Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Sampling with adaptive drawing probabilities

by Bardia Panahbehagh, Yves Tillé and Azad Khanzadi

Release date: June 30, 2023

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2023
Vol. 49, No. 1, pp. 191-211
Statistics Canada, Catalogue No. 12-001-X

Sampling with adaptive drawing probabilities

Bardia Panahbehagh, Yves Tillé and Azad Khanzadi[1]

Abstract

In this paper, with and without-replacement versions of adaptive proportional to size sampling are presented. Unbiased estimators are developed for these methods and their properties are studied. In the two versions, the drawing probabilities are adapted during the sampling process based on the observations already selected. To this end, in the version with-replacement, after each draw and observation of the variable of interest, the vector of the auxiliary variable will be updated using the observed values of the variable of interest to approximate the exact selection probability proportional to size. For the without-replacement version, first, using an initial sample, we model the relationship between the variable of interest and the auxiliary variable. Then, utilizing this relationship, we estimate the unknown (unobserved) population units. Finally, on these estimated population units, we select a new sample proportional to size without-replacement. These approaches can significantly improve the efficiency of designs not only in the case of a positive linear relationship, but also in the case of a non-linear or negative linear relationship between the variables. We investigate the efficiencies of the designs through simulations and real case studies on medicinal flowers, social and economic data.

Key Words: Adaptive sampling; Efficiency; Regression models; Sampling design.

1. Introduction

In probability proportional to size sampling (PS), the sample units are selected proportional to size of an auxiliary variable. The sampling design with unequal probabilities with-replacement, PPS, is first introduced by Hansen and Hurwitz (1943). Madow (1949), Narain (1951) and Horvitz and Thompson (1952) proposed without-replacement versions of PPS as ΠPS. Many different schemes have been proposed for ΠPS of which 50 of them are listed in Brewer and Hanif (1983) and Tillé (2006, 2020). Almost all of these methods use the π-estimator (Narain, 1951; Horvitz and Thompson, 1952) to derive an unbiased estimator of the population total and its variance estimator. Generally ΠPS is more efficient than PPS, however PPS offers advantages over ΠPS with respect to simplicity of the sample selection and the variance estimator calculations.

Our goal is to improve PPS and ΠPS designs based on an adaptive approach. The word "adaptive" refers to the use of information from sampled units in the sampling process (Seber and Salehi, 2013). In adaptive designs, it is not possible to select the final sample before starting the sampling process. The concept of an adaptive design is to use the information from the observed sample units to obtain as much information as possible about the population. The proposed approaches are easy to implement. In Section 2, adaptive PPS and, in Section 3, adaptive ΠPS sampling are presented. Section 4 and 5 contain simulations and real case studies to evaluate the effectiveness of PPS sampling and ΠPS sampling, respectively. Conclusions are drawn in Section 6.

1. Bardia Panahbehagh, Institute of Statistics, University of Neuchatel, Neuchatel, Switzerland and Department of Mathematics, Kharazmi University, Tehran, Iran. E-mail: panahbehagh@khu.ac.ir; Yves Tillé, Institute of Statistics, University of Neuchatel, Neuchatel, Switzerland; Azad Khanzadi, Department of Economics, Razi University, Kermanshah, Iran.

2. Adaptive PPS (APPS) sampling

Assume that we have a finite population whose set of labels is denoted by $U = \{1, \ldots, k, \ldots, N\}$. The variable of interest is $\mathbf{y} = (y_1, \ldots, y_k, \ldots, y_N)^\top$ and the auxiliary variable is $\mathbf{x} = (x_1, \ldots, x_k, \ldots, x_N)^\top$. Both variables are assumed to be positive and non-zero, i.e., $\mathbf{y}, \mathbf{x} \in \mathbb{R}^N_{>0}$. Suppose that the parameter of interest is the total of the variable of interest,

$$t_y = \sum_{k \in U} y_k.$$

The total of the auxiliary variable is denoted by

$$t_x = \sum_{k \in U} x_k.$$

Also, for any subset A of U with cardinality N_A, we define

$$t_{yA} = \sum_{k \in A} y_k, \quad \bar{y}_A = \frac{1}{N_A} \sum_{k \in A} y_k \quad \text{and} \quad t_{xA} = \sum_{k \in A} x_k.$$

The basic idea behind APPS is to update the vector of auxiliary variables based on the information of the observed variable of interest after each draw. To take an APPS sample of size n, we proceed as described in Algorithm 1.

Algorithm 1. Adaptive PPS (APPS)

Define

$$\mathbf{p}_1 = \frac{\mathbf{x}}{t_x} = (p_{11}, \ldots, p_{1k}, \ldots, p_{1N})^\top = \left(\frac{x_1}{t_x}, \ldots, \frac{x_k}{t_x}, \ldots, \frac{x_N}{t_x} \right)^\top. \tag{2.1}$$

Define $s_0 = \{\ \}$

For $i = 1, \ldots, n$ do

- Select a unit (say j) in U with probabilities $\mathbf{p}_i = (p_{i1}, \ldots, p_{ik}, \ldots, p_{iN})^\top$.
- Define $s_i = s_{i-1} \cup \{j\}$.
- Compute $\mathbf{p}_{i+1} = \left(p_{(i+1)1}, \ldots, p_{(i+1)k}, \ldots, p_{(i+1)N} \right)^\top$, where

$$p_{(i+1)k} = \begin{cases} \dfrac{y_k t_{xs_i}}{t_x t_{ys_i}} & \text{if } k \in s_i \\ p_{1k} & \text{if } k \notin s_i \end{cases}, \quad \text{for all} \quad k \in U. \tag{2.2}$$

In Algorithm 1, the first two units are selected with-replacement using $\mathbf{p}_1$ and $\mathbf{p}_2 (= \mathbf{p}_1)$ respectively and we observe their y values. Indeed, according to (2.2) in Algorithm 1, after observing the y value of the first sample unit (say j) we have

$$p_{2k} = \begin{cases} \dfrac{y_j x_j}{t_x y_j} = \dfrac{x_j}{t_x} = p_{1j} & \text{if } k = j \\ p_{1k} & \text{if } k \neq j \end{cases}, \quad \text{for all} \quad k \in U,$$

or briefly $\mathbf{p}_2 = \mathbf{p}_1$. Therefore at least the y values of two different units are required to update the drawing probabilities vector $(\mathbf{p})$. For the third unit onwards, based on the observed y values, we update the vector of drawing probabilities. Each unit is then selected using a different drawing probabilities vector.

Result 1. *In APPS, for each* $i = 1, \ldots, n$,

$$\sum_{k \in U} p_{ik} = 1,$$

and

$$\hat{t}_{\text{APPS}} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_{k_i}}{p_{ik_i}}$$

is an unbiased estimator of t_y *with variance*

$$V\left(\hat{t}_{\text{APPS}}\right) = E\left[\frac{1}{n^2} \sum_{i=1}^{n} \sum_{k=1}^{N} \left(\frac{y_k}{p_{ik}} - t_y\right)^2 p_{ik}\right].$$

An unbiased estimator of the variance is given by:

$$\hat{V}\left(\hat{t}_{\text{APPS}}\right) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_{k_i}}{p_{ik_i}} - \hat{t}_{\text{APPS}}\right)^2,$$

where p_{ik_i} *is* k_i^{th} *unit of* $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iN})^{\top}$.

For the proof of Result 1, see Appendix A.

Setting the drawing probabilities exactly proportional to size of y, i.e., $p_k = y_k / t_y$, will lead to an unbiased estimator for t_y with zero variance,

$$\hat{t}_{\text{APPS}} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_{k_i}}{p_{ik_i}} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_{k_i}}{y_{k_i} / t_y} = t_y.$$

By following the procedure of Algorithm 1 step by step, the drawing probability for unit k approaches the ideal probability proportional to size based on y. As evidence, consider that if units k and ℓ have been selected at least once in steps up to and including i, then in all the steps after step i, the ratio of their drawing probabilities is equal to y_k / y_ℓ, which is the same as the ideal case,

$$\frac{p_{jk}}{p_{j\ell}} = \left(\frac{y_\ell\, t_{xs_j}}{t_x\, t_{ys_j}}\right)^{-1} \frac{y_k\, t_{xs_j}}{t_x\, t_{ys_j}} = \frac{y_k / t_y}{y_\ell / t_y}, \quad j > i.$$

3. Adaptive ΠPS (AΠPS) sampling

In general, without-replacement designs are more efficient than with-replacement designs of the same size due to the inclusion of unduplicated information. AΠPS is a kind of adaptive version of ΠPS. To take a AΠPS sample of size n, we proceed as described in Algorithm 2.

Algorithm 2. Adaptive ΠPS (AΠPS)

1. Based on a conventional design (like Simple Random Sampling without-replacement (SRSWOR) or ΠPS) an initial sample s_0 of size n_0 will be selected.
2. Using s_0, y is modeled, for example, by a polynomial of order M of x to detect the potentially non-linear relationship between x and y. In other words, we assume a superpopulation model as

   $$y_k = \beta_0 + \beta_1 x_k + \beta_2 x_k^2 + \ldots + \beta_M x_k^M + \varepsilon_k, \; k \in U,$$

   where ε_k is a random variable independent of x_k with $E(\varepsilon_k) = 0$ and then

   $$\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 + \ldots + \hat{\beta}_M x_k^M, \; k \in U^* = U \setminus s_0,$$

   where $\hat{\beta}_m, m = 1, \ldots, M,$ can be estimated using the least square method in finite population sampling. If $\hat{y}_k$ is negative or null, it is replaced by a small positive value of the y, so as not to have zero inclusion probabilities.
3. Based on $\hat{y}_k, k \in U^*$ we select a ΠPS of size $n^* = n - n_0$, say s^*.

In Algorithm 2, one can obviously use any parametric or non-parametric model instead of a linear model to obtain a forecast of y_k. Our sampling method will be all the more efficient if the prediction of y_k is accurate. The predicted values must be positive. With an AΠPS sampling design, we can estimate the population total by

$$\hat{t}_{A\Pi\text{PS}} = \sum_{k \in s^*} \frac{y_k}{\Pi_k^*} + \sum_{k \in s_0} y_k,$$

with

$$\Pi_k^* = E\left(I_k \mid s_0\right) = \min\left(c^* \hat{y}_k, 1\right),$$
$$\text{where constant } c^* \text{ is defined by } \sum_{k \in U^*} \min\left(c^* \hat{y}_k, 1\right) = n^*, \tag{3.1}$$

where I_k is an indicator function which takes 1 if unit k is selected as a unit of s^*. With inclusion probabilities exactly proportional to the size of y as in (3.1), provided that

$$0 < \frac{n^* y_k}{t_{yU^*}} \leq 1, \quad \text{for all} \quad k \in U^*,$$

we will have

$$\Pi_k^* = \frac{n^* y_k}{t_{yU^*}}, \quad \text{for all} \quad k \in U^*,$$

which will lead to an unbiased estimator for t_y with zero variance,

$$\hat{t}_{A\Pi\text{PS}} = \sum_{k \in s^*} \frac{y_k}{\Pi_k^*} + \sum_{k \in s_0} y_k = \sum_{k \in s^*} \frac{y_k}{n^* y_k / t_{yU^*}} + \sum_{k \in s_0} y_k = t_{yU^*} + t_{ys_0} = t_y.$$

Then, if we can estimate the y values with high accuracy based on Algorithm 2 and using initial sample s_0, we can estimate t_y with high efficiency.

Result 2. *In $A\Pi\,PS$,*

$$\hat{t}_{A\Pi\text{PS}} = \sum_{k \in s^*} \frac{y_k}{\Pi_k^*} + \sum_{k \in s_0} y_k,$$

is an unbiased estimator of t_y with variance

$$V\left(\hat{t}_{A\Pi\text{PS}}\right) = E\left(\sum_{k \in U^*} \sum_{\ell \in U^*} \frac{y_k}{\Pi_k^*} \frac{y_\ell}{\Pi_\ell^*} \Delta_{k\ell}^*\right),$$

and provided that all the $\Pi_{k\ell}^*$ are strictly positive (which depends on the sampling design used in U^*), an unbiased estimator of variance is

$$\hat{V}\left(\hat{t}_{A\Pi\text{PS}}\right) = \sum_{k \in s^*} \sum_{\ell \in s^*} \frac{y_k}{\Pi_k^*} \frac{y_\ell}{\Pi_\ell^*} \frac{\Delta_{k\ell}^*}{\Pi_{k\ell}^*},$$

where $\Delta_{k\ell}^* = \Pi_{k\ell}^* - \Pi_k^* \Pi_\ell^*$ and $\Pi_{k\ell}^* = E\left(I_k I_\ell \mid s_0\right)$.

For the proof of Result 2, see Appendix B.

In $\hat{t}_{A\Pi\text{PS}}$, for extreme cases where the size of s^* is too small, we may exaggerate the role of s^* in estimation relative to s_0. Then we can adjust the estimator by adding a weighting parameter, say $0 \le \alpha \le 1$ as follows:

$$\hat{t}_{A\Pi\text{PS}\alpha} = N\left[\alpha \frac{1}{N-n_0} \sum_{k\in s^*} \frac{y_k}{\Pi_k^*} + (1-\alpha)\frac{1}{n_0}\sum_{k\in s_0} y_k\right].$$

Result 3. *In* $A\Pi\text{PS}\alpha$, *if we select* s_0 *by SRSWOR, then*

(i) $\hat{t}_{A\Pi\text{PS}\alpha} = \hat{t}_{A\Pi\text{PS}}$, *for* $\alpha = (N-n_0)/N$,

(ii) $\hat{t}_{A\Pi\text{PS}\alpha}$ *is unbiased,* $E\left(\hat{t}_{A\Pi\text{PS}\alpha}\right) = t_y$,

(iii) *with the following variance*

$$\begin{aligned} V\left(\hat{t}_{A\Pi\text{PS}\alpha}\right) &= N^2 \frac{\alpha^2}{(N-n_0)^2} E\left(\sum_{k\in U^*}\sum_{\ell\in U^*} \frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right) \\ &+ N^2\left(1-\frac{\alpha}{1-f_0}\right)^2\left(\frac{1-f_0}{n_0}S_y^2\right), \end{aligned}$$

where $f_0 = n_0/N$,

(iv) *and an unbiased estimator of the variance, provided that all the* $\Pi_{k\ell}^*$ *are strictly positive, is*

$$\begin{aligned} \hat{V}\left(\hat{t}_{A\Pi\text{PS}\alpha}\right) &= N^2 \frac{\alpha^2}{(N-n_0)^2} \sum_{k\in s^*}\sum_{\ell\in s^*} \frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\frac{\Delta_{k\ell}^*}{\Pi_{k\ell}^*} \\ &+ N^2\left(1-\frac{\alpha}{1-f_0}\right)^2\left(\frac{1-f_0}{n_0}s_{0y}^2\right), \end{aligned}$$

where

$$S_y^2 = \frac{1}{N-1}\sum_{k\in U}\left(y_k - \bar{y}_U\right)^2 \quad \textit{and} \quad s_{0y}^2 = \frac{1}{n_0-1}\sum_{k\in s_0}\left(y_k - \bar{y}_{s_0}\right)^2,$$

(v) *The optimal value for* α *to minimize the variance of the estimator is*

$$\alpha^* = (1-f_0)\frac{\frac{1-f_0}{n_0}S_y^2}{E\left(\frac{1}{N^2}\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right) + \frac{1-f_0}{n_0}S_y^2}. \tag{3.2}$$

4. Simulations for APPS Sampling

In order to evaluate the designs, we have run simulations on real data. All the simulations in Section 4 and Section 5 have been implemented using Monte Carlo methods with 2,000 iterations. We use a real case study of medicinal flowers and real data from the statistical center of Iran between 2015-2016 (https://www.amar.org.ir) to evaluate the results of Section 2. To compare the designs, the efficiency is defined by

$$\text{Efficiency} = F_{\bullet} = \frac{V\left(N\bar{y}_s\right)}{V\left(\hat{t}_{\bullet}\right)}, \tag{4.1}$$

where $\bar{y}_s$ is the sample mean in Simple Random Sampling with-replacement (SRSWR) with size n and $\hat{t}_{\bullet}$ indicates the Hansen-Hurwitz estimator in PPS, APPS with n draws or π -estimator in ΠPS. In each case, we indicate the variable of interest and the auxiliary variable. Drawing probabilities for PPS and APPS are calculated based on (2.1) and (2.2) respectively. Also inclusion probabilities for ΠPS are calculated based on the auxiliary variable using (3.1). For inclusion probabilities in (3.1) we used U and m instead of U^* and n^* respectively. As APPS and ΠPS are with and without-replacement designs respectively, in order to have a fair comparison, the cost of the sample needs to be as equal as possible for all of the designs. For this purpose, in each iteration an APPS is implemented first, and then for ΠPS, the sample size, is set to the number of distinct units obtained with n draws in APPS. To implement the ΠPS in this section we used the eliminatory method based on `UPtille` function available in the `R` package `sampling` (Tillé and Matei, 2015).

For the simulations, we considered two kinds of data:

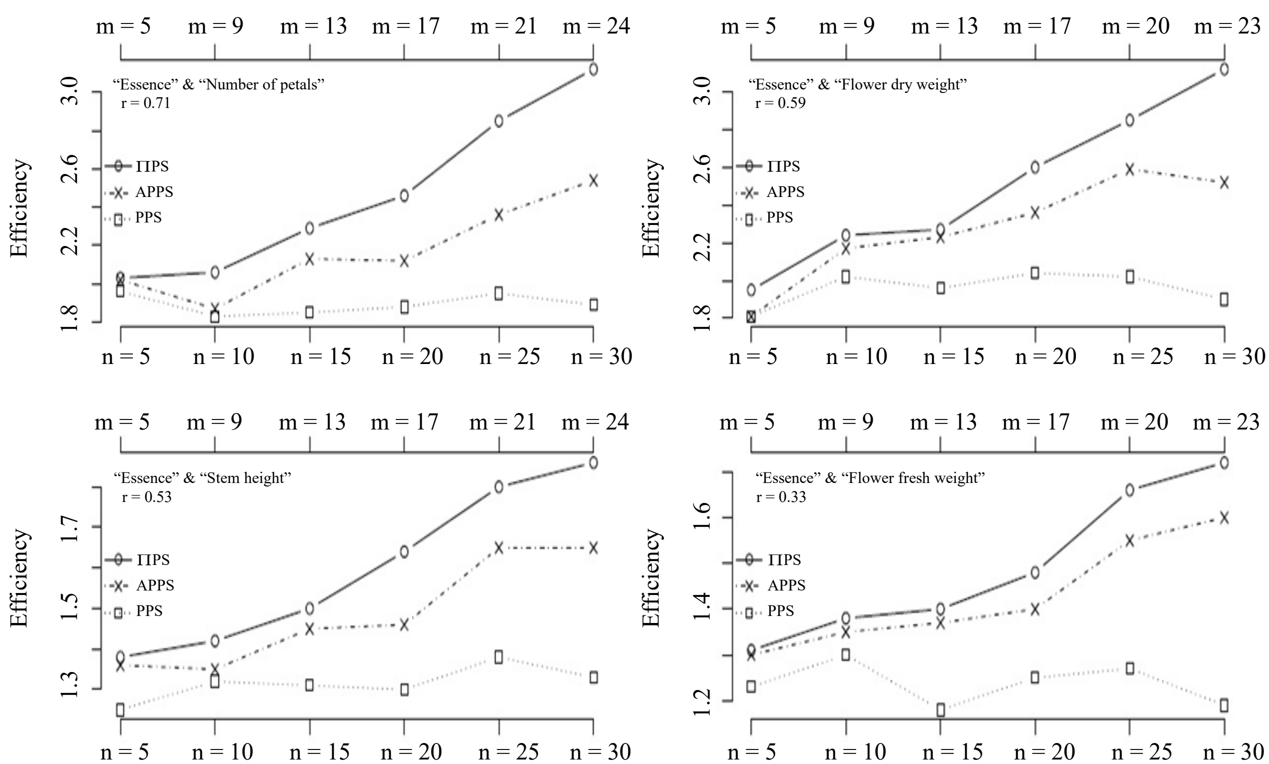
1. *Medicinal Flowers*: The data come from a real case study on chamomile flower (Panahbehagh, Bruggemann, Parvardeh, Salehi and Sabzalian, 2018) as the medicinal use of flowers. The population mean of the "Essence" is the parameter of interest with $t_y = 44.4$ and $N = 60$. In practice, the variable of interest is not know prior to sampling so we use four readily available auxiliary variables with various correlations with the variable of interest. The four auxiliary variables and correlations are "Flower fresh weight" with 0.33, "Flower dry weight" with 0.59, "Stem height" with 0.53 and "Number of petals" with 0.71. The results are presented in Figure 4.1, where the correlations are denoted by r.
2. *Social Data*: These data are from the Statistical Center of Iran gathered from 31 provinces of Iran in 2015-2016 (www.amar.org.ir). Marriage-Divorce and Academic degrees data are official statistics covering all target populations, based on the "National Organization for Civil Registration" and the "Ministry of Science, Research and Technology" respectively. In addition, the provincial population sizes are based on the 2016 census in Iran. We considered four situations having an auxiliary and a variable of interest:
   - The registered number of "Divorce less than 1 year" and "Marriage" as the variable of interest and the auxiliary variable respectively,
   - The registered number of "Divorce less than 1 year" and "Divorce" as the variable of interest and the auxiliary variable respectively,

– The registered number of “Bachelors” and “Diplomas” as the variable of interest and the auxiliary variable respectively,
– The registered number of “Masters and higher” and “Diplomas” as the variable of interest and the auxiliary variable respectively.

The results are presented in Figure 4.2.

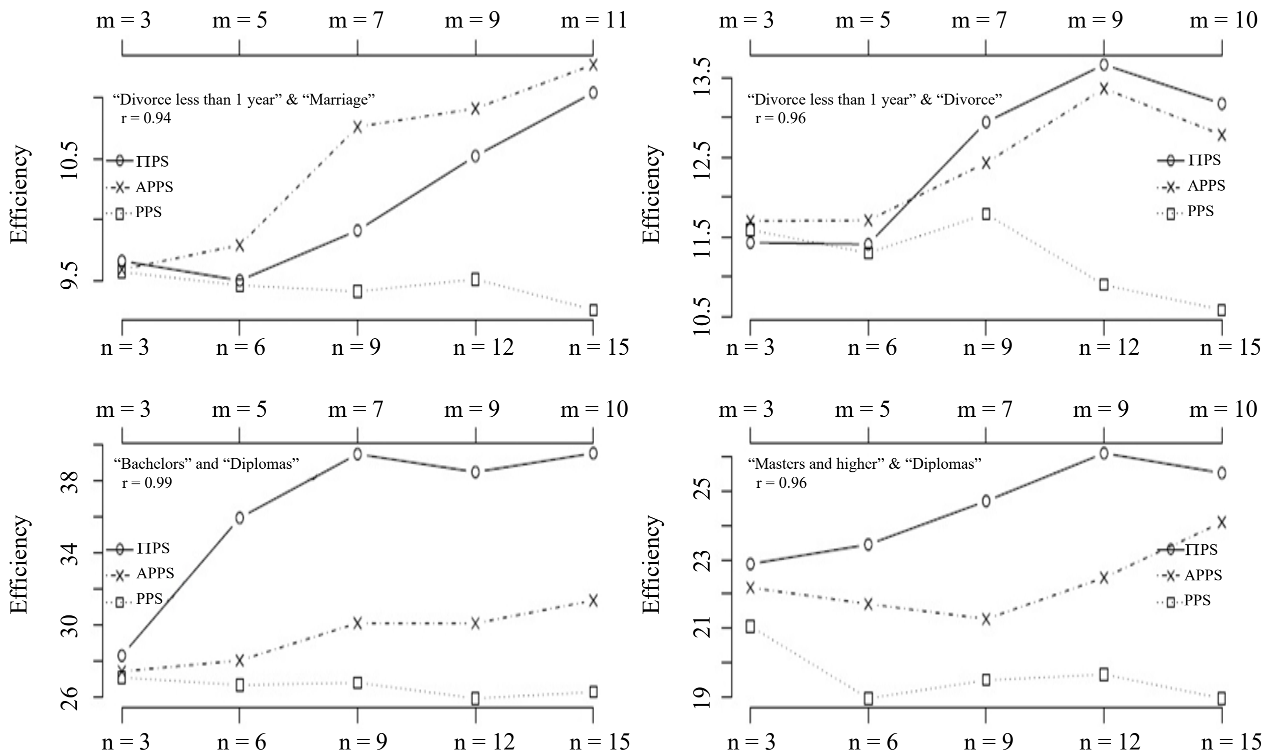
As can be seen in Figure 4.1, comparing the vertical axis, generally the higher the correlation, the higher the efficiency. By comparing Figure 4.2 and Figure 4.1, the efficiency increases dramatically for the social data compare to the medicinal flowers, which was predictable given the correlations of more than 0.90 in the former. A positive relationship between correlation and efficiency was expected because when the correlation is high, the drawing probabilities vector approximates the exact sampling probability proportional to size more accurately.

Figure 4.1 Efficiencies of ΠPS, APPS and PPS relative to SRSWR for the medicinal flowers data with different auxiliary variables. *m* is the size of ΠPS which is the Monte Carlo expectation of the number of distinct units in the respective with-replacement designs (PPS and APPS) of size *n*. At the top-left of each plot, the variable of interest and the auxiliary variable are indicated, with the respective Pearson correlation, indicated by *r*.



Note : Simple random sampling with-replacement (SRSWR); Proportional to size (PS); Adaptive probability proportional to size sampling (APPS); Probability proportional to size sampling (PPS).

Figure 4.2 Efficiencies of ΠPS, APPS and PPS relative to SRSWR for the social data with different auxiliary variables and different correlations. *m* is the size of ΠPS which is the Monte Carlo expectation of the number of distinct units in the respective with-replacement designs (PPS and APPS) of size *n*. At the top-left of each plot, the variable of interest and the auxiliary variable are indicated, with the respective Pearson correlation, indicated by *r*.



Note : Simple random sampling with-replacement (SRSWR); Proportional to size (PS); Adaptive probability proportional to size sampling (APPS); Probability proportional to size sampling (PPS).

In Figure 4.1, for the medicinal flowers data, APPS is more efficient than PPS in all cases. The efficiency of PPS fluctuates slightly with the variation of n, which shows that increasing the sample size, improves both SRSWR and PPS at the same level. But at the same time, the efficiency of APPS generally increases with increasing sample size. In fact, in APPS, the larger the sample size, the more updated the auxiliary variable units, and therefore the more accurate the exact proportional to size approximation. Furthermore, although the efficiency of APPS is much closer to ΠPS compared to PPS in most cases, the efficiency of ΠPS, particularly for large sample sizes, is higher than the other two in all Figure 4.1 cases and discrepancy increases with increasing n. As a final point in Figure 4.1, the efficiency of APPS is often about the same as ΠPS if the sample size is less than around 15% of the population size (a reasonable sample size).

Most of the results in Figure 4.2 are similar to the results in Figure 4.1. For the social data, like the medicinal flowers data, APPS is more efficient than PPS in all cases and the efficiency of APPS generally increases (with some fluctuations) as the sample size increases. But interestingly, unlike the medicinal flowers, APPS is more efficient than ΠPS in some cases. This is interesting because it is much easier to implement and calculate the estimators in APPS than in ΠPS. Eventually, the efficiency of PPS fluctuates again, but this time it tends to fall slightly as n increases.

5. Simulations for $A\Pi PS$

Following the notation used in (4.1), "•" indicates the particular strategy, ΠPS, $A\Pi PS$ or $A\Pi PS\alpha$ and $\bar{y}_s$ is the sample mean in SRSWOR with size n. Regarding the note in step 2 of Algorithm 2 related to negative values of $\hat{y}$, we replace them with 0.0001 in the simulations.

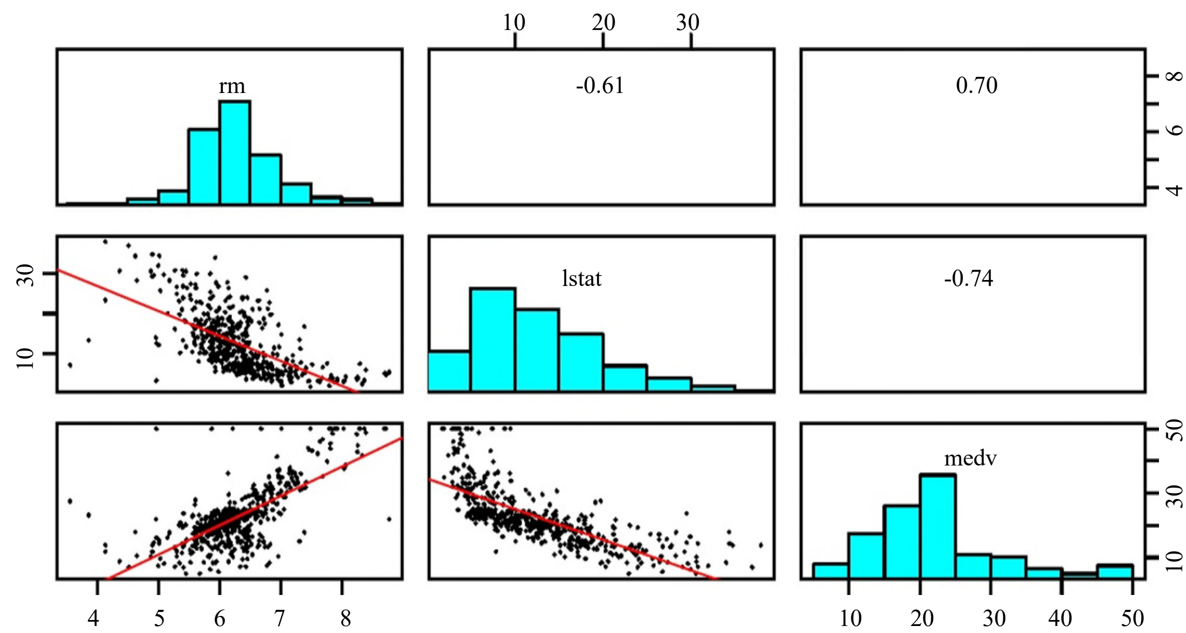
Also for ΠPS and step 3 of Algorithm 2 in $A\Pi PS$, we used the maximum entropy design based on `UPmaxentropy` function available in the `R` package `sampling` (Tillé and Matei, 2015).

5.1 Boston data

In this subsection, we analyze a dataset for the city of Boston (see Figure 5.1). Three different housing value variables for suburbs of Boston (Harrison and Rubinfeld, 1978; Belsley, Kuh and Welsch, 1980) are available in `R` package `MASS` as:

- rm: Average number of rooms per dwelling,
- lstat: Percentage of population in weak and deprived economic situation in Boston Suburbs,
- medv: Median value of owner-occupied homes in 1,000$s.

Figure 5.1 The relationship among three variables of interest, rm, lstat and medv, for the Boston data. In this 3×3 matrix of plots, the lower off-diagonal draws scatter plots with fitted linear least squares regressions, the diagonal represents histograms with the name of the variables and the upper off-diagonal reports the Pearson correlations.



Note : rm: Average number of rooms per dwelling; lstat: Percentage of population in weak and deprived economic situation in Boston Suburbs; medv: Median value of owner-occupied homes in 1,000$s.

In urban and residential areas, the larger the dimensions of a house, the more rooms one can expect to have. Also, the larger the dimensions of the house, the higher the value of the house. Therefore, there is a positive relationship between the dimensions of houses and the average number of rooms in each house and a positive relationship between the average number of rooms and the value of the house. In addition, economically disadvantaged people typically live in smaller houses, so the higher the proportion of disadvantaged people in a residential area, the greater the demand for small houses, and therefore the average number of rooms per house in that area will be lower. It follows that in a residential area there will be a negative relationship between the proportion of disadvantaged people and the average number of rooms in each house.

To model the variable of interest y on the auxiliary variable $x,$ we used

$$\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 + \hat{\beta}_3 x_k^3, \tag{5.1}$$

where the coefficients are estimated based on the least squares error method. Here based on Result 3, the estimator of the optimal value of α given in (3.2) is used, where $\hat{S}_y^2 = s_{0y}^2$ and

$$\hat{E}\left(\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right) = \sum_{k\in s^*}\sum_{\ell\in s^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\frac{\Delta_{k\ell}^*}{\Pi_{k\ell}^*}.$$

The results of the simulations on the Boston data are presented in Table 5.1. In all cases, $\text{A}\Pi\text{PS}\alpha$ is better than $\text{A}\Pi\text{PS}$ and, in almost all cases, is more efficient than SRSWOR (except for some cases with small n and n_0). Also, for $\text{A}\Pi\text{PS}$ and $\text{A}\Pi\text{PS}\alpha$ the efficiency generally increases with increasing n and n_0. In each model, the R^2's for different cases fluctuated slightly around a certain value, and predictably, the values appear to be independent of the initial sample size. Then, we have only reported the median values of the R^2's for different cases in the table.

As expected, the higher the absolute value of the correlation between x and $y,$ the higher the R^2. Consequently, as $\text{A}\Pi\text{PS}$ and $\text{A}\Pi\text{PS}\alpha$ use the regression model to predict y values, the higher the R^2, the higher the efficiencies of $\text{A}\Pi\text{PS}$ and $\text{A}\Pi\text{PS}\alpha$. Furthermore, ΠPS is better than SRSWOR only for rm-medv, which has a positive and almost linear relationship with some outliers, and ΠPS is less efficient than SRSWOR for the other two models with a negative (albeit strong) correlations. Due to the use of a regression model, $\text{A}\Pi\text{PS}$ and $\text{A}\Pi\text{PS}\alpha$ are not affected by the sign of the correlations. In the rm-medv model, ΠPS is better than the others but for large sample size, $\text{A}\Pi\text{PS}\alpha$ could approach ΠPS.

Looking into Monte Carlo's results in detail, $\text{A}\Pi\text{PS},$ by exaggerating the role of s^* (as discussed in Section 3) in certain iterations, results in very biased estimates for the parameter. Since, it cannot be as efficient as SRSWOR for medv-rm and lstat-rm model, in the next simulation on economic data, we simply compare the efficiencies of $\text{A}\Pi\text{PS}\alpha$ and ΠPS designs.

Table 5.1
Efficiencies of ΠPS, AΠPS and AΠPSα with $M = 3$ for Boston data. For each case, the variable of interest y and the auxiliary variable x are specified. Initial and final sample sizes are denoted by n_0 and n respectively, F indicates efficiency and R^2 is R-squared of model $\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 + \hat{\beta}_3 x_k^3$

	n	n_0	R^2	$F_{\Pi PS}$	$F_{A\Pi PS}$	$F_{A\Pi PS\alpha}$
x = medv, y = rm, N = 506	50	15	0.63	1.65	0.59	0.83
		20			0.70	1.05
		25			0.62	1.22
	75	20		1.65	0.81	1.01
		30			0.84	1.27
		40			0.89	1.43
	100	25		1.50	0.96	1.15
		30			0.78	1.34
		50			0.89	1.36
x = stat, y = rm, N = 506	50	15	0.49	0.72	0.62	0.91
		20			0.40	1.03
		25			0.63	1.05
	75	20		0.71	0.77	0.97
		30			0.77	1.10
		40			0.71	1.19
	100	25		0.76	0.93	1.15
		30			0.86	1.23
		50			0.79	1.18
x = medv, y = lstat, N = 506	50	15	0.70	0.11	1.20	1.51
		20			1.12	1.51
		25			1.08	1.64
	75	20		0.11	1.43	1.78
		30			1.41	1.93
		40			1.23	1.76
	100	25		0.10	1.40	1.91
		30			1.42	1.79
		50			1.41	1.93

Note : rm: Average number of rooms per dwelling; lstat: Percentage of population in weak and deprived economic situation in Boston Suburbs; medv: Median value of owner-occupied homes in 1,000$s.

5.2 Economic data

Data from four different economic variables for 180 countries, partially available from 1980 to 2006, were used to evaluate the results of Section 3. The data are collected on the website of the World Bank (2021). The four variables considered in this simulation are:

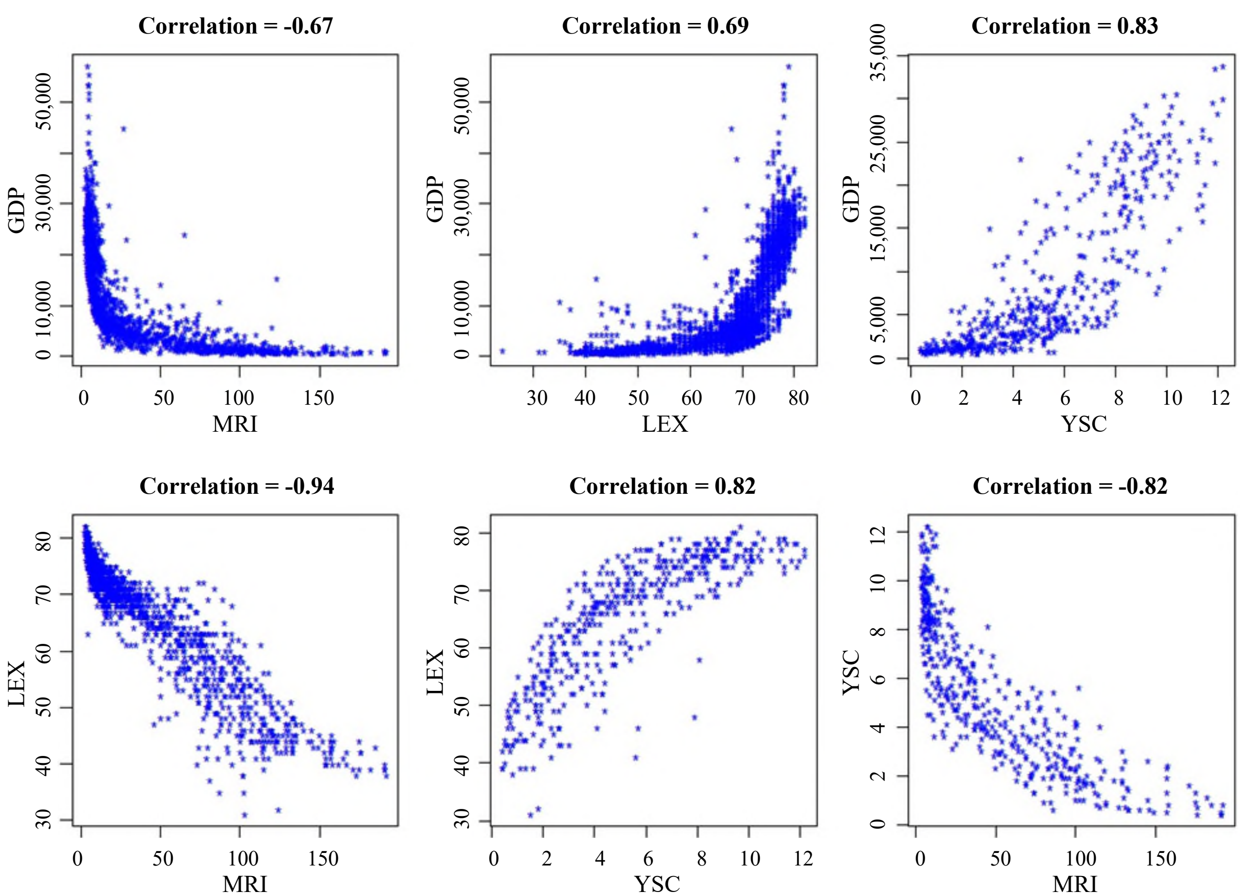
- GDP: Gross domestic product per capita based on purchasing power parity. GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2,000 international dollars.
- MRI: Mortality rate of infants per 1,000 births is the number of infants dying before reaching one year of age.

- LEX: Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.
- YSC: Average schooling years in the total population aged over 25.

One of the factors of production is human resources, and the higher the quality and quantity of human resources, the higher the productivity and output of the economy. The quality of human resources can be enhanced by improving their health and well-being. Improving healthcare leads to increased life expectancy and reduced mortality. In addition, the training of human resources leads to their promotion in the fields of science and technology. Therefore, live expectancy and average years of schooling have a positive relationship with GDP per capita, and mortality rate has a negative relationship with GDP per capita.

The relationship among the four variables are presented in Figure 5.2. The population size N varies for different pairs due to the exclusion of missing data.

Figure 5.2 The relationship among the four variables in economic data. Scatter plots for the variables are shown with the Pearson correlations for the two variables at the top of each plot.



Note : Mortality rate of infants (MRI); Life expectancy at birth (LEX); Average schooling years (YSC); Gross domestic product (GDP).

The results presented in Table 5.2 can be summarized as follows:

- $\text{A}\Pi\text{PS}\alpha$ is more efficient than SRSWOR in all cases, but ΠPS is very inefficient for cases with non-linear or negative relationships. In all cases, except for model YSC-GDP, $\text{A}\Pi\text{PS}$ is more efficient than Π PS.
- MRI-GDP and LEX-GDP show almost the same pattern but with different signs. $\text{A}\Pi\text{PS}\alpha$ is efficient in both of them and is more efficient in the model with higher absolute correlation. But ΠPS is efficient for the positive relationship (LEX-GDP) and very inefficient in the negative relationship (MRI-GDP).
- For YSC-LEX, although the relationship is positive and almost linear (with $R^2 = 0.82^2 = 0.67$ for $\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k$), but contrary to $\text{A}\Pi\text{PS}\alpha$ which is an efficient design, ΠPS is very inefficient compared to SRSWOR.
- The correlations in both models YSC-LEX and MRI-YSC are the same, but according to R^2, it seems that regression equation (5.1) can predict $\hat{y}$ in the latter model better than the former model. Therefore the mean of the efficiencies in model MRI-YSC (2.93) is higher than model YSC-LEX (2.47).
- MRI-LEX has the highest R^2, and for large initial and final sample sizes, the efficiency of $\text{A}\Pi\text{PS}\alpha$ is the highest compared to other relationships with the same initial and final sample sizes.
- In general, increasing the sample size leads to an increase in the efficiency of $\text{A}\Pi\text{PS}\alpha$.
- For $\text{A}\Pi\text{PS}\alpha$, in all cases (except model MRI-GDP), the highest efficiency (on average) is for the largest sample size $(n = 150)$.

Table 5.2
Efficiencies of ΠPS and AΠPS with $M = 3$ for Economic data. For each case, the variable of interest y and the auxiliary variable x are specified. Initial and final sample sizes are denoted by n_0 and n respectively, F indicates efficiency and R^2 is R-squared of model $\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 + \hat{\beta}_3 x_k^3$. The population size N, is different for different pairs due to the exclusion of missing data

	n	n_0	R^2	$F_{\Pi\text{PS}}$	$F_{A\Pi\text{PS}\alpha}$
x = MRI, y = GDP, N = 1,522	80	20	0.70	0.07	2.18
		30			2.30
		40			2.35
	100	30		0.08	2.39
		40			2.12
		50			2.13
	150	40		0.07	1.88
		60			1.94
		80			2.22
x = MRI, y = LEX, N = 1,619	80	20	0.85	0.01	1.32
		30			2.67
		40			2.94
	100	30		0.01	2.02
		40			2.81
		50			2.89
	150	40		0.01	2.62
		60			4.17
		80			3.22

Note : Mortality rate of infants (MRI); Gross domestic product (GDP); Life expectancy at birth (LEX); Average schooling years (YSC).

Table 5.2 (continued)
Efficiencies of ΠPS and AΠPS with $M = 3$ for Economic data. For each case, the variable of interest y and the auxiliary variable x are specified. Initial and final sample sizes are denoted by n_0 and n respectively, F indicates efficiency and R^2 is R-squared of model $\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_k + \hat{\beta}_2 x_k^2 + \hat{\beta}_3 x_k^3$. The population size N, is different for different pairs due to the exclusion of missing data

	n	n_0	R^2	$F_{\Pi\text{PS}}$	$F_{A\Pi\text{PS}\alpha}$
$x =$ LEX, $y =$ GDP, $N =$ 2,357	80	20	0.76	1.31	2.46
		30			2.68
		40			2.38
	100	30		1.29	2.63
		40			2.57
		50			2.14
	150	40		1.41	2.86
		60			2.58
		80			2.51
$x =$ YSC, $y =$ LEX, $N =$ 452	80	20	0.75	0.08	2.29
		30			2.71
		40			2.30
	100	30		0.08	2.66
		40			2.34
		50			2.50
	150	40		0.07	2.82
		60			2.54
		80			2.10
$x =$ YSC, $y =$ GDP, $N =$ 487	80	20	0.71	3.46	2.30
		30			2.33
		40			2.23
	100	30		4.10	2.87
		40			2.97
		50			2.72
	150	40		7.04	4.49
		60			3.88
		80			2.99
$x =$ MRI, $y =$ YSC, $N =$ 428	80	20	0.78	0.05	2.82
		30			2.89
		40			2.56
	100	30		0.04	2.73
		40			2.94
		50			2.63
	150	40		0.04	3.65
		60			3.26
		80			2.90

Note : Mortality rate of infants (MRI); Gross domestic product (GDP); Life expectancy at birth (LEX); Average schooling years (YSC).

6. Conclusions

Two adaptive versions of PS, with- and without- replacement have been presented. Both versions are based on information observed in the process of sampling, and help the sampler to obtain a more efficient sample, leading to more accurate estimates. Compared to the conventional versions, these adaptive versions of PS require no additional information and only need time to analyze the initial sample to decide on the next steps in the sampling process.

APPS is easy to implement, more efficient than its conventional version, PPS, and sometimes more efficient than ΠPS. The simulations show that APPS is always more efficient than the PPS. In addition, increasing the sample size gives APPS the ability to update more units of the auxiliary variable, resulting in increased efficiency. Besides these advantages, APPS has two weaknesses: the design is with-replacement and the sample units must be selected one by one in order.

On the other hand, $\text{A}\Pi\text{PS}$ is a without-replacement design that must be implemented in two phases. In the first phase, an initial sample is selected and y is modeled on x and in the second phase, the final sample is selected based on the predicted y values. The relationship between x and y is modeled using the sample information based on Taylor expansion theory in the first phase of sampling. Next a proportional to size scheme is used in the second phase of sampling. The simulations confirm that $\text{A}\Pi\text{PS}$ is an efficient and reliable design.

Acknowledgements

The authors would like to thank two referees and the associate editor for constructive and very insightful comments that helped them improve this article.

Appendix A

Proof of Result 1

To prove that $\sum_{k \in U} p_{ik} = 1,$ we have

$$\begin{aligned}\sum_{k \in U} p_{ik} &= \sum_{k \in s_{i-1}} p_{ik} + \sum_{k \in U \setminus s_{i-1}} p_{ik} = \sum_{k \in s_{i-1}} \frac{y_k \, t_{xs_i}}{t_x \, t_{ys_i}} + \sum_{k \in U \setminus s_{i-1}} \frac{x_k}{t_x} \\ &= \frac{t_{xs_i}}{t_x t_{ys_i}} t_{ys_i} + \frac{t_{xU \setminus s_i}}{t_x} = \frac{t_{xs_i}}{t_x} + \frac{t_{xU \setminus s_i}}{t_x} = 1.\end{aligned}$$

For unbiasedness of $\hat{t}_{\text{APPS}},$ we have

$$\hat{t}_{\text{APPS}} = \frac{1}{n} \sum_{i=1}^{n} Z_i,$$

where $Z_i = y_{k_i} / p_{ik_i}$ and $Z_i \mid s_{i-1} \sim f_i(z),$ with

$$f_i(z) = \begin{cases} p_{ik} & \text{if } \; z = \dfrac{y_k}{p_{ik}}, k = 1, \ldots, N \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$E(Z_i) = E\left[E\left(Z_i \mid s_{i-1}\right)\right] = E\left(\sum_{k=1}^{N} \frac{y_k}{p_{ik}} p_{ik}\right) = E\left(t_y\right) = t_y,$$

and therefore

$$E\left(\hat{t}_{\text{APPS}}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left[E\left(Z_i \middle| s_{i-1}\right)\right] = t_y.$$

For the variance, we have

$$V\left(\hat{t}_{\text{APPS}}\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} V\left(Z_i\right) + \sum_{i=1}^{n}\sum_{j \neq i=1}^{n} C\left(Z_i, Z_j\right)\right],$$

where C is a symbol for covariance function. Then

$$\begin{aligned} V\left(\hat{t}_{\text{APPS}}\right) &= \frac{1}{n^2}\left\{\sum_{i=1}^{n}\left[EV\left(Z_i \middle| s_{i-1}\right) + VE\left(Z_i \middle| s_{i-1}\right)\right]\right. \\ &\quad \left. + \; 2\sum_{i=1}^{n}\sum_{j<i}^{n}\left(EC\left(Z_i, Z_j \mid s_{i-1}\right) + C\left[E\left(Z_i \middle| s_{i-1}\right), E\left(Z_j \middle| s_{i-1}\right)\right]\right)\right\}. \quad \text{(A.1)} \end{aligned}$$

Now, we have

$$\begin{aligned} V\left[E\left(Z_i \middle| s_{i-1}\right)\right] &= V\left(t_y\right) = 0, \\ C\left[E\left(Z_i \middle| s_{i-1}\right), E\left(Z_j \middle| s_{i-1}\right)\right] &= C\left(t_y, Z_j\right) = 0, \end{aligned}$$

and

$$E\left[V\left(Z_i \middle| s_{i-1}\right)\right] = E\left[\sum_{k=1}^{N}\left(\frac{y_k}{p_{ik}} - t_y\right)^2 p_{ik}\right].$$

Also for $i > j$, Z_j is constant given s_{i-1}, and therefore $C\left(Z_i, Z_j \mid s_{i-1}\right) = 0.$ Finally, using (A.1) we have

$$V\left(\hat{t}_{\text{APPS}}\right) = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{N}\left(\frac{y_k}{p_{ik}} - t_y\right)^2 p_{ik}\right].$$

Moreover,

$$\hat{V}\left(\hat{t}_{\text{APPS}}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_{k_i}}{p_{ik_i}} - \hat{t}_{\text{APPS}}\right)^2,$$

is an unbiased estimator for $V\left(\hat{t}_{\text{APPS}}\right)$, because

$$\sum_{i=1}^{n}\left(\frac{y_{k_i}}{p_{ik_i}} - \hat{t}_{\text{APPS}}\right)^2 = \sum_{i=1}^{n}\left(\frac{y_{k_i}}{p_{ik_i}}\right)^2 - n\hat{t}_{\text{APPS}}^2,$$

and

$$\begin{aligned} E\left(\frac{y_{k_i}}{p_{ik_i}}\right)^2 &= V\left(\frac{y_{k_i}}{p_{ik_i}}\right) + E^2\left(\frac{y_{k_i}}{p_{ik_i}}\right) \\ &= EV\left(\frac{y_{k_i}}{p_{ik_i}} \mid s_{i-1}\right) + VE\left(\frac{y_{k_i}}{p_{ik_i}} \mid s_{i-1}\right) + t_y^2 \\ &= E\left[\sum_{k=1}^{N}\left(\frac{y_k}{p_{ik}} - t_y\right)^2 p_{ik}\right] + 0 + t_y^2, \\ E\left(\hat{t}^2_{\text{APPS}}\right) &= V\left(\hat{t}_{\text{APPS}}\right) + E^2\left(\hat{t}_{\text{APPS}}\right) = V\left(\hat{t}_{\text{APPS}}\right) + t_y^2, \end{aligned}$$

then

$$\begin{aligned} E\left[\sum_{i=1}^{n}\left(\frac{y_{k_i}}{p_{ik_i}} - \hat{t}_{\text{APPS}}\right)^2\right] &= E\left[\sum_{i=1}^{n}\sum_{k=1}^{N}\left(\frac{y_k}{p_{ik}} - t_y\right)^2 p_{ik}\right] + nt_y^2 - nV\left(\hat{t}_{\text{APPS}}\right) - nt_y^2 \\ &= n^2 V\left(\hat{t}_{\text{APPS}}\right) - nV\left(\hat{t}_{\text{APPS}}\right) = n(n-1)V\left(\hat{t}_{\text{APPS}}\right), \end{aligned}$$

and therefore $E\hat{V}\left(\hat{t}_{\text{APPS}}\right) = V\left(\hat{t}_{\text{APPS}}\right)$.

Appendix B

Proof of Result 2

To prove $\hat{t}_{A\Pi\text{PS}}$ is unbiased and driving its variance, we have

$$\begin{aligned} E\left(\hat{t}_{A\Pi\text{PS}}\right) &= E\left[E\left(\hat{t}_{A\Pi\text{PS}} \mid s_0\right)\right] = E\left(\sum_{k\in U^*}\frac{y_k}{\Pi_k^*}E\left(I_k^* \mid s_0\right) + \sum_{k\in s_0} y_k\right) \\ &= E\left(\sum_{k\in U^*} y_k + \sum_{k\in s_0} y_k\right) = E\left(t_y\right) = t_y, \end{aligned}$$

and

$$\begin{aligned} V\left(\hat{t}_{A\Pi\text{PS}}\right) &= EV\left(\hat{t}_{A\Pi\text{PS}} \mid s_0\right) + VE\left(\hat{t}_{A\Pi\text{PS}} \mid s_0\right) \\ &= E\left(\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right) + V\left(t_y\right) \\ &= E\left(\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right), \end{aligned}$$

respectively. Note that U^*, Π_k^* and $\Delta_{k\ell}^*$ are random based on the design.

Also, to prove that $\hat{V}\left(\hat{t}_{A\Pi\text{PS}}\right)$ is an unbiased estimator of the variance, we have

$$\begin{aligned} E\left[\hat{V}\left(\hat{t}_{A\Pi \text{PS}}\right)\right] &= E\left[E\left(\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\frac{\Delta_{k\ell}^*}{\Pi_{k\ell}^*}I_k I_\ell \mid s_0\right)\right] \\ &= E\left[\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\frac{\Delta_{k\ell}^*}{\Pi_{k\ell}^*}E\left(I_k I_\ell \mid s_0\right)\right] \\ &= E\left(\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right). \end{aligned}$$

Appendix C

Proof of Result 3

(i) This part of Result 3 will be easily proved by replacing α with $\left(N-n_0\right)/N$.

(ii) To prove that $E\left(\hat{t}_{A\Pi \text{PS}\alpha}\right)=t_y$, it is enough to note that when s_0 is a SRSWOR of size n_0, then $U \setminus s_0$ is a SRSWOR of size $N-n_0$. Therefore

$$\begin{aligned} E\left(\hat{t}_{A\Pi \text{PS}\alpha}\right) &= E\left[E\left(\hat{t}_{A\Pi \text{PS}\alpha} \mid s_0\right)\right] \\ &= NE\left(\frac{\alpha}{N-n_0}\sum_{k\in U\setminus s_0}\frac{y_k}{\Pi_k^*}E\left(I_k^* \mid s_0\right)+\frac{1-\alpha}{n_0}\sum_{k\in s_0}y_k\right) \\ &= N\left[\alpha E\left(\bar{y}_{U\setminus s_0}\right)+\left(1-\alpha\right)E\left(\bar{y}_{s_0}\right)\right]=N\bar{y}_U=t_y. \end{aligned}$$

(iii) For calculating the variance of $\hat{t}_{A\Pi \text{PS}\alpha}$ we have

$$\begin{aligned} V\left(\hat{t}_{A\Pi \text{PS}\alpha}\right) &= EV\left(\hat{t}_{A\Pi \text{PS}\alpha} \mid s_0\right)+VE\left(\hat{t}_{A\Pi \text{PS}\alpha} \mid s_0\right) \\ &= N^2E\left(\frac{\alpha^2}{\left(N-n_0\right)^2}\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right)+N^2V\left(\alpha\bar{y}_{U\setminus s_0}+\left(1-\alpha\right)\bar{y}_{s_0}\right) \\ &= N^2E\left[\frac{\alpha^2}{\left(N-n_0\right)^2}\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right]+N^2\left(1-\frac{\alpha}{1-f_0}\right)^2\left(\frac{1-f_0}{n_0}S_y^2\right). \end{aligned}$$

(iv) The proof of this part is the same as Result 2 and the fact that in SRSWOR, $E\left(s_{0y}^2\right)=S_y^2$.

(v) The optimal value α^* is obtained by calculating the derivative of $V\left(\hat{t}_{A\Pi \text{PS}\alpha}\right)$ with respect to α and solving the resulting equation

$$\frac{dV\left(\hat{t}_{A\Pi \text{PS}\alpha}\right)}{d\alpha} = 0,$$

which leads to

$$\frac{\alpha}{\left(N-n_0\right)^2}E\left[\sum_{k\in U^*}\sum_{\ell\in U^*}\frac{y_k}{\Pi_k^*}\frac{y_\ell}{\Pi_\ell^*}\Delta_{k\ell}^*\right] - \frac{1}{1-f_0}\left(1-\frac{\alpha}{1-f_0}\right)\left(\frac{1-f_0}{n_0}S_y^2\right) = 0.$$

Thus

$$\alpha = (1 - f_0) \frac{\frac{1-f_0}{n_0} S_y^2}{E\left(\frac{1}{N^2} \sum_{k \in U^*} \sum_{\ell \in U^*} \frac{y_k}{\Pi_k^*} \frac{y_\ell}{\Pi_\ell^*} \Delta_{k\ell}^*\right) + \frac{1-f_0}{n_0} S_y^2}.$$

In order to show that the calculated α minimizes the variance, it is easy to show that the second derivative of $V\left(\hat{t}_{A\Pi\text{PS}\alpha}\right)$, given $f_0 < 1$ and $S_y^2 > 0$, is strictly positive:

$$\frac{d^2 V\left(\hat{t}_{A\Pi\text{PS}\alpha}\right)}{d\alpha^2} = \frac{1}{\left(N - n_0\right)^2} E\left[\sum_{k \in U^*} \sum_{\ell \in U^*} \frac{y_k}{\Pi_k^*} \frac{y_\ell}{\Pi_\ell^*} \Delta_{k\ell}^*\right] + \frac{1}{\left(1 - f_0\right)^2} \left(\frac{1 - f_0}{n_0} S_y^2\right) > 0.$$

References

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, Inc.

Brewer, K.R.W., and Hanif, M. (1983). *Sampling with Unequal Probabilities*. New York: Springer.

Hansen, M.H., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.

Harrison, D., and Rubinfeld, D. (1978). Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81-102.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Madow, W.G. (1949). On the theory of systematic sampling, II. *Annals of Mathematical Statistics*, 20, 333-354.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3, 169-174.

Panahbehagh, B., Bruggemann, R., Parvardeh, A., Salehi, M.M. and Sabzalian, M.R. (2018). An unbalanced ranked set sampling to get more than one sample from each set. *Journal of Survey Statistics and Methodology*, 6, 256-305.

Seber, G.A.F., and Salehi, M.M. (2013). *Adaptive Sampling Designs*. Springer, Berlin, Heidelberg.

Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer.

Tillé, Y. (2020). *Sampling and Estimation from Finite Populations*. Wiley, Hoboken.

Tillé, Y., and Matei, A. (2015). *Sampling: Survey Sampling*. R package version 2.7.

World Bank (2021). World bank open database (free and open access to global development data). https://scholar.harvard.edu/files/shleifer/files/data_friedman.xls. Accessed: 2021-11-10.