



On the Maximum Likelihood Estimation of Population and Domain Means

Janusz L. Wywiał¹ 

Accepted: 10 June 2023 / Published online: 12 July 2023
© The Author(s) 2023

Abstract

Estimating of population and domain means based on model-design approaches is considered in this paper. Population elements randomly belong to domains. A joint distribution of the variable under study and an auxiliary variable is assumed. Data are observed in a sample selected from a fixed population. The partition of the sample elements into domains of the population is also known. Outside of the sample, values of the auxiliary variable are known but their partition among the domains is not known. The domain means are estimated based on the likelihood function of the data observed in the sample and outside of it. The maximum likelihood estimation method provides regression-type estimators of domain means of the variable under study. They are dependent on posterior probabilities that observations of the auxiliary variable belong to particular domains. Moreover, the weighted means of the domain averages estimators are used to estimation of the population mean. The accuracy of the evaluated estimators and the ordinary estimator is compared using a simulation analysis. The results of this paper could be useful in economic, demographic and sociological surveys.

Keywords Domain mean · Maximum likelihood estimation · Regression estimator · Ratio estimator · Posterior probability

1 Introduction

Domains are usually treated as fixed and mutually disjoint subsets of the population. We consider the case when a population element belong to a domain with some probability. Therefore, the size of the domain is random. Our problem is estimation domain means and the population mean of a variable under study y on the basis of a

✉ Janusz L. Wywiał
janusz.wywial@ue.katowice.pl

¹ Department of Statistics, Econometrics and Mathematics, University of Economics in Katowice, Katowice, Poland

random sample selected from a whole population. In the population all values of an auxiliary variable x are known, while values of y are observed only in the sample. In the sample domains are identifiable, while outside of the population they are not. Under the outlined assumptions let us consider the following motivation examples. Let us consider the population of firms that have taken out bank loans for investments. The values of granted loans are observations of the variable y , while the observations of the variable x are the values of companies' capital. Any company with the probability $p_h, h = 1, \dots, H$ may default on its loan. The h index identifies the domain consisted of companies classified approximately at the same credit risk. The total distribution of x, y variables in the population will be treated as a mixture of the distributions of these variables in domains weighted by the probabilities p_h . The empirical Sect. 3.2 of the article presents other examples.

Żądło [14] assumed that population elements randomly belong to domains. He presented several examples. For instance, he considered the estimation of the income of different enterprises when they randomly belonged to different investment intervals. Elections are another example of domains. In this case a domain consists of people who vote for a specific party. Often, the selection of a particular party is random because a lot of voters are not committed to voting for any particular party. The model for generating an accounting error (see [13]) can also exemplify domains. An observed value of an accounting document is treated as the outcome of the random variable, which is a mixture of two distribution functions. One of these distribution functions generates the true accounting amount, and the second generates an accounting amount contaminated with an error. Documents without errors belong to the first domain and documents polluted with accounting errors belong to the second domain. Hence, documents randomly belong to the domains. This idea, which is based on distribution mixtures, is developed in this paper.

Many auxiliary variables are usually observed during national censuses. Moreover, variables under study (observed during a census) can be used as auxiliary variables in survey sampling on a subsequent occasion. Therefore, we can expect these variables to be highly correlated. Let us note that apart from the above examples, there are many populations where all values of the auxiliary variable are observed. These can be found in economic, demographic, agricultural and other official registers.

In this paper, the model or model-randomization approaches are taken into account (see, e.g., [8] or [9]). Estimating domain means is usually better, when the estimation is supported by data on auxiliary variables observed outside the sample but usually under the assumption that their distribution among domains is known (see several monographs on Small Area Sampling and, e.g., [10]). The models formulated in this paper are close to those considered by Chambers and Skinner [2]. Estimators of domain averages are derived by means of the maximum pseudo-likelihood method. More precisely, a variant of the likelihood method of estimation based on incomplete data of the variable under study is adopted to estimate distribution mixture parameters. Our analysis is mainly supported by monographs [4, 6, 7].

The most important results of the paper are as follows:

- The pseudo-likelihood function is formulated for estimation of the mixture distribution parameters in the case when data are observed in the sample selected according various inclusion probabilities (Sect. 2.2).
- On the basis of this function, the regression and ratio type estimators of domain means are derived in the case of bivariate normal components of the distribution mixture (Sect. 3.1 and Appendix).
- These results are generalized into the case of a multidimensional auxiliary variable (Sect. 3.1).
- The linear combination of the regression (ratio) estimators is used to estimation the population mean (Sect. 3.1).
- Examples of the simulation analysis of the estimation accuracy are prepared (Sect. 3.2).

2 General Results

2.1 Model-Design Approach

Let us denote by U a population of size N partitioned into H mutually disjoint domains denoted by $U_h, h = 1, \dots, H, 1 < H < N$. Let $[y_k, \mathbf{x}_k, \mathbf{z}_{k*}]$ be the k -th observation of the variable under study, an auxiliary variable vector, and a vector identifying domains where $\mathbf{x}_k = [x_{k,1} \dots x_{k,m}]$, $1 \leq m < N$ and $\mathbf{z}_{k*} = [z_{k,1} \dots z_{k,h} \dots z_{k,H}]$, $k = 1, \dots, N$. Let $\mathbf{z}^{(h)}$ be a row vector, in which all H elements are equal to zero except the h -th element, which is equal to one, and this identifies the h -th domain. When $\mathbf{z}_{k*} = \mathbf{z}^{(h)}$, then the k -th population element is in the h -th domain.

Let us assume that $[y_k \mathbf{x}_k \mathbf{z}_{k*}]$ is an observation of a random vector $[Y_k \mathbf{X}_k \mathbf{Z}_{k*}]$ attached to a k -th population element, where $\mathbf{X}_k = [X_{k,1} \dots X_{k,m}]$ and $\mathbf{Z}_{k*} = [Z_{k,1} \dots Z_{k,H}]$. The random vectors $[Y_k \mathbf{X}_k \mathbf{Z}_{k*}], k \in U$, are independent, and each of them has the same probability distribution. Let $P(\mathbf{Z}_{k*} = \mathbf{z}^{(h)}) = p_h, h = 1, \dots, H, \sum_{h=1}^H p_h = 1$. Random variable \mathbf{Z}_{k*} has multinomial distribution with parameters $(1, p_1, \dots, p_H)$ (see, e.g., [7]). Event $\{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k\}$ with specific feature $\{\mathbf{Z}_{k*} = \mathbf{z}^{(h)}\}$ written as $\{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k, \mathbf{Z}_{k*} = \mathbf{z}^{(h)}\}$ concerns the h -th domain and

$$\{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k\} = \bigcup_{h=1}^H \{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k, \mathbf{Z}_{k*} = \mathbf{z}^{(h)}\}.$$

The events $\{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k, \mathbf{Z}_{k*} = \mathbf{z}^{(h)}\}$ and $\{Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k, \mathbf{Z}_{k*} = \mathbf{z}^{(t)}\}$ are mutually exclusive for all $h \neq t$ and $h = 1, \dots, H, t = 1, \dots, H$. This and the total probability theorem lets us write the following:

$$F(y_k, \mathbf{x}_k) = P(Y_k < y_k, \mathbf{X}_k < \mathbf{x}_k) = \sum_{h=1}^H F(y_k, \mathbf{x}_k | \mathbf{Z}_{k*} = \mathbf{z}^{(h)}) p_h$$

where $F(y_k, \mathbf{x}_k | \mathbf{Z}_{k*} = \mathbf{z}^{(h)})$ is the conditional distribution function. In the case where variables $[Y_k \mathbf{X}_k]$ are continuous, we have:

$$f(y_k, \mathbf{x}_k) = \sum_{h=1}^H f(y_k, \mathbf{x}_k | \mathbf{Z}_{k*} = \mathbf{z}^{(h)}) p_h$$

where $f(y_k, \mathbf{x}_k | \mathbf{Z}_{k*} = \mathbf{z}^{(h)})$, $h = 1, \dots, G$, $k \in U$, are density functions. This leads to the conclusion that our model defines the following distribution function: $F(y, \mathbf{x}) = \prod_{k \in U} F(y_k, \mathbf{x}_k)$ or density function $f(y, \mathbf{x}) = \prod_{k \in U} f(y_k, \mathbf{x}_k)$.

According to the assumptions of this model, the random variable $\sum_{k=1}^N \mathbf{Z}_{k*}$ has multinomial probability distribution with parameters $[N, p_1, \dots, p_H]$. Moreover, the column vector $\mathbf{Z}_{*h} = [Z_{1,h} \dots Z_{N,h}]^T$ identifies the h -th domain of size $N_h = \sum_{k \in U} Z_{k,h}$ where $0 \leq N_h \leq N$ and the expected domains sizes are $E(N_h) = Np_h$, $h = 1, \dots, H$ because N_h has binomial distribution with parameters (N, p_h) . Let us note that the introduced definitions lead to the conclusion that the sizes and consistencies of the domains are random. Hence, the multinomial probability model leads to partitions of the population into disjoint subsets called domains. Therefore, each outcome of partitioning the population into domains could be different.

Our main aim is to estimate the expected (domain mean) value $\mu_h = E(Y_k | \mathbf{Z}_{k*} = \mathbf{z}^{(h)})$ and the probabilities p_h , $h = 1, \dots, H$. Additionally, estimators of the expected value (population mean) $\mu = \sum_{h=1}^H p_h \mu_h$ are proposed.

In order to do this, sample s of size $n \leq N$ is selected from population U according to a sampling design denoted by $P(s) \geq 0$, $s \in \mathcal{S}$, where \mathcal{S} is sampling space and $\sum_{s \in \mathcal{S}} P(s) = 1$. Inclusion probabilities of the sampling design are defined by $\pi_k = \sum_{\{s: k \in s, s \in \mathcal{S}\}} P(s)$, $k = 1, \dots, N$. Let $\underline{s} = U - s$ be the complement of s in U . Moreover, let $s = \bigcup_{h=1}^H s_h$, where $s_h \subseteq U_h$, n_h is the size of s_h , $n = \sum_{h=1}^H n_h$ is size of s . We assume that $1 < n_h \leq N_h$ for $h = 1, \dots, H$. If $s = U$, then \underline{s} is the empty set.

2.2 Maximum Likelihood Estimation

Identifying a domain is possible after observation of variable \mathbf{Z}_{k*} in sample s . The density function of the conditional distribution of $[Y_k \mathbf{X}_k | \mathbf{Z}_{k*}]$ provided $\mathbf{Z}_{k*} = \mathbf{z}^{(h)}$ will be denoted by $f_h(y_k, \mathbf{x}_k, \theta_h)$, $h = 1, \dots, H$ where $\theta_h = [\theta_{h,1} \dots \theta_{h,m}]$, $\theta_h \subseteq R^m$, $\theta = [\theta_1 \dots \theta_h \dots \theta_H]$. Therefore, the observed values of the variables in the whole population are defined by the following distribution mixture:

$$f(y_k, \mathbf{x}_k, \Theta) = \sum_{h=1}^H p_h f_h(y_k, \mathbf{x}_k, \theta_h), \quad k \in U \tag{1}$$

where $\Theta = \{\mathbf{p} \cup \theta\}$, $\mathbf{p} = [p_1 \dots p_H]$. We assume that only values $\mathbf{x}_1, \dots, \mathbf{x}_k, \dots, \mathbf{x}_N$ are observed in the whole population before selecting a sample. The marginal

distribution of \mathbf{X}_k is as follows:

$$g(\mathbf{x}_k, \Theta_x) = \int_R f(y_k, \mathbf{x}_k, \Theta) dy_k = \sum_{h=1}^H p_h g_h(\mathbf{x}_k, \theta_{x,h}), \quad k \in U$$

where $g_h(\mathbf{x}_k, \theta_{x,h}) = \int_R f_h(y_k, \mathbf{x}_k, \theta_h) dy_k$, $\theta_{x,h} \subseteq \theta_h$ and $\Theta_x \subseteq \Theta$. Moreover, let: $\theta_x = [\theta_{x,1} \cdots \theta_{x,H}]$, $\Theta_x = \{\theta_x, \mathbf{p}\}$.

The sample contains the following data on variable values: $[y_k \mathbf{x}_k, \mathbf{z}_{k*}]$ of random variables $(Y_k \mathbf{X}_k, \mathbf{Z}_{k*})$, $k \in s$. Let $\mathbf{d}_s = \{[y_k \mathbf{x}_k, \mathbf{z}_{k*}], k \in s\}$ and $\mathbf{x}_s = \{\mathbf{x}_{k*}, k \in \underline{s}\}$. Hence, the sample contains complete data on the distribution mixture, while outside of the sample, the data are incomplete.

When the sample is selected according to preassigned inclusion probabilities, the pseudo-likelihood approach (see, [3, 8, 12]) leads to the following function:

$$l(\mathbf{d}_s, \mathbf{x}_s) = l_1(\mathbf{d}_s) + l_2(\mathbf{x}_s) \tag{2}$$

where the complete and incomplete log-likelihood functions are as follows, respectively:

$$\begin{cases} l_1(\mathbf{d}_s) = \sum_{h=1}^H \ln(p_h) \sum_{k \in s_h} \frac{1}{\pi_k} + \sum_{h=1}^H \sum_{k \in s_h} \frac{\ln(f_h(y_k, \mathbf{x}_k, \theta_h))}{\pi_k}, \\ l_2(\mathbf{x}_s) = \sum_{k \in \underline{s}} \frac{\ln(g(\mathbf{x}_k, \Theta_x))}{1 - \pi_k}. \end{cases} \tag{3}$$

where n_h is the size of $s_h \subseteq U_h$, which is the sub-sample of $s = \bigcup_{h=1}^H s_h$, $N_h \geq n_h > 1$, $n = \sum_{h=1}^H n_h$. We can easily show that $E_P(l_1(\mathbf{d}_s)) = l_1(\mathbf{d}_U)$ and $E_P(l_2(\mathbf{x}_s)) = l_2(\mathbf{x}_U)$ where

$$l_1(\mathbf{d}_U) = \sum_{h=1}^H N_h \ln(p_h) + \sum_{h=1}^H N_h \ln(f_h(y_k, \mathbf{x}_k, \theta_h)), \quad l_2(\mathbf{x}_U) = \sum_{k \in U} \ln(g(\mathbf{x}_k, \Theta_x)).$$

This means that the sample log-likelihood functions $l_1(\mathbf{d}_s)$ and $l_2(\mathbf{x}_s)$ are design-unbiased estimators of the population log-likelihood functions $l_1(\mathbf{d}_U)$ and $l_2(\mathbf{x}_U)$, respectively.

Usually, looking for the maximum of the log-likelihood function $l(\mathbf{d}_s, \mathbf{x}_s)$ is very complex and not exact. An approximation method has to be applied to solve the problem. Therefore, we use the more simple iteration method known as the EM-algorithm (see [4, 6, 7]). According to this method, function $l(\mathbf{d}_s, \mathbf{x}_s)$ is replaced with the following:

$$l^{(t)}(\mathbf{d}_s, \mathbf{x}_s) = l_1(\mathbf{d}_s) + l_2^{(t)}(\mathbf{x}_s) \tag{4}$$

where

$$l_2^{(t)}(\mathbf{x}_s) = \sum_{h=1}^H \tau_h^{(t)} \ln(p_h) + \sum_{h=1}^H \sum_{k \in \underline{s}} \frac{\tau_{h,k}^{(t)} \ln(g_h(\mathbf{x}_k, \theta_{x,h}))}{1 - \pi_k}, \tag{5}$$

$$\begin{cases} \hat{\tau}_h^{(t)} = \hat{\tau}_h(\hat{\Theta}_x^{(t)}) = \sum_{k \in \underline{s}} \frac{\tau_{h,k}^{(t)}}{1 - \pi_k}, \\ \tau_{h,k}^{(t)} = \tau_h(\mathbf{x}_k, \hat{\Theta}_x^{(t)}) = \frac{p_h g_h(\mathbf{x}_k, \hat{\Theta}_x^{(t)})}{g(\mathbf{x}_k, \hat{\Theta}_x^{(t)})}, \end{cases} \tag{6}$$

$\sum_{h=1}^H \tau_{h,k}^{(t)} = 1$ and $\hat{\tau}_{h,k}^{(t)}$ is the posterior probability that the k -element ($k \in \underline{s}$) belongs to the h -th domain. Moreover, $\hat{\tau}_h^{(t)}$ is the estimator of the expected size of the h -domain in the set \underline{s} . The Appendix provides outline of how to get optimal values of parameters $\hat{\Theta}_x^{(t+1)}$ and the following estimators of probabilities p_h :

$$\hat{p}_h^{(t+1)} = \frac{\hat{N}_h + \hat{\tau}_h^{(t)}}{\hat{N} + \hat{\tau}^{(t)}}, \quad h = 1, \dots, H. \tag{7}$$

where

$$\hat{N}_h = \sum_{k \in \underline{s}_h} \frac{1}{\pi_k}, \quad \hat{N} = \sum_{h=1}^H \hat{N}_h = \sum_{k \in \underline{s}} \frac{1}{\pi_k}, \quad \hat{\tau}^{(t)} = \sum_{h=1}^H \hat{\tau}_h^{(t)}.$$

Statistics \hat{N} and $\hat{\tau}^{(t)}$ are estimators of N . In general, estimators of $\hat{\Theta}^{(t+1)}$ could be obtained as roots of the first subsystem of the equation system (21). Moreover, $\tilde{N}_h^{(t)} = N \hat{p}_h^{(t)}$ is the estimator of the expected values of the domain size $N p_h$. The initial values of $\hat{\Theta}^{(t)}$ and $\hat{p}_h^{(t)}$ are equal to the roots of system $\frac{\partial l^{(t)}(\mathbf{d}_s, \mathbf{x}_s)}{\partial p_h} = \mathbf{0}$ and $\hat{p}_h^{(t)} = \frac{\tilde{N}_h}{N}, h = 1, \dots, H$.

When $\pi_k, k \in U$ depend on variables from \mathbf{X} , the likelihood function under the condition that $\mathbf{X} = \mathbf{x}$ needs to be consider. Several aspects of this problem were discussed by Pfeffermann [8] on the basis of large literature. Therefore, in order to simplify our considerations, we assume that the inclusion probabilities $\pi_k, k \in U$ as well as $p_h, h = 1, \dots, H$ could depend on the non-random auxiliary variable that is different from observations of variables from \mathbf{X} .

The simple random sample drawn without replacement does not depend on the auxiliary variables. In this case $\pi_k = \frac{n}{N}$ for $k \in U$, and the estimator expressed by (7) simplifies to the following form:

$$\hat{p}_h^{(t+1)} = \frac{1}{2}(\bar{p}_h + \bar{\tau}_h), \quad \bar{p}_h = \frac{n_h}{n}, \quad \bar{\tau}_h^{(t)} = \frac{1}{N - n} \sum_{k \in \underline{s}} \tau_{h,k}^{(t)}, \quad h = 1, \dots, H. \tag{8}$$

3 Estimation for a Bivariate Normal Model

3.1 Estimators

We assume that the components of the distribution mixture are two dimensional normal components with parameters: $N(\mu_{y,h}, \mu_{x,h}, \sigma_{y,h}^2, \sigma_{x,h}^2, \rho_h), h = 1, \dots, H$.

In the Appendix, we evaluate estimators of domain means $\mu_{y,h}$ and the fraction of the population elements in domains $p_h, h = 1, \dots, H$ according to the EM estimation algorithm and expressions (4)–(8). From Expressions (6) and (7), let us write for $t = 0, 1, 2, \dots$ the following:

$$\begin{cases} \tau_h^{(t)} = \sum_{k \in \underline{s}} \frac{\tau_{h,k}^{(t)}}{1 - \pi_k}, \\ \tau_{h,k}^{(t)} = \frac{\hat{p}_h^{(t)} g_h(x_k, \hat{x}_h^{(t)}, \sigma_{x,s,h}^{2(t)})}{\sum_{i=1}^H \hat{p}_i^{(t)} g_i(x_k, \hat{x}_i^{(t)}, \sigma_{x,s,i}^{2(t)})}, \quad \tau_{h,k}^{(0)} = \frac{\bar{p}_h g_h(x_k, \bar{x}_{s_h}, \sigma_{x,s_h}^2)}{\sum_{i=1}^H \bar{p}_i g_i(x_k, \bar{x}_{s_i}, \sigma_{x,s_i}^2)}. \end{cases} \tag{9}$$

where \hat{p}_h and $p_h^{(0)} = \bar{p}$ are explained by expressions (7) and (8).

$$\begin{aligned} \hat{x}_h^{(t+1)} &= w^{(t)} \bar{x}_{s_h} + (1 - w_h^{(t)}) \bar{x}_{\underline{s},h}^{(t)}, & (10) \\ \bar{x}_{\underline{s},h}^{(t)} &= \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} x_k \frac{\tau_{h,k}^{(t)}}{1 - \pi_k}, \quad w^{(t)} = \frac{\hat{N}_h}{\hat{N}_h + \tau_h^{(t)}}, \quad \bar{x}_h^{(0)} = \bar{x}_{s_h}, \\ \sigma_{x,s_h} &= \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{(x_k - \bar{x}_{s_h})^2}{\pi_k}, \quad \sigma_{y,s_h} = \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{(x_k - \bar{y}_{s_h})^2}{\pi_k}, \\ \sigma_{xy,s_h} &= \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{(x_k - \bar{x}_{s_h})(y_k - \bar{y}_{s_h})}{\pi_k}, \quad \bar{x}_{s_h} = \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{x_k}{\pi_k}, \quad \bar{y}_{s_h} = \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{y_k}{\pi_k}. \end{aligned} \tag{11}$$

The following regression-type estimators of $\mu_{y,h}$ are derived in the Appendix:

$$\begin{aligned} \hat{y}_h^{(t+1)} &= \bar{y}_{s_h} - \frac{\sigma_{xy,s_h}}{\hat{\sigma}_{x,h}^{2(t+1)}} (\bar{x}_{s_h} - \hat{x}_h^{(t+1)}) \quad \text{or} \\ \hat{y}_h^{(t+1)} &= \bar{y}_{s_h} - (1 - w_h^{(t)}) \frac{\sigma_{xy,s_h}}{\hat{\sigma}_{x,h}^{2(t+1)}} (\bar{x}_{s_h} - \bar{x}_{\underline{s},h}^{(t)}), \end{aligned} \tag{12}$$

$$\begin{aligned} \tilde{y}_h^{(t+1)} &= \bar{y}_{s_h} - \frac{\sigma_{xy,s_h}}{\sigma_{x,s_h}^2} (\bar{x}_{s_h} - \hat{x}_h^{(t+1)}) \quad \text{or} \\ \tilde{y}_h^{(t+1)} &= \bar{y}_{s_h} - (1 - w_h^{(t)}) \frac{\sigma_{xy,s_h}}{\sigma_{x,s_h}^2} (\bar{x}_{s_h} - \bar{x}_{\underline{s},h}^{(t)}), \end{aligned} \tag{13}$$

where $t = 0, 1, 2, \dots$,

$$\hat{\sigma}_{x,h}^{2(t+1)} = w_h^{(t)} \sigma_{x,s_h}^2 + (1 - w_h^{(t)}) \sigma_{x,\underline{s},h}^{2(t)}, \quad \hat{\sigma}_{x,h}^{2(0)} = \sigma_{x,s_h}^2, \tag{14}$$

$$\sigma_{x,\underline{s},h}^{2(t)} = \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} \frac{(x_k - \bar{x}_{\underline{s},h}^{(t)})^2}{1 - \pi_k} \tau_{h,k}^{(t)}, \tag{15}$$

When the constant of the linear regression y on x is approximately equal to zero, we can use the following ratio-type estimator:

$$\check{y}_h^{(t+1)} = \bar{y}_{s_h} \frac{\hat{x}_h^{(t+1)}}{\bar{x}_{s_h}} = w_h^{(t)} \bar{y}_{s_h} + (1 - w_h^{(t)}) \bar{y}_{s_h} \frac{\bar{x}_{s,h}^{(t)}}{\bar{x}_{s_h}}, \tag{16}$$

Particularly, in the case of a simple random sample drawn without replacement, when $\pi_k = \frac{n}{N}$ for all $k \in U$, we have:

$$\begin{aligned} \hat{N}_h &= N \bar{p}_h = N \frac{n_h}{N}, \quad \tau_h^{(t)} = N \bar{\tau}_h^{(t)}, \quad \bar{\tau}_h^{(t)} = \frac{1}{N - n} \sum_{k \in \underline{s}} \tau_{h,k}^{(t)}, \quad w^{(t)} = \frac{\bar{p}_h}{\bar{p}_h + \bar{\tau}_h^{(t)}}, \\ \bar{x}_{s_h} &= \frac{1}{n_h} \sum_{k \in s_h} x_k, \quad \bar{y}_{s_h} = \frac{1}{n_h} \sum_{k \in s_h} y_k, \quad \sigma_{x,s_h}^2 = \sigma_{xx,s_h}, \quad \sigma_{y,s_h}^2 = \sigma_{yy,s_h}, \\ \sigma_{xy,s_h} &= \frac{1}{n_h} \sum_{k \in s_h} (x_k - \bar{x}_{s_h})(y_k - \bar{y}_{s_h}), \\ \bar{x}_{s,h}^{(t)} &= \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} x_k \tau_{h,k}^{(t)}, \quad \sigma_{x,s,h}^{2(t)} = \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} (x_k - \bar{x}_{s,h}^{(t)})^2 \tau_{h,k}^{(t)}, \end{aligned} \tag{17}$$

Generalization of the proposed regression-type estimators into the case of a multi-dimensional auxiliary variable is as follows. Let

$$\begin{aligned} \hat{\mathbf{x}}^{(t+1)} &= w^{(t)} \bar{\mathbf{x}}_{s_h} + (1 - w^{(t)}) \bar{\mathbf{x}}_{s,h}^{(t)}, \\ \bar{\mathbf{x}}_{s_h} &= [\bar{x}_{1s_h} \dots \bar{x}_{is_h} \dots \bar{x}_{ms_h}], \quad \bar{\mathbf{x}}_{s,h}^{(t)} = [\bar{x}_{1s,h}^{(t)} \dots \bar{x}_{is,h}^{(t)} \dots \bar{x}_{ms,h}^{(t)}], \\ \bar{x}_{i,s_h} &= \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{x_{k,i}}{\pi_k}, \quad \bar{x}_{i,s,h}^{(t)} = \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} \frac{x_{k,i} \tau_{h,k}^{(t)}}{1 - \pi_k}, \quad \tau_{h,k}^{(t)} \\ &= \frac{\hat{p}_h^{(t)} g_h(\mathbf{x}_k, \hat{\mathbf{x}}^{(t+1)}, \hat{\Sigma}_{xxh}^{(t)})}{\sum_{e=1}^H \hat{p}_e^{(t)} g_e(\mathbf{x}_k, \hat{\mathbf{x}}^{(t+1)}, \hat{\Sigma}_{xxe}^{(t)})}. \end{aligned}$$

Let \mathbf{I}_a be the unit matrix of degree a and \mathbf{J}_a be the a -element column vector which all elements are equal to one. The rows of the matrix \mathbf{X} could be rewritten in such a way that

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_{\underline{s}} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \mathbf{U}_s \\ \mathbf{U}_{\underline{s}} \end{bmatrix}, \quad \mathbf{U}_{\underline{s}} = \mathbf{M} \mathbf{X}_{\underline{s}}, \quad \mathbf{M} = \mathbf{I}_{N-n} - \frac{1}{N-n} \mathbf{J}_{N-n} \mathbf{J}_{N-n}^T, \\ \boldsymbol{\pi} &= \begin{bmatrix} \boldsymbol{\pi}_s \\ \boldsymbol{\pi}_{\underline{s}} \end{bmatrix}, \quad \underline{\boldsymbol{\pi}} = \mathbf{J}_N - \boldsymbol{\pi}, \quad D(\boldsymbol{\pi}) = \text{diag}(\boldsymbol{\pi}), \\ \hat{\Sigma}_{xx,h}^{(t)} &= w^{(t)} \Sigma_{xx,s_h} + (1 - w^{(t)}) \hat{\Sigma}_{xx,s,h}^{(t)}, \end{aligned}$$

$$\begin{aligned} \Sigma_{xx,s_h} &= \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{(\mathbf{x}_k - \bar{\mathbf{x}}_{s_h})^T (\mathbf{x}_k - \bar{\mathbf{x}}_{s_h})}{\pi_k}, & \Sigma_{xy,s_h} &= \frac{1}{\hat{N}_h} \sum_{k \in s_h} \frac{(\mathbf{x}_k - \bar{\mathbf{x}}_{s_h})^T (y_k - \bar{y}_{s_h})}{\pi_k} \\ \Sigma_{xx,\underline{s},h}^{(t)} &= \frac{1}{\tau_h^{(t)}} \sum_{k \in \underline{s}} \frac{(\mathbf{x}_k - \bar{\mathbf{x}}_{\underline{s},h}^{(t)})^T (\mathbf{x}_k - \bar{\mathbf{x}}_{\underline{s},h}^{(t)})}{1 - \pi_k} \tau_{h,k}^{(t)} \\ &= \frac{1}{\hat{N}} \mathbf{U}_{\underline{s}}^T \mathbf{U}_{\underline{s}} D^{-1} (\boldsymbol{\pi}_{\underline{s}}) = \frac{1}{\hat{N}} \mathbf{X}_{\underline{s}}^T \mathbf{M} \mathbf{X}_{\underline{s}} D^{-1} (\boldsymbol{\pi}_{\underline{s}}) \end{aligned}$$

These let us generalize the estimators defined by expressions (12) and (13) as follows:

$$\begin{aligned} \hat{y}_h^{(t+1)} &= \bar{y}_{s_h} - \hat{\Sigma}_{xx,h}^{-1} \Sigma_{xy,s_h}^T (\bar{\mathbf{x}}_{s_h} - \hat{\mathbf{x}}_h^{(t+1)}) \text{ or} \\ \hat{y}_h^{(t+1)} &= \bar{y}_{s_h} - (1 - w_h^{(t)}) \hat{\Sigma}_{xx,h}^{-1} \Sigma_{xy,s_h}^T (\bar{\mathbf{x}}_{s_h} - \bar{\mathbf{x}}_{\underline{s},h}^{(t)}), \end{aligned} \tag{18}$$

$$\begin{aligned} \tilde{y}_h^{(t+1)} &= \bar{y}_{s_h} - \Sigma_{xx,s_h}^{-1} \Sigma_{xy,s_h}^T (\bar{\mathbf{x}}_{s_h} - \hat{\mathbf{x}}_h^{(t+1)}) \text{ or} \\ \tilde{y}_h^{(t+1)} &= \bar{y}_{s_h} - (1 - w_h^{(t)}) \Sigma_{xx,s_h}^{-1} \Sigma_{xy,s_h}^T (\bar{\mathbf{x}}_{s_h} - \bar{\mathbf{x}}_{\underline{s},h}^{(t)}), \end{aligned} \tag{19}$$

where $t = 0, 1, 2, \dots$ and $\bar{\mathbf{x}}_{\underline{s},h}^{(0)} = \bar{\mathbf{x}}_{s_h}, \hat{\Sigma}_{xx,h}^{(0)} = \Sigma_{xx,s_h}$.

Usually, the estimation process is stopped when the number of iterations t reaches the preassigned level T . Some other stopping rules are discussed, e.g., in [6, 7]. These works also considered several procedures which assess accuracy of estimators such as bootstrap methods.

Finally, let us show that the estimators evaluated in the previous paragraph, which are given by expressions (12)–(14), (16), (18), (19), (7) and (8), let us construct the following estimators of the population mean:

$$\begin{aligned} \hat{y}_h^{(t+1)} &= \sum_{h=1}^H \hat{p}_h^{(t+1)} \hat{y}_h^{(t+1)}, & \tilde{y}_h^{(t+1)} &= \sum_{h=1}^H \hat{p}_h^{(t+1)} \tilde{y}_h^{(t+1)}, \\ \check{y}_h^{(t+1)} &= \sum_{h=1}^H \hat{p}_h^{(t+1)} \check{y}_h^{(t+1)}. \end{aligned} \tag{20}$$

where $t = 0, 1, 2, \dots$

3.2 Simulation Study

Let simple random samples $\{s_j, j = 1, \dots, M\}$ be independently drawn without replacement from the whole population of size N . We assume that each of them is partitioned between H -domains in such a way that $s_j = s_{1,j} \cup \dots \cup s_{h,j} \cup \dots \cup s_{H,j}$ and $2 \leq n_h \leq n - 2(H - 1), h = 1, \dots, H$. Values of relative efficiency coefficient for estimator of the mean in the h -domain, $h = 1, \dots, H$, are defined as the following ratio:

$$e(t_{s_h}) = \frac{mse(t_{s_h})}{v(\bar{y}_{s_h})} 100\%, \quad h = 1, \dots, H$$

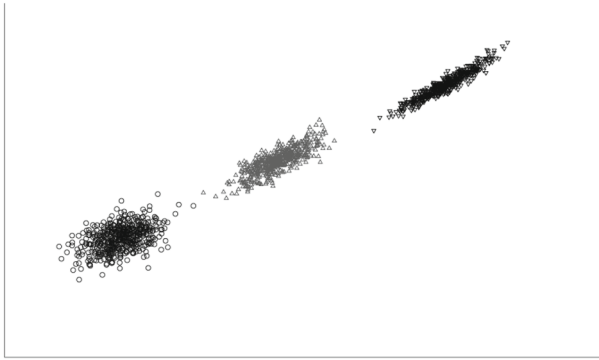


Fig. 1 Spread of the generated data

where $mse(t_{s_h}) = \frac{1}{M} \sum_{j=1}^M (t_{s_{h,j}} - \bar{y}_h)^2$, $v(\bar{y}_{s_h}) = \frac{1}{M} \sum_{j=1}^M (\bar{y}_{s_{h,j}} - \bar{y}_h)^2$, $\bar{y}_h = \frac{1}{N_h} \sum_{j=1}^M y_{k,i}$, $h = 1, \dots, H$. The relative bias of estimators is defined as follows:

$$b(t_{s_h}) = \frac{|\bar{t}_{s_h} - \bar{y}_h|}{mse(t_{s_h})} 100\%,$$

$$\bar{t}_{s_h} = \frac{1}{M} \sum_{j=1}^M t_{s_{h,j}}, \quad h = 1, \dots, H.$$

We assume that $M = 10000$.

Example 1 Let us consider the following simple set of data on a two-dimensional random variable generated according to two-dimensional normal distribution. The set consists of three domains of the same size equal to 500. Hence, a population of size 1500 is divided into three domains. The data in the h -domain are generated according to normal distribution $N(\mu_{x,h}, \mu_{y,h}, v_{x,h}, v_{y,h}, \rho_h)$. We will consider the following population partitioned into domains. The domain parameters of the population are: $N(8, 4, 1, 1, 0.5)$, $N(14, 11.2, 1, 1, 0.8)$ and $N(20, 19, 1, 1, 0.95)$. The spread of artificially generated data is shown in Fig. 1.

The simple random sample ($s = s_1 \cup s_2 \cup s_3$) is drawn without replacement from the whole population of size N . We assume that the size of each $s_{h,j}$ is $2 \leq n_h \leq n - 2(H - 1)$, $h = 1, 2, 3$, $j = 1, \dots, M$. In the second column of Table 1, the domains are identified by integers 1, 2 and 3. Columns 3–6 give the relative efficiency coefficient values $e(\cdot)$ for the domains.

Estimators \bar{y}_{s_h} , $\tilde{y}_h^{(t)}$ and $\check{y}_h^{(t)}$ are less accurate than $\hat{y}_h^{(t)}$. In the second domain estimators $\tilde{y}_h^{(t)}$ and $\check{y}_h^{(t)}$ are more accurate than \bar{y}_{s_h} and comparable with the accuracy of $\hat{y}_h^{(t)}$ for $n = 75, 150$ and in the third domain for $n = 150$. All considered estimators are practically unbiased because their relative biases (evaluated as the ratio of the bias module by the square root of the mean square error of the estimator) are not larger than 0.1%. Therefore, the biases are not shown in Table 1. Accuracy of the estimators increases when the sample size also increases. When the correlation coefficient

Table 1 Relative efficiency coefficients. *Source:* Own calculations

T_{sh}	n	15	45	75	150	
1	2	3	4	5	6	
	1	94.6	88.9	86.2	86.2	
$\hat{y}_h^{(t)}$	2	73.3	57.2	52.4	48.2	
	3	64.2	44.5	35.9	29.5	
$\tilde{y}_h^{(t)}$	1	286	202	131	98.9	
	2	115	76.7	58.0	43.3	
$\check{y}_h^{(t)}$	3	599	405	159	40.6	
	1	241	205	135	97.4	
$\hat{p}_h^{(t)}$	2	107	76.0	57.4	47.8	
	3	569	406	158	40.7	
$\hat{y}_h^{(t)}$	1	101	33.3	31.9	21.3	
	2	127	59.0	38.3	21.5	
$\hat{y}_h^{(t)}$	3	104	45.5	30.5	20.7	
			83.3	37.3	27.8	20.9
$\tilde{y}_h^{(t)}$			45.3	28.4	24.8	20.7
$\check{y}_h^{(t)}$			41.7	26.7	24.6	20.6

The artificial population

between the auxiliary variable and the variable under study in a particular domain increases, then the accuracy of the estimator also increases. Regression estimator $\hat{y}_h^{(t)}$ is significantly more accurate than the ordinary sub-sample mean \bar{y}_{sh} . Statistic $\hat{y}_h^{(t)}$ seems to be the most universal of the considered estimators and therefore should be preferred.

The relative biases of $\hat{p}_h^{(t)}$, $h = 1, \dots, H$, are not larger than 0.5%. Accuracies of these also increase when the sample size increases. They are better than ordinary sample frequencies \bar{p}_h for $n \geq 45$. Hence, the considered procedure could also be used to estimate of the probabilities p_h , $h = 1, \dots, H$ of distribution mixtures. The several last rows of Table 1 let us say that all three estimators of the population average are significantly better than the simple sample mean. Moreover, the ratio-type estimator is the most accurate.

Example 2 The second population consists of data published in [11] about Swedish municipalities. We consider data about three variables REV84 (real estate values from 1984), RMT85 (revenues from municipal taxation in 1985) and ME84 (municipal employees in 1984). We take into account these data without the largest outliers. The size of the considered population is 281. The population was partitioned into three domains according to quantiles 30% and 70% of variable REV84. This provided the following sizes of domains: $N_1 = 86$, $N_2 = 109$ and $N_3 = 86$. Real estate valuation depends on market fluctuations. Therefore, the same property today may be in the first domain, but tomorrow it may be in a different domain. Therefore, belonging to a domain can be treated as random.

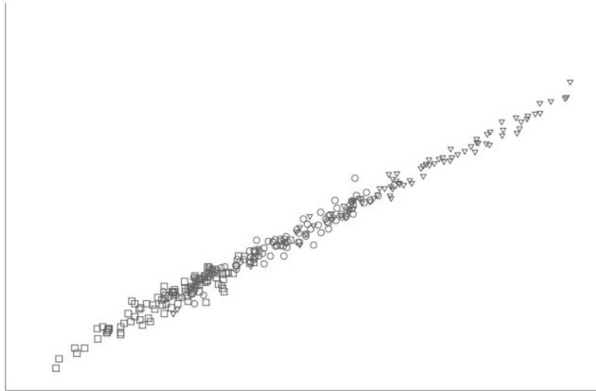


Fig. 2 Spread of logRMT85 and logME84

Table 2 Estimation accuracy of population mean. *Source:* Own calculations

n	8	14	28
1	2	3	4
e(.)			
$\hat{y}^{(t)}$	85.1	58.4	46.2
$\tilde{y}^{(t)}$	30.0	24.9	21.8
$\check{y}^{(t)}$	49.0	43.0	43.9
b(.)			
$\hat{y}^{(t)}$	8.0	7.9	5.2
$\tilde{y}^{(t)}$	1.7	1.0	3.3
$\check{y}^{(t)}$	31.6	18.0	9.4

The population of 281 Swedish municipal units

The distributions of variables RMT85 and ME84 have too much asymmetry on the right side, and they differ too significantly from normal distribution. Therefore, we considered their logarithmic transformation, and the spread of this is shown Fig. 2. The domain mean values of logRMT85 were $\mu_1 = 6.704$, $\mu_2 = 7.5.20$ and $\mu_3 = 8.528$. The simulation of estimation accuracy was based on the simple random samples drawn without replacement. The sizes of the samples were: 8 (2.85% the population size), 14 (4.98%) and 28(2.96%). Table 2 shows only the accuracies of the estimation of population mean because the estimators of the domain means were less accurate than the simple random sample mean. Analysis of Table 2 lets us say that all of the estimators of the population mean that are taken into account, are more accurate than the simple random sample mean. The accuracy of the second regression estimator is the best among the considered ones, and similarly its relative bias is the smallest.

Example 3 Let us consider data about current and starting salaries of employees that are available in the SPSS statistical packages as the example dataset. The set consists of 474 observations. The two data domains are identified. The first domain of 390 observations is the set of clerks and the second one consists of 84 managers. In general,

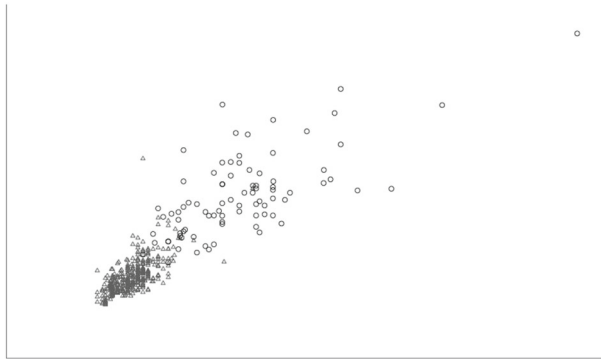


Fig. 3 Spread of the data on starting and current salaries

Table 3 Estimation accuracy of the population mean. *Source:* Own calculations

n	15	24	48
1	2	3	4
e(.)			
$\hat{y}^{(t)}$	85.7	86.6	114
$\tilde{y}^{(t)}$	122	38.2	60.9
$\check{y}^{(t)}$	41.3	36.0	32.3
b(.)			
$\hat{y}^{(t)}$	47.0	51.7	78.9
$\tilde{y}^{(t)}$	50.9	15.1	4.0
$\check{y}^{(t)}$	20.0	10.3	10.5

Population of employees

an employee randomly belongs to one of these domains, because one day he could be a manager and the next day he could be a clerk, and vice versa. The starting and current salaries in the first domain (clerks) are \$14164 and \$28054, respectively. The starting and current salaries in the second domain (managers) are \$28091 and \$63978, respectively. The spread of the data partitioned into domains is shown in Fig. 3. The following sizes of samples were taken into account: 15 (3.2%), 24 (5.1%) and 48 (10.1%). The results of the simulation are shown in Table 3. Similar to example 2, this table shows only the accuracy of the estimation of the population mean because the estimators of domain means were less accurate than the simple random sample mean.

Analysis of Table 3 leads to the conclusions that the all estimators are more accurate than the simple sample mean for sample size $n > 24$ except the ratio estimator because its deff coefficients are less than 100% for all sample sizes. The accuracy of the ratio estimator decreases when the sample size increases. The relative biases of the estimators are quite large.

Analysis of all the tables and figures lets us say that estimation of domain means is possible only when data observed in domains are well separated. The more optimistic

conclusion is that the proposed estimators of population means are always more accurate than the simple random sample in all considered cases, when the sample size is at least 5% of the population size. Their biases are also rather acceptable. Of those estimators of the population mean, the second regression estimator and the ratio estimator are the best and could be used in practical research.

The presented simulation analyzes will be continued in a wider scope in the next article. In particular, different mixtures of at least three-dimensional probability distributions will be considered. In addition, various modifications to the estimators used herein will be proposed, leading to a more accurate estimation of the domain averages.

4 Conclusions

Three estimators of domain means use additional auxiliary data in order to improve estimation accuracy. Properties of the maximum likelihood method let us to derive new estimators of domain and population means. The simulation analysis shows which is the best estimator for occasions when the domains are sufficiently well separated. This separation may not to be very obvious when estimating the population mean. In this case, all of the estimators of population means were better than the simple random sample mean. The considered estimation method lets us also estimate the probabilities of the distribution mixtures. Generalization of the regression estimators for a multidimensional auxiliary variable was also shown.

Some other generalization or modifications of the estimation procedure are possible. Auxiliary variables observed in censuses or in official registers can be used to improve the efficiency of estimating means. We can consider distributions other than normal as elements of the mixture. For instance, expenditures or incomes in domains could be modeled by means of asymmetric distributions like lognormal or gamma distributions.

Acknowledgements This paper is the result of a grant supported by the National Science Centre, Poland, No. 2016/21/B/HS4/00666.

Declarations

Conflict of interest The author declare that he has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Derivation of parameters $\hat{\Theta}_x^{(t+1)}$

$\hat{\Theta}_x^{(t+1)}$ maximize the log-likelihood function $l^{(t)}(\mathbf{d}_s, \mathbf{x}_s)$ [given by expression (4)] provided $\sum_{h=1}^H p_h = 1$. Equivalently $\hat{\Theta}_x^{(t+1)}$ maximize the Lagrangian function: $\psi_x^{(t)}(\Theta) = l^{(t)}(\mathbf{d}_s, \mathbf{x}_s) + \lambda(1 - \sum_{h=1}^H p_h)$. According to the pseudo-likelihood approach (see, e.g., [8, 12]), $\hat{\Theta}_x^{(t+1)}$ is the solution of the following pseudo-log-likelihood estimation equation system:

$$\begin{cases} \frac{\partial \psi_x^{(t)}(\Theta)}{\partial \Theta} = \frac{\partial l_1(\mathbf{d}_s)}{\partial \Theta} + \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \Theta_x} = \mathbf{0}, \\ \frac{\partial l^{(t)}(\mathbf{d}_s, \mathbf{x}_s)}{\partial p_h} = \lambda, \quad h = 1, \dots, H \end{cases} \tag{21}$$

where

$$\begin{cases} \frac{\partial l_1(\mathbf{d}_s)}{\partial \Theta} = \sum_{h=1}^H \sum_{k \in s_h} \frac{\partial \ln(f_h(y_k, \mathbf{x}_k, \Theta))}{\partial \Theta}, \\ \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \Theta_x} = \sum_{h=1}^H \sum_{k \in s_h} \frac{\partial \ln(g_h(\mathbf{x}_k, \Theta_x))}{\partial \Theta_x} \tau_{h,k}^{(t)}, \end{cases} \tag{22}$$

$$\frac{\partial l^{(t)}(\mathbf{d}_s, \mathbf{x}_s)}{\partial p_h} = \frac{\partial l_1(\mathbf{d}_s)}{\partial p_h} + \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial p_h}, \quad h = 1, \dots, H, \tag{23}$$

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial p_h} = \frac{\hat{N}_h}{p_h}, \quad \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial p_h} = \frac{\hat{\tau}_h^{(t)}}{p_h}. \tag{24}$$

These equations let us write: $\hat{N}_h + \hat{\tau}_h^{(t)} = \lambda p_h$ for $h = 1, \dots, H$. This leads to $\lambda = \sum_{h=1}^H (\hat{N}_h + \hat{\tau}_h^{(t)})$ and finally to the expression (7).

Derivation of estimators under an assumed normal distribution mixture

Under an assumed mixture of normal distribution, the components of the model expressed by (1) become as follows:

$$f_h(y_k, x_k, \theta_h) = \frac{1}{2\pi \sqrt{\sigma_{x,h}^2 \sigma_{y,h}^2 (1 - \rho_h^2)}} \exp\{Q(y_k, x_k, \theta_h)\} \tag{25}$$

where $\theta_h = [\mu_{y,h} \ \mu_{x,h} \ \sigma_{y,h} \ \sigma_{x,h} \ \sigma_{x,y,h}]$, $\rho_h = \frac{\sigma_{xy,h}}{\sigma_{x,h} \sigma_{y,h}}$ and

$$Q(y_k, x_k, \theta_h) = -\frac{1}{2(1 - \rho_h^2)} \left(\frac{(x_k - \mu_{x,h})^2}{\sigma_{x,h}^2} - 2\rho_h \frac{(x_k - \mu_{x,h})(y_k - \mu_{y,h})}{\sqrt{\sigma_{x,h}^2 \sigma_{y,h}^2}} + \frac{(y_k - \mu_{y,h})^2}{\sigma_{y,h}^2} \right),$$

$$g(\mathbf{x}_k, \Theta_x) = \sum_{h=1}^H p_h g_h(\mathbf{x}_k, \theta_{x,h}), \tag{26}$$

$$g_h(x_k, \theta_{x,h}) = g_h(x_k, \mu_{x,h}, \sigma_{x,h}^2) = \frac{1}{\sqrt{2\pi}} (\sigma_{x,h}^2)^{-\frac{1}{2}} \exp \left\{ -\frac{(x_k - \mu_{x,h})^2}{2\sigma_{x,h}^2} \right\}.$$

The log-likelihood functions defined by expressions (3)–(6) take the following forms:

$$l_1(\mathbf{d}_s) = \sum_{h=1}^H \hat{N}_h \ln(p_h) - \hat{N} \ln(2\pi) + \tag{27}$$

$$-\frac{1}{2} \sum_{h=1}^H \hat{N}_h \left(\ln(\sigma_{x,h}^2) + \ln(\sigma_{y,h}^2) + \ln(1 - \rho_h^2) \right) + \sum_{h=1}^H Q_{s_h}(y_k, x_k, \theta_h),$$

where

$$Q_{s_h}(y_k, x_k, \theta_h) = -\frac{\hat{N}_h}{2(1 - \rho_h^2)} \left(\frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} - 2\rho_h \frac{\sigma_{*xy,s_h}}{\sqrt{\sigma_{x,h}^2 \sigma_{y,h}^2}} + \frac{\sigma_{*y,s_h}^2}{\sigma_{y,h}^2} \right), \tag{28}$$

$$\sigma_{*xy,s_h} = \frac{1}{\hat{N}_h} \sum_{k \in s_h} (x_k - \mu_{x,h})(y_k - \mu_{y,h}), \quad \sigma_{*x,s_h}^2 = \sigma_{*xx,s_h}, \quad \sigma_{*y,s_h}^2 = \sigma_{*yy,s_h}.$$

$$l_2^{(t)}(\mathbf{x}_s) = \sum_{h=1}^H \ln(p_h) \tau_h^{(t)} - \frac{1}{2} \sum_{h=1}^H \sum_{k \in s} \frac{\tau_{h,k}^{(t)}}{1 - \pi_k} \left(\ln(2\pi) + \ln(\sigma_{x,h}^2) + \frac{(x_k - \mu_{x,h})^2}{\sigma_{x,h}^2} \right) \tag{29}$$

$$= \sum_{h=1}^H \ln(p_h) \tau_h^{(t)} - \frac{\tau^{(t)}}{2} \ln(2\pi) - \frac{1}{2} \sum_{h=1}^H \ln(\sigma_{x,h}^2) \tau_h^{(t)} - \frac{1}{2} \sum_{h=1}^H \frac{\sigma_{*x,s,h}^{2(t)} \tau_h^{(t)}}{\sigma_{x,h}^2},$$

$$\sigma_{*x,s,h}^{2(t)} = \frac{1}{\tau_h^{(t)}} \sum_{k \in s} \frac{(x_k - \mu_{x,h})^2}{1 - \pi_k} \tau_{h,k}^{(t)}. \tag{30}$$

The first derivatives of the likelihood function expressed by (27) in the case of normal distribution, as given by (25), are evaluated as follows (see Kendall and Stuart [5], page 57):

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial \mu_{x,h}} = \frac{\hat{N}_h}{\sigma_{x,h}(1 - \rho_h^2)} \left(\frac{\bar{x}_{s_h} - \mu_{x,h}}{\sigma_{x,h}} - \rho_h \frac{\bar{y}_{s_h} - \mu_{y,h}}{\sigma_{y,h}} \right), \tag{31}$$

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial \mu_{y,h}} = \frac{\hat{N}_h}{\sigma_{y,h}(1 - \rho_h^2)} \left(\frac{\bar{y}_{s_h} - \mu_{y,h}}{\sigma_{y,h}} - \rho_h \frac{\bar{x}_{s_h} - \mu_{x,h}}{\sigma_{x,h}} \right), \tag{32}$$

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial \sigma_{x,h}^2} = -\frac{\hat{N}_h}{2\sigma_{x,h}^2(1 - \rho_h^2)} \left((1 - \rho_h^2) - \frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} + \rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} \right), \tag{33}$$

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial \sigma_{y,h}^2} = -\frac{\hat{N}_h}{2\sigma_{y,h}^2(1-\rho_h^2)} \left((1-\rho_h^2) - \frac{\sigma_{*y,s_h}^2}{\sigma_{y,h}^2} + \rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} \right), \tag{34}$$

$$\frac{\partial l_1(\mathbf{d}_s)}{\partial \rho_h} = \frac{\hat{N}_h}{1-\rho_h^2} \left\{ \rho_h - \frac{1}{1-\rho_h^2} \left[\rho_h \left(\frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} + \frac{\sigma_{*y,s_h}^2}{\sigma_{y,h}^2} \right) - (1+\rho_h^2) \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} \right] \right\}. \tag{35}$$

The first derivatives of the likelihood function expressed by (29) in the case of normal distribution are as follows:

$$\frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \mu_{x,h}} = \sum_{k \in \underline{s}} \frac{(x_h - \mu_{x,h})\tau_{h,k}^{(t)}}{\sigma_{x,h}^2} = \frac{(\bar{x}_{s,h}^{(t)} - \mu_{x,h})\tau_h^{(t)}}{\sigma_{x,h}^2}, \tag{36}$$

$$\frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \sigma_{x,h}^2} = -\frac{\tau_h^{(t)}}{2\sigma_{x,h}^2} \left(1 - \frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} \right) \tag{37}$$

where $\tau_{h,k}^{(t)}$, $\tau_h^{(t)}$, $\bar{x}_{s,h}^{(t)}$ and σ_{*x,s_h} are given by expressions (9), (11) and (30).

Equations $\frac{\partial l^{(t)}(\mathbf{x}_U)}{\partial \mu_{x,h}} = \frac{\partial l_1(\mathbf{d}_s)}{\partial \mu_{x,h}} + \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \mu_{x,h}} = 0$ and $\frac{\partial l^{(t)}(\mathbf{x}_U)}{\partial \mu_{y,h}} = \frac{\partial l_1(\mathbf{d}_s)}{\partial \mu_{y,h}} = 0$ are equivalent to the following, respectively:

$$\frac{\hat{N}_h}{\sigma_{x,h}(1-\rho_h^2)} \left(\frac{\bar{x}_{s_h} - \mu_{x,h}}{\sigma_{x,h}} - \rho_h \frac{\bar{y}_{s_h} - \mu_{y,h}}{\sigma_{y,h}} \right) + \frac{(\bar{x}_{s,h} - \mu_{x,h})\tau_h^{(t)}}{\sigma_{x,h}^2} = 0 \tag{38}$$

and

$$\frac{\bar{y}_{s_h} - \mu_{y,h}}{\sigma_{y,h}} = \rho_h \frac{\bar{x}_{s_h} - \mu_{x,h}}{\sigma_{x,h}}. \tag{39}$$

When in (38) we replace $\frac{\bar{y}_{s_h} - \mu_{y,h}}{\sigma_{y,h}}$ with the right side of equation (39), then, after simplification, we have:

$$\mu_{x,h}(\hat{N}_h + \tau_h^{(t)}) = \hat{N}_h \bar{x}_{s_h} + \tau_h^{(t)} \bar{x}_{s,h}^{(t)}.$$

This leads to the iterative estimator $\hat{x}_{s,h}^{(t+1)}$, given by (10) and (11).

Equation $\frac{\partial l_1(\mathbf{d}_s)}{\partial \mu_{y,h}} = 0$ is equivalent to the following:

$$\mu_{y,h} = \bar{y}_{s_h} - \frac{\sigma_{xy,h}}{\sigma_{x,h}^2} (\bar{x}_{s_h} - \mu_{x,h}).$$

After replacing $\mu_{x,h}$ with $\hat{x}_{x,h}^{(t+1)}$, we have:

$$\hat{\mu}_{y,h}^{(t+1)} = \bar{y}_{s_h} - \frac{\sigma_{xy,h}}{\sigma_{x,h}^2} (\bar{x}_{s_h} - \hat{x}_{x,h}^{(t+1)}) = \bar{y}_{s_h} - (1-w_h^{(t)}) \frac{\sigma_{xy,h}}{\sigma_{x,h}^2} (\bar{x}_{s_h} - \bar{x}_{s,h}^{(t)}), \tag{40}$$

Now, when we replace $\sigma_{xy,h}$ and $\sigma_{x,h}^2$ with σ_{xy,s_h} and σ_{x,s_h}^2 , respectively, we obtain the estimator $\tilde{y}_h^{(t+1)}$ expressed by (13).

Equations $\frac{\partial l^{(t)}(\mathbf{x}_U)}{\partial \sigma_{y,h}^2} = \frac{\partial l_1(\mathbf{d}_s)}{\partial \sigma_{y,h}^2} = 0$, $\frac{\partial l^{(t)}(\mathbf{x}_U)}{\partial \sigma_{x,h}^2} = \frac{\partial l_1(\mathbf{d}_s)}{\partial \sigma_{x,h}^2} + \frac{\partial l_2^{(t)}(\mathbf{x}_s)}{\partial \sigma_{x,h}^2} = 0$ are equivalent to the following:

$$1 - \rho_h^2 = \frac{\sigma_{*y,s_h}^2}{\sigma_{y,h}^2} - \rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}}, \tag{41}$$

$$1 - \rho_h^2 = \frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} - \rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} + \frac{(1 - \rho_h^2)\tau_h^{(t)}}{n_h} \left(\frac{\sigma_{*x,s,h}^2}{\sigma_{x,h}^2} - 1 \right) \tag{42}$$

and after multiplying $\frac{\partial l_1(\mathbf{d}_s)}{\partial \rho_h} = 0$ [see equation (35)] by $\frac{1 - \rho_h^2}{n_h \rho_h}$ we have:

$$1 - \rho_h^2 = \frac{\sigma_{*y,s_h}^2}{\sigma_{y,h}^2} + \frac{\sigma_{*x,s_h}^2}{\sigma_{x,h}^2} - \frac{1 + \rho_h^2}{\rho_h} \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}}. \tag{43}$$

Similar to Kendall and Stuart [5] pp. 57–58, let us add Eqs. (41) and (42). Next, after subtracting Eq. (43) from this sum, we have:

$$1 - \rho_h^2 = \frac{1 - \rho_h^2}{\rho_h} \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} + \frac{(1 - \rho_h^2)\tau_h^{(t)}}{n_h} \left(\frac{\sigma_{*x,s,h}^2}{\sigma_{x,h}^2} - 1 \right).$$

The above equation is multiplied by $\frac{\rho_h^2}{1 - \rho_h^2}$ and simplified to the form:

$$\rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}} = \rho_h^2 \left(1 - \frac{\tau_h^{(t)}}{n_h} \left(\frac{\sigma_{*x,s,h}^2}{\sigma_{x,h}^2} - 1 \right) \right).$$

In Eq. (42) we replace $\rho_h \frac{\sigma_{*xy,s_h}}{\sigma_{x,h}\sigma_{y,h}}$ with the right side of the above equation. After some simplification, this lets us write:

$$\sigma_{x,h}^2 = w_h^{(t)} \sigma_{*x,s_h}^2 + (1 - w_h^{(t)}) \sigma_{*x,s,h}^2.$$

In expressions (28) and (30), mean μ_h is replaced by \bar{x}_{s_h} and $\bar{x}_{s_h}^{(t)}$, respectively. This leads to expression (14) and (12).

References

1. Cassel CM, Särndal CE, Wretman JH (1977) Foundation of inference in survey sampling. Wiley, New York
2. Chambers RL, Skinner CJ (2003) Analysis of survey data. Wiley, Chichester

3. Chen J, Sitter RR (1999) A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Stat Sin* 9:385–406
4. Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J R Stat Soc Ser B* 39:1–38
5. Kendall MG, Stuart A (1961) *The advanced theory of statistics. Vol. 2: inference and relationship.* Charles Griffin & Company Limited, London
6. McLachlan G, Krishnan T (1977) *The EM algorithm and extensions.* Wiley, New York
7. McLachlan G, Peel D (2000) *Finite mixture models.* Wiley, New York
8. Pfeffermann D (1993) The role of sampling weights when modelling survey data. *Int Stat Rev* 61(2):317–337
9. Skinner CJ, Holt D, Smith TMF (1989) *Analysis of complex surveys.* Wiley, New York
10. Rao JNK, Molina I (2015) *Small area estimation.* Wiley, Hoboken
11. Särndal CE, Swensson B, Wretman J (1992) *Model assisted survey sampling.* Springer, New York
12. Thompson ME (1997) *Theory of sample surveys.* Chapman & Hall, London
13. Wywiał J L (2018) Application of two gamma distribution mixture to financial auditing. *Sankhya B* 80(1):1–18
14. Żądło T (2006) On prediction of total value in incompletely specified domains. *Aust N Z J Stat* 48(3):269–283

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.