0.05	0	0.03	0.03	0.03	0	400
	0.16	0.16	0.17	0.03	0.09	0.1	0.1	0.02	0.06	0.06	0.06	0.01	0.03	0.03	0.03	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E8. Difference of mean absolute bias between the non-probability and the combined estimator of unequal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	**-0.01**	0.09	0.11	0.1	0.01	0.01	0.01	0.01	0	**-0.02**	**-0.02**	**-0.02**	0	0	0	0	10
	0	0.13	0.14	0.12	0.08	0.07	0.09	0.08	0.05	0.03	0.04	0.04	0.03	0.01	0.01	0.02	100
	0.11	0.17	0.17	0.17	0.08	0.09	0.1	0.11	0.06	0.05	0.06	0.05	0.04	0.03	0.03	0.03	400
	0.06	0.19	0.16	0.16	0.09	0.09	0.09	0.09	0.06	0.06	0.06	0.06	0.04	0.03	0.03	0.03	900
K = 4	0.05	0.06	0.06	0.06	0.03	0.04	0.05	0.04	0.02	0.03	0.02	0.03	0	0	0	0	10
	0.15	0.14	0.14	0.14	0.08	0.07	0.07	0.07	0.05	0.04	0.04	0.04	0.02	0.02	0.02	0.02	100
	0.14	0.16	0.16	0.16	0.09	0.1	0.09	0.1	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	400
	0.16	0.16	0.17	0.17	0.09	0.1	0.1	0.09	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	900
K = 10	0.09	0.09	0.09	0.03	0.04	0.03	0.03	0.01	0.01	0.01	0.01	0	0	0	0	0	10
	0.16	0.14	0.14	0.07	0.08	0.08	0.08	0.04	0.05	0.04	0.04	0.02	0.02	0.02	0.02	0.01	100
	0.15	0.16	0.16	0.09	0.1	0.09	0.09	0.05	0.06	0.06	0.06	0.03	0.03	0.03	0.03	0.02	400
	0.17	0.17	0.17	0.1	0.1	0.1	0.1	0.05	0.07	0.06	0.06	0.04	0.03	0.03	0.03	0.02	900
K = 15	0.03	0.03	0.03	**-0.04**	0.02	0.02	0.02	**-0.03**	0.01	0.01	0.01	**-0.02**	0.01	0.01	0	**-0.01**	10
	0.13	0.14	0.14	0.01	0.08	0.08	0.08	0	0.05	0.04	0.04	0	0.02	0.02	0.02	0	100
	0.15	0.16	0.16	0.02	0.1	0.09	0.09	0.01	0.06	0.06	0.06	0	0.03	0.03	0.03	0	400
	0.16	0.17	0.16	0.02	0.1	0.1	0.1	0.01	0.07	0.06	0.06	0.01	0.03	0.03	0.03	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E9. Mean absolute bias of the combined estimator of equal size categories and weak selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.11	0.05	0.05	0.07	0.07	0.09	0.08	0.08	0.02	0.03	0.03	0.02	0.05	0.04	0.04	0.04	10
	0.04	0.06	0.06	0.06	0.02	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.04	0.03	0.04	0.03	0.02	0.02	0.02	0.02	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.06	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	10
	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.07	0.06	0.06	0.04	0.05	0.04	0.04	0.03	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.02	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.07	0.05	0.05	0.01	0.06	0.05	0.05	0.03	0.04	0.03	0.03	0.01	0.03	0.02	0.02	0.01	10
	0.04	0.04	0.04	0.02	0.03	0.03	0.03	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0	100
	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E10. Mean absolute bias of the combined estimator of equal size categories and weak selectivity - Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.02	0.06	0.06	0.06	0.05	0.05	0.03	0.04	0.06	0.07	0.07	0.07	0.05	0.03	0.04	0.03	10
	0.02	0.05	0.05	0.05	0.03	0.04	0.04	0.04	0.02	0.01	0.02	0.02	0.02	0.01	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	900
K = 4	0.07	0.07	0.06	0.06	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	10
	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.09	0.09	0.08	0.06	0.05	0.05	0.04	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01	10
	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	100
	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
K = 15	0.07	0.07	0.07	0.03	0.06	0.05	0.05	0.03	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.01	10
	0.04	0.03	0.04	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0	100
	0.01	0.02	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E11. Mean absolute bias of the combined estimator of unequal size categories and weak selectivity - Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.27	0.25	0.26	0.26	0.08	0.11	0.11	0.11	0.03	0.03	0.03	0.02	0.01	0.02	0.02	0.02	10
	0.04	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.12	0.12	0.12	0.12	0.05	0.04	0.04	0.04	0.04	0.03	0.03	0.04	0.03	0.02	0.02	0.02	10
	0.03	0.03	0.03	0.03	0.04	0.03	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.11	0.1	0.1	0.09	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.01	0.01	10
	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
K = 15	0.09	0.09	0.09	0.06	0.05	0.05	0.05	0.02	0.04	0.03	0.03	0.01	0.02	0.02	0.02	0.01	10
	0.02	0.03	0.03	0.01	0.03	0.03	0.03	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0	100
	0.02	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E12. Mean absolute bias of the combined estimator of unequal size categories and weak selectivity - Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.02	0.06	0.06	0.06	0.03	0.03	0.03	0.02	0.05	0.05	0.05	0.05	0.03	0.03	0.03	0.03	10
	0.02	0.03	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	100
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0	0	900
K = 4	0.06	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02	10
	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	900
K = 10	0.07	0.07	0.07	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	10
	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	900
K = 15	0.08	0.07	0.07	0.05	0.07	0.06	0.06	0.05	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	10
	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E13. Mean absolute bias of the combined estimator of Equal size categories and severe selectivity - Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.11	0.18	0.19	0.18	0.06	0.04	0.05	0.05	0.04	0.04	0.05	0.05	0.07	0.05	0.05	0.05	10
	0.07	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	100
	0.02	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.1	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	10
	0.06	0.06	0.06	0.06	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.14	0.14	0.14	0.13	0.08	0.08	0.09	0.06	0.07	0.07	0.07	0.05	0.04	0.04	0.04	0.03	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.09	0.09	0.09	0.04	0.09	0.09	0.09	0.04	0.06	0.06	0.06	0.02	0.05	0.04	0.04	0.02	10
	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E14. Mean absolute bias of the combined estimator of equal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.12	0.04	0.04	0.03	0.07	0.1	0.1	0.1	0.09	0.08	0.08	0.09	0.05	0.05	0.05	0.05	10
	0.04	0.01	0.02	0.02	0.03	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.01	0.02	0.02	0.02	100
	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.11	0.11	0.1	0.11	0.07	0.08	0.08	0.08	0.07	0.06	0.07	0.07	0.04	0.03	0.03	0.03	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	100
	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.1	0.11	0.11	0.09	0.08	0.08	0.08	0.06	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.03	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.12	0.12	0.13	0.07	0.08	0.08	0.08	0.04	0.06	0.05	0.05	0.02	0.04	0.03	0.03	0.01	10
	0.04	0.03	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E15. Mean absolute bias of the combined estimator of unequal size categories and severe selectivity. Model A

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.07	0.15	0.13	0.13	0.13	0.11	0.1	0.1	0.07	0.07	0.07	0.06	0.03	0.03	0.04	0.03	10
	0.03	0.03	0.04	0.04	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 4	0.11	0.11	0.12	0.11	0.08	0.07	0.08	0.08	0.06	0.06	0.06	0.06	0.04	0.04	0.04	0.04	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 10	0.16	0.16	0.16	0.13	0.09	0.1	0.1	0.07	0.06	0.06	0.06	0.05	0.04	0.03	0.04	0.03	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	900
K = 15	0.13	0.14	0.14	0.07	0.08	0.08	0.09	0.04	0.07	0.07	0.06	0.02	0.04	0.04	0.04	0.02	10
	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

Table E16. Mean absolute bias of the combined estimator of unequal size categories and severe selectivity. Model B

	c = 3				c = 5				c = 8				c = 15				PS
K = 1	0.02	0.06	0.06	0.06	0.03	0.03	0.03	0.02	0.05	0.05	0.05	0.05	0.03	0.03	0.03	0.03	10
	0.02	0.03	0.02	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	100
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0	0	900
K = 4	0.06	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.02	10
	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	900
K = 10	0.07	0.07	0.07	0.05	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.02	0.03	0.02	0.02	0.01	10
	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	900
K = 15	0.08	0.07	0.07	0.05	0.07	0.06	0.06	0.05	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	10
	0.04	0.04	0.04	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	100
	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	400
	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	0	900
NPS	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	100	1000	2000	6000	

8. REFERENCES

Baker, R., Brick, J., Bates, N., Battaglia, M., Couper, M., Dever, J., Gile, K., and Tourangeau, R. (2013), "Summary Report of the AAPOR Task Force on Non-probability Sampling," *Journal of Survey Statistics and Methodology*, 1, 90–143.

Brick, J. M. (2014), "Explorations in non-probability sampling using the web," *Proceedings of Statistics Canada Symposium 2014. Beyond traditional survey taking: Adapting to a changing world.* Retrieved from: https://www.statcan.gc.ca/eng/conferences/symposium2014/program/14252-eng.pdf

Elliott, M., and Haviland, A. (2007), "Use of a Web-Based Convenience Sample to Supplement a Probability Sample," *Survey Methodology*, 33, 211–215.

Elliott, M. R., and Valliant, R. (2017), "Inference for Nonprobability Samples," *Statistical Science*, 32, 249–264.

Linder, F., D. van Roon and B. Bakker (2014), "Combining Data from Administrative Sources and Sample Surveys; the Single-Variable Case. Case Study: Educational Attainment." Report for Work Package 4.2 of the ESSnet project Data Integration.

Lohr, S.L. (2019)*, "Sampling: Design and Analysis (2nd ed.),"* Boca Raton, Florida: Chapman and Hall/CRC.

Pannekoek, J., and De Waal, T. (1998), "Synthetic and Combined Estimators in Statistical Disclosure Control," *Journal of Official Statistics,* 14, 399-410.

Pfeffermann, D. (2002), "Small Area Estimation – New Developments and Directions," *International Statistical Review,* 70, 125-143.

Pfeffermann, D. (2013), "New Important Developments in Small Area Estimation," *Statistical Science*, 28, 40–68.

Rao, J., (2003). Small area estimation. New York: Wiley.

Särndal, C-E, Swensson, B., & Wretman, J. H. (1992). Model assisted survey sampling. New York: Springer.

Smit, V.I.C. (2021), "Correcting Selectivity in Datasets with Pseudo-Weights: A Simulation Study," Master Thesis, Leiden University, The Netherlands. Available at: https://doi.org/10.13140/RG.2.2.28253.74726

Valliant, R. (2020), "Comparing Alternatives for Estimation from Nonprobability Samples," *Journal of Survey Statistics and Methodology*, 8, 231–263.

Van den Brakel, J. (2019), "New data sources and inference methods for statistics," Discussion paper. Statistics Netherlands.

Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., and Blom, A. G. (2020), "Integrating Probability and Nonprobability Samples for Survey Inference," *Journal of Survey Statistics and Methodology*, 8, 120-147.