Visualization of generalized mean estimators using auxiliary information in survey sampling

Rodolphe Priam

Visualization of generalized mean estimators using auxiliary information in survey sampling*

R. Priam (rpriam@gmail.com)

May 6, 2019

Abstract

The mean estimators depend on multiple auxiliary variables and unknown parameters in a finite population setting. We propose a new generic approach for modeling multivariate mean estimators. Our approach brings naturally a graphical analysis for comparing the mean estimators.

Keywords: ratio estimator, auxiliary variable, mean squared error, bias

1 Introduction

In survey theory, the general purpose is to find a relevant value for an aggregated stastistics non available at the population level for diverse reasons such as cost, time or feasability. For instance, it is cumbersome to query a whole population of a large country for a social survey each week. Similarly, the variable in stake may not be even observable directly thus the real population not reachable. For these reasons a sample is drawn from the finite population according to some method.

From the sample, the unknown aggregated statistics of the population -such as a mean the most often- needs to be infered, say estimated. In order to assist the mean estimation, auxiliary variables (or variates) bring additional information when available as some of their statistics are known at the level of the population. This helps for reduction of the mean squared error (mse) for the related mean estimators by eventually increasing the bias. The idea is that the correction applied to the sample mean of an auxiliary variable in order to retrieve its population mean can be (exactly) applied to our sample mean of interest for the characteristic or variable unknown at the population level. Several methods have been invented and developped in this domain of research, with in particular the ratio method with a multiplicative correction (Cochran, 1940) and the regression method with an additive correction (Rao, 1991).

For multiple auxiliary variables some proposed estimators (Allen, Singh, and Smarandache, 2003; Diana and Perri, 2007; Vishwakarma and Kumar, 2015) in the literature have their expression related to an additional (Olkin, 1958; Rao and Mudholkar, 1967; Singh and Espejo, 2003; Abu-Dayyeh, Ahmed, Ahmed, and Muttlak, 2003), multiplicative (Singh, 1967; Srivastava, 1971; Abu-Dayyeh et al., 2003), quotient (Shukla, 1966; John, 1969). These estimators are able to improve the one variable estimator by reducing the variance when several auxiliary variables are available.

A new generalizing class for the ratio estimators is introduced and the expressions of the corresponding mean squared errors are proposed for visualization purpose. The plan is as follows. In section 2 a new parametric class of univariate estimator is proposed by introducing a polynomial expansion. In section 3

*Only the published paper at [*Communications in Statistics - Theory and Methods*, 2019] is complete. A complementary contents is available via a separated document.

existing multivariate ratio estimators are generalized and listed, their mean squared errors are approximated via two differents ways in an unifying analytical expression. In section 4, the approximated mse of several estimators defined from the proposed parametric class in section 2 are written in closed form and minimized. In section 5 the experiments demonstrate the interest of the approach for the comparison of several estimators with real populations. In section 6 a discussion and the perspectives conclude the paper.

2 Parametric univariate ratio estimator

Let denote Y the variable of interest and X an auxiliary variable which is correlated with Y. The population mean $\bar{X}$ of X is known while the population mean $\bar{Y}$ of Y is unknown. The observation y_i for Y and x_i for X are available for each sampled unit: the sample $(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$ is a random variable of size n on pairs of variable (X,Y) drawn by simple random sampling for instance from a population of size N. Let define $\bar{x}=\sum_i^n x_i/n$ and $\bar{y}=\sum_i^n y_i/n$.

2.1 Definition

When $f(.;.)$ is an one variable function with an eventual vector of parameters θ, we defined the *parametric ratio estimator* as written as follows:

$$\bar{y}_{\mathrm{R}_f} \quad = \quad f_\theta(\bar{x};\bar{X})\bar{y}. \tag{1}$$

Let denote $\delta_x=\frac{\bar{x}-\bar{X}}{\bar{X}}$ and $\delta_y=\frac{\bar{y}-\bar{Y}}{\bar{Y}}$ where $E_s[\delta_x]=0$ and $E_s[\delta_y]=0$. Let define $a=a(\theta)$ and $b=b(\theta)$ eventually constant and taking their values according to the chosen function $f_\theta(.;.)$. An explicit 2^{nd} order serie approximation leads to:

$$f_\theta(\bar{x};\bar{X}) \quad = \quad 1+a(\theta)\delta_x+b(\theta)\delta_x^2+\cdots. \tag{2}$$

This is mostly related to (Diana, Giordan, and Perri, 2011) with explicit variables a and b too, but for the derivative also to (Srivastava, 1971; Srivastava and Jhajj, 1981, 1983) for instance. These reseaches consider also second order approximations but without a and b being fully variables as proposed herein. Note that a serie approximation is also proposed in other publications but similarly w.r.t. differences instead of relative differences. In general the value 1 is met as the lowest coefficient in the serie of $f_\theta(.;.)$ and our approach can be extended to other cases as a perspective. For $f_\theta(\bar{x};\bar{X})=\bar{X}/\bar{x}$ when the usual ratio estimator is considered, the value of a is just -1 and as a constant it does not depend on a parameter θ but in for some cases of functions it does. With $R=\bar{Y}/\bar{X}$, this functional approximation brings the following bias and mse for the parametric ratio estimator as,

$$\begin{array}{lcl} \mathrm{bias}_{(a,b)}\left[\bar{y}_{\mathrm{R}_f}\right] & \doteq & \bar{Y}\left(\frac{a}{\bar{X}\bar{Y}}Cov_s(\bar{x},\bar{y})+\frac{b}{\bar{X}^2}Var_s(\bar{x})\right) \\ \mathrm{mse}_{(a,b)}\left[\bar{y}_{\mathrm{R}_f}\right] & \doteq & a^2R^2Var_s(\bar{x})+2aRCov_s(\bar{x},\bar{y})+Var_s(\bar{y}). \end{array} \tag{3}$$

Note that the expectations, variances and covariances are according to the sample which is the random variable. Their expressions depend on the chosen sampling, typically a simple random sampling with or without replacement which leads to their usual expressions. At the first order approximation, this mean squared error may be smaller than the variance of the usual sample mean estimator $\bar{y}$ when $a<-2Cov_s(\bar{x},\bar{y})/RVar_s(\bar{x})$. Before presenting the multivariate case, it must be noticed the linearity of the bias when a is a free parameter and the second-degree polynomial form of the mse as a function of a, such as they may be both minimized w.r.t. a and b. This results into an estimator related to a regression estimator when the optimal expressions are replaced in the second-order approximation of $f_\theta(.;.)$ above. In the following the quantities a and b are supposed constant or depending on θ, an eventual scalar or vectorial parameter.

2.2 Exemples of values for $a(\theta)$ and $b(\theta)$

A list of the corresponding values for a and b for several functions (from the literature) is presented in the table 1. The functions are defined in (Singh, 1965) for f_1, in (Singh, 1965) for f_2, in (Bahl and Tuteja, 1991) for f_3 and f_4, in (Muneer, Shabbir, and Khalil, 2017) for f_5, in (Khoshnevisan, Singh, Chauhan, Sawan, and Smarandache, 2007; Yadav and Kadilar, 2013) for f_6, in (Haq and Shabbir, 2013) for f_7, in (Diana and Perri, 2010) for f_8, f_9 is inspired from f_7 and f_8 while f_{10} is adapted from (Bhushan and Kumari, 2018) for positive means.

Table 1: Examples of values for a and b.

	$f_\theta(\bar{x};\bar{X})$	θ	$a(\theta)$	$b(\theta)$
f_1	$\frac{\bar{x}}{\bar{X}}$		1	0
f_2	$\frac{\bar{X}}{\bar{x}}$		-1	1
f_3	$e^{\left[\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right]}$		$\frac{1}{2}$	$-\frac{1}{8}$
f_4	$e^{\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right]}$		$-\frac{1}{2}$	$\frac{3}{8}$
f_5	$\alpha\left[2-e^{\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}}\right]+(1-\alpha)e^{\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}}$	α	$-\frac{1}{2}$	$\frac{3}{8}-\frac{\alpha}{4}$
f_6	$\left[\frac{c\bar{X}+d}{\tau(c\bar{x}+d)+(1-\tau)(c\bar{X}+d)}\right]^g$	(c,d)	$-\frac{g\tau c\bar{X}}{d+c\bar{X}}$	$\frac{g(g+1)}{2}\left(\frac{\tau c\bar{X}}{d+c\bar{X}}\right)^2$
f_7	$e^{\left[\frac{c\bar{X}+d}{\tau(c\bar{x}+d)+(1-\tau)(c\bar{X}+d)}-1\right]}$	(c,d)	$-\frac{\tau c\bar{X}}{d+c\bar{X}}$	$\frac{3}{2}\left(\frac{\tau c\bar{X}}{d+c\bar{X}}\right)^2$
f_8	$\frac{\bar{X}+\gamma}{\bar{x}+\gamma}$	γ	$-\frac{\bar{X}}{\gamma+\bar{X}}$	$\left(\frac{\bar{X}}{\gamma+\bar{X}}\right)^2$
f_9	$e^{\left[\frac{\bar{X}+\gamma}{\bar{x}+\gamma}-1\right]}$	γ	$-\frac{\bar{X}}{\gamma+\bar{X}}$	$\frac{3}{2}\left(\frac{\bar{X}}{\gamma+\bar{X}}\right)^2$
f_{10}	$1+\gamma\ln\left(\frac{\bar{x}}{\bar{X}}\right)$	γ	γ	$-\frac{\gamma}{2}$

Next section, these functions f_θ are introduced further in the ratio estimators for the case when two auxiliary variables are available instead of just one.

3 Generalized bivariate estimators and mse

In this section, we review several bivariate estimators from the statistical literature in order to propose an expression of their mse when the function $f_\theta(.;)$ from the previous section 2 is considered. Some estimators like the quotient in (Shukla, 1966; John, 1969) or the classes of estimators in (Solanki and Singh, 2015) are not presented herein and are an appealing perspective. Considering the highly extensive research in the literature on modeling of ratio estimators, the purpose is not to review the whole literature in the domain but to show on several selected models the interest of a generic modeling associated to a visualisation of the estimators. Several auxiliary variables seem more appealing than just one because adding more information in the estimators is often able to reduce the bias or the variability.

In a bidimensional setting, the notation are as follows. Two auxiliary variables X_j are available with population mean $\bar{X}_j$ and sample mean $\bar{x}_j$ for $j=1$ and $j=2$. Hence $x_i=(x_{i1},x_{i2})$ is bidimensional. Let define $\bar{x}_j=\sum_i^n x_{ij}/n$ and $\bar{y}=\sum_i^n y_i/n$ and similarly $\bar{X}_j$ and $\bar{Y}$ for the population means. Let denote the $p=2$ free parameters α_j aggregated in the vector $\alpha=(\alpha_1,\alpha_2)^T$, such that the contraint $\sum_{j=1}^2\alpha_j=1$ is usually introduced. Let denote $\delta_{x_j}=(\bar{x}_j-\bar{X}_j)/\bar{X}_j$ such as $f_{\theta_j}(.;.)$ is obtained by replacing in $f_\theta(.;.)$ θ by a new vector θ_j eventually different for each j if not equal to θ and also replacing δ_x by δ_{x_j}. Let also denote $C_{0j}=S_{yx_j}/\bar{X}_j\bar{Y}$, $C_{jk}=S_{x_jx_k}/\bar{X}_j\bar{X}_k$, $C_0^2=S_y^2/\bar{Y}^2$, and $C_j^2=S_{x_j}^2/\bar{X}_j^2$, $Cov_s(\bar{y},\bar{x}_j)=\lambda_n S_{yx_j}$, $Cov_s(\bar{x}_j,\bar{x}_k)=\lambda_n S_{x_jx_k}$,

$Var_s(\bar{y}) = \lambda_n S_y^2$, $Var_s(\bar{x}_j) = \lambda_n S_{x_j}^2$, and $\lambda_n = (1-f)/n$ where $f = n/N$. Let also denote the correlations $\rho_{0j} = S_{yx_j}/S_y S_{yx_j}$ and $\rho_{12} = S_{x_1x_2}/S_{x_1}S_{x_2}$. Several kinds of bivariate ratio estimators are obtained by the combination of functions $f_\theta(.;.)$ as explained nextafter.

3.1 Exemples of generalized estimators

When denoting new coefficients, $a_j = a_j(\theta_j)$ and $b_j = b_j(\theta_j)$ per function $f_{\theta_j}(.;.)$, let define the following estimators for two auxiliary variables X_1 and X_2.

- The additive parametric ratio estimator is defined via a weighted sum of ratio estimators as follows,

$$\bar{y}_{R_f^a} = (\alpha_1 f_{\theta_1}(\bar{x}_1;\bar{X}_1) + \alpha_2 f_{\theta_2}(\bar{x}_2;\bar{X}_2))\,\bar{y}. \tag{4}$$

  Some related literature is in (Olkin, 1958; Rao and Mudholkar, 1967; Singh and Espejo, 2003; Abu-Dayyeh et al., 2003). When there are or not constraints on α to sum to one, the closed-form expression for $\hat{\alpha}$ is found in (Abu-Dayyeh et al., 2003) for the cases of quotients. See also (Kumar and Vishwakarma, 2017) for a more complex sampling design.

- The multiplicative parametric ratio estimator is defined via a weighted product of ratio estimators as follows,

$$\bar{y}_{R_f^m} = f_{\theta_1}(\bar{x}_1;\bar{X}_1)^{\alpha_1} f_{\theta_2}(\bar{x}_2;\bar{X}_2)^{\alpha_2}\bar{y}. \tag{5}$$

  Some related literature is in (Singh, 1967), (Srivastava, 1971; Abu-Dayyeh et al., 2003). As for the additive case, there are several possible cases. It may be chosen a geometrical mean with $\alpha_j = 1/p$ (but negative weights could be better if the sign of a_j is wrong) or just $\alpha_j = 1$ as met in some estimators in the literature. When $a_j = \pm 1$, the closed-form expressions for the mse are found in (Singh, 1965, 1967) for the case of quotients.

- The parametric combinations of additive and multiplicative ratio estimators are written as follows:

$$\begin{aligned}
\bar{y}_{R_f^{am}} &= \left(\alpha_+ + \alpha_- f_{\theta_j}(\bar{x}_1;\bar{X}_1) f_{\theta_2}(\bar{x}_2;\bar{X}_2)\right)\bar{y} \\
\bar{y}_{R_f^{am2}} &= \left(\alpha_+ f_{\theta_1}(\bar{x}_1;\bar{X}_1) + \alpha_- f_{\theta_1}^{-1}(\bar{x}_1;\bar{X}_1)\right) f_{\theta_2}(\bar{x}_2;\bar{X}_2)\bar{y} \\
\bar{y}_{R_f^{am3}} &= \left(\alpha_+ f_{\theta_1}(\bar{x}_1;\bar{X}_1) f_{\theta_2}(\bar{x}_2;\bar{X}_2) + \alpha_- f_{\theta_1}^{-1}(\bar{x}_1;\bar{X}_1) f_{\theta_2}^{-1}(\bar{x}_2;\bar{X}_2)\right)\bar{y} \\
\bar{y}_{R_f^{am4}} &= \left(\alpha_+ f_{\theta_{1+}}(\bar{x}_1;\bar{X}_1) f_{\theta_{2+}}(\bar{x}_2;\bar{X}_2) + \alpha_- f_{\theta_{1-}}(\bar{x}_1;\bar{X}_1) f_{\theta_{2-}}(\bar{x}_2;\bar{X}_2)\right)\bar{y}.
\end{aligned} \tag{6}$$

  The first one is related to (Upadhyaya, Singh, and Vos, 1985; Singh, Yadav, and Pal, 2018) for mean estimation. The second one is a parametric bivariate ratio-product estimator extending the estimator in (Singh and Espejo, 2003) by averaging a function $f_\theta(.;.)$ and its inverse while adding a product for the second variable. This model generalizes several contribution to the literature. When $X_1 = X_2$ such as for variance estimation in (Muneer, Khalil, Shabbir, and Narjis, 2018), this is a particular case of the additive parametric ratio estimator introduced just above. The third one is related to (Yasmeen, Amin, and Hanif, 2015; Adichwal, Sharma, and Singh, 2017). See (Singh and Yadav, 2018) for a comparison of this kind of estimators. The fourth one is related to (Solanki and Singh, 2015; Singh and Yadav, 2018), this generalized estimator looks more general than the three other ones.

- A parametric bivariate combined ratio and regression estimator is written:

$$\bar{y}_{R_f^c} = (k_0\bar{y} + k_1\delta_{x_1}) f_\theta(\bar{x}_2;\bar{X}_2)\,. \tag{7}$$

For combining a regression estimator with several auxiliary variables, diverse models have been developped in the literature. Pioneer reseach with this kind of estimator can be found in (Kadilar and Cingi, 2004; Gupta and Shabbir, 2008; Haq and Shabbir, 2013). The case when two different auxiliary variables are introduced, one in the regression part and one in the ratio part is presented further herein. Such estimator is presented by (Muneer et al., 2017) for particular values $a = -0.5$ and $b = 3/8$ or $b = 1/8$. And in (Hanif, Hamad, and Shahbaz, 2009) with $k_0 = 1$ and with the same function $f_\theta(.;.)$ than in the ratio-product estimator (Singh and Espejo, 2003), its authors have shown that its minimum mse is the same than for the difference estimator.

- Other parametric bivariate combined ratio and regression estimators are written:

$$\begin{aligned}
\bar{y}_{\mathrm{R}_f^{c2}} &= (k_0\bar{y} + k_1\delta_{x_1} + k_2\delta_{x_2}) f_{\theta_1}(\bar{x}_1;\bar{X}_1) f_{\theta_2}(\bar{x}_2;\bar{X}_2) \\
\bar{y}_{\mathrm{R}_f^{c3}} &= k_0\bar{y} f_{\theta_1}(\bar{x}_1;\bar{X}_1)^{\alpha_1} f_{\theta_2}(\bar{x}_2;\bar{X}_2)^{\alpha_2} + k_1\delta_{x_1} + k_2\delta_{x_2} \\
\bar{y}_{\mathrm{R}_f^{c4}} &= k_0\bar{y} + k_1\delta_{x_1} f_{\theta_2}(\bar{x}_2;\bar{X}_2) + k_2\delta_{x_2} f_{\theta_1}(\bar{x}_1;\bar{X}_1)\,.
\end{aligned} \tag{8}$$

For combining a regression estimator with a generic function in the ratio part, see for instance (Lu, 2017; Shabbir, Gupta, and Ahmed, 2018) for the first alternative model and (Kadilar and Cingi, 2005) for the second alternative model. With generic functions introduced in the ratio part, this underlines the generality of the parametric estimators through the function $f_\theta(.;.)$. When $k_0 \neq 1$, the mse of the combined estimator $\bar{y}_{\mathrm{R}_f^{c2}}$ generalizes several ones from the literature. When $f_{\theta_1} = f_{\theta_2} = 1$, the estimator $\bar{y}_{\mathrm{R}_f^{c2}}$ reduces to the one in (Rao, 1991; Lu, 2017) denoted $\bar{y}_{\mathrm{R}_{rao91}} = k_0\delta_y + k_1\delta_{x_1} + k_2\delta_{x_2}$ which reduces the mse of $\bar{y}_{\mathrm{R}_{diff}}$ when $k_0 = 1$ as explained next section. Note that other expressions such as in (Shahzad, Hanif, and Koyuncu, 2018) for instance are not considered herein.

3.2 Usual approximated mse (to second order) of estimators

In this subsection, we consider the case when two auxiliary variables are available and also third order terms in the approximations of $f_{\theta_1}(.;.)$ and $f_{\theta_2}(.;.)$ respectively with c_1 and c_2 as the coefficients associated to $\delta_{x_1}^3$ and $\delta_{x_2}^3$, such that a generalized estimator $\bar{y}_{\mathrm{R}_{est}}$ is as follows:

$$\begin{aligned}
\bar{y}_{\mathrm{R}_{est}} - \bar{Y} \;\doteq\; & T + U_0\delta_y + \sum_j U_j\delta_{x_j} + \sum_j V_{jj}\delta_{x_j}^2 + \sum_j V_{0j}\delta_y\delta_{x_j} + \sum_{j,k;k>j} V_{jk}\delta_{x_j}\delta_{x_k} \\
& + \sum_j W_{jjj}\delta_{x_j}^3 + \sum_{j,k} W_{jkk;k\neq j}\delta_{x_j}\delta_{x_k}^2 + \sum_j W_{0jj}\delta_y\delta_{x_j}^2 + \sum_{j,k;k>j} W_{0jk}\delta_y\delta_{x_j}\delta_{x_k}\,.
\end{aligned} \tag{9}$$

The expressions of the corresponding values for T, U_0, U_j, V_{jj}, V_{0j}, V_{jk}, W_{jjj}, W_{0jj}, W_{jkk} and W_{0jk}, depend on the given estimator. The expectation of the expression just before leads to the bias of the generalized estimator. For the mean squared error it is obtained that:

$$\begin{aligned}
(\bar{y}_{\mathrm{R}_{est}} - \bar{Y})^2 \;\doteq\; & \left[T + U_0\delta_y + \sum_j U_j\delta_{x_j}\right]^2 \\
& + 2\left[T + U_0\delta_y + \sum_j U_j\delta_{x_j}\right]\left[\sum_j V_{jj}\delta_{x_j}^2 + \sum_j V_{0j}\delta_y\delta_{x_j} + \sum_{j,k;k>j} V_{jk}\delta_{x_j}\delta_{x_k}\right] \\
& + 2T\left[\sum_j W_{jjj}\delta_{x_j}^3 + \sum_{j,k;k\neq j} W_{jkk}\delta_{x_j}\delta_{x_k}^2 + \sum_j W_{0jj}\delta_y\delta_{x_j}^2 + \sum_{j,k;k>j} W_{0jk}\delta_y\delta_{x_j}\delta_{x_k}\right].
\end{aligned} \tag{10}$$

The expectation w.r.t. the sampling may lead to the mean squared error in the general case which is a new result to our knowledge. This general expression remains relevant even when $T = 0$ which is the case of the multiplicative parametric ratio estimator for instance. When the third order terms are removed, this leads to

the approximated mean squared error:

$$\begin{aligned}
\mathrm{mse}\left[\bar{y}_{\mathrm{R}_{est}}\right] &\doteq \mathrm{amse}\left[\bar{y}_{\mathrm{R}_{est}}\right] \\
\mathrm{amse}\left[\bar{y}_{\mathrm{R}_{est}}\right] &= \lambda_n U_0^2 C_0^2 + \lambda_n \sum_{j,k} U_j U_k C_{jk} + 2T\lambda_n \sum_{j,k;k>j} V_{jk} C_{jk} + 2T\lambda_n \sum_j V_{jj} C_j^2 \\
&\quad + 2\lambda_n \sum_j (TV_{0j} + U_j U_0) C_{0j} + T^2 \\
&= \lambda_n U_0^2 C_0^2 + 2\lambda_n (TV_{11} + 0.5U_1^2) C_1^2 + 2\lambda_n (TV_{22} + 0.5U_2^2) C_2^2 + T^2 \\
&\quad + 2\lambda_n (TV_{12} + U_1 U_2) C_{12} + 2\lambda_n (TV_{01} + U_1 U_0) C_{01} + 2\lambda_n (TV_{02} + U_2 U_0) C_{02}\,.
\end{aligned} \tag{11}$$

The first row in the expression above may be relevant for more than one auxiliary variable while the second row provides directly the mean squared error one when the unknown values for T, U_0, U_1, U_2, V_1, V_2, V_{01}, V_{02}, and V_{12} are filled for two auxiliary variables. Note that instead of directly minimizing the amse (11) w.r.t. these parameters, the ratio framework prefers to introduce constraints which depend on a smaller set of unknown parameters as considered in section 4.

3.3 Linearizing approximated mse of estimators

The random variable $\bar{y}_{\mathrm{R}_{est}}$ is a function of $\delta = (\delta_y, \delta_{x_1}, \delta_{x_2})^T$, as follows:

$$\bar{y}_{\mathrm{R}_{est}} \doteq Z(\delta)\,. \tag{12}$$

An expression of the variance via a linearization is proposed for more completenesss. The corresponding mean squared error may be computed as follows at the first order:

$$\begin{aligned}
\mathrm{mse}\left[\bar{y}_{\mathrm{R}_{est}}\right] &\doteq \mathrm{amse}_L\left[\bar{y}_{\mathrm{R}_{est}}\right] \\
\mathrm{amse}_L\left[\bar{y}_{\mathrm{R}_{est}}\right] &= \frac{\partial Z}{\partial \delta^T} \mathrm{Var}[\delta] \frac{\partial Z}{\partial \delta} + \mathrm{bias}^2\left[\bar{y}_{\mathrm{R}_{est}}\right].
\end{aligned} \tag{13}$$

The positivity of the variance and the squared lead to the positivity of this amse. This new approximation of the mse adds several terms to the usual one in order to insure its positivity:

$$\begin{aligned}
\mathrm{amse}_L\left[\bar{y}_{\mathrm{R}_{est}}\right] &= \mathrm{amse}\left[\bar{y}_{\mathrm{R}_{est}}\right] + \mathrm{bmse}_L\left[\bar{y}_{\mathrm{R}_{est}}\right] \\
\mathrm{bmse}_L\left[\bar{y}_{\mathrm{R}_{est}}\right] &= \lambda_n^2 \left(V_{11} C_1^2 + V_{22} C_2^2 + V_{01} C_{01} + V_{02} C_{02} + V_{12} C_{12}\right)^2.
\end{aligned} \tag{14}$$

Next, several examples of mean squared errors are computed with the usual and the linearizing approaches for two auxiliary variables.

4 Approximation of the mean squared error for several estimators

The following results on the mse are obtained for our proposed generic estimators.

4.1 Approximated mse of the additive and multiplicative estimators

The mse for the additive bivariate estimator and the multiplicative estimators are written from:

$$\begin{aligned}
\bar{y}_{\mathrm{R}_f^a} &\doteq \left\{ \sum_{j=1}^{2} \alpha_j + \sum_{j=1}^{2} \alpha_j a_j \delta_{x_j} + \sum_{j=1}^{2} \alpha_j b_j \delta_{x_j}^2 \right\} \bar{y}. \\
\bar{y}_{\mathrm{R}_f^m} &\doteq \left\{ 1 + \sum_{j=1}^{2} \alpha_j a_j \delta_{x_j} + \alpha_1 a_2 \alpha_1 a_2 \delta_{x_1} \delta_{x_2} + \sum_{j=1}^{2} \alpha_j \frac{2b_j + (\alpha_j - 1) a_j^2}{2} \delta_{x_j}^2 \right\} \bar{y}.
\end{aligned} \tag{15}$$

This results into the following expressions entering the general amse in (11) listed in table 2. The corresponding expressions for the approximated mean squared error are discusses nextafter.

Table 2: Parameterization for the additive and multiplicative estimators.

	T	U_0	U_j	V_{jj}	V_{0j}	V_{12}
$\bar{y}_{\mathrm{R}_f^a}$	$(\alpha_1+\alpha_2-1)\bar{Y}$	$(\alpha_1+\alpha_2)\bar{Y}$	$a_j\alpha_j\bar{Y}$	$b_j\alpha_j\bar{Y}$	$a_j\alpha_j\bar{Y}$	0
$\bar{y}_{\mathrm{R}_f^m}$	0	$\bar{Y}$	$a_j\alpha_j\bar{Y}$	$\alpha_j\frac{2b_j+(\alpha_j-1)a_j^2}{2}\bar{Y}$	$a_j\alpha_j\bar{Y}$	$a_1\alpha_1a_2\alpha_2\bar{Y}$

Case $\alpha_1 = 1-\alpha_2 = \alpha$
The additive and multiplicative estimators have identical mse at the first order:

$$\mathrm{mse}^{(\alpha_1+\alpha_2=1)}\left[\bar{y}_{\mathrm{R}_f^a}\right] \doteq \mathrm{mse}^{(\alpha_1+\alpha_2=1)}\left[\bar{y}_{\mathrm{R}_f^m}\right] \doteq \lambda_n\bar{Y}^2\left\{C_0^2+2\sum_{j=1}^{2}\alpha_j a_j C_{0j}+\sum_{j,k=1}^{2}\alpha_j\alpha_k a_j a_k C_{jk}\right\}. \tag{16}$$

A direct computation or the general expression (11) with $U_0=\bar{Y}$, $U_j=\bar{Y}\alpha_j a_j$ while $T=0$ leads to this approximated mse. The optimal value of α minimizing the mse of $\bar{y}_{\mathrm{R}_f^a}$ and $\bar{y}_{\mathrm{R}_f^m}$ is found by replacing the expression of α_1 and α_2 depending of α in the mse $\left[\bar{y}_{\mathrm{R}_f^a}\right]$, and derivating w.r.t. α. The solution generalizes the usual one when $a_j=-1$ for a quotient, it is written:

$$\alpha_{\text{opt}} \quad = \quad \left\{a_2C_{02}-a_1C_{01}+a_2^2C_2^2-a_1a_2C_{12}\right\}\left\{a_1^2C_1^2+a_2^2C_2^2-2a_1a_2C_{12}\right\}^{-1}. \tag{17}$$

Note the identical analytical expression in (Abu-Dayyeh et al., 2003) with quotients when a_j is for an exponentiation. When the optimal values $\alpha_{1;\text{opt}}=\alpha_{\text{opt}}$ and $\alpha_{2;\text{opt}}=1-\alpha_{\text{opt}}$ are inserted in the expression of the mean squared error, it is denoted $\mathrm{mse}_{(\min)}^{(\alpha_1+\alpha_2=1)}$ for both $\bar{y}_{\mathrm{R}_f^a}$ and $\bar{y}_{\mathrm{R}_f^m}$. The resulting mse are minimum for these two estimators under constraints.

Case $\alpha_1+\alpha_2\neq 1$
The additive and multiplicative estimators have different mse. The additive bivariate estimator and its combinations with the multiplicative one are not considered further herein, and left as a perspective. For the multiplicative bivariate estimator with any value of a_1, a_2, b_1 and b_2, the mean squared error obtained after finding the optimal values for α_1 and α_2 without constraints or by recognizing the estimator in the first terms from (15) is written for the usual case:

$$\mathrm{mse}_{(\min)}^{(\alpha_1+\alpha_2\neq 1)}\left[\bar{y}_{\mathrm{R}_f^m}\right] \doteq \mathrm{mse}\left[\bar{y}_{\mathrm{R}_{diff}}\right], \tag{18}$$

where,

$$\mathrm{mse}\left[\bar{y}_{\mathrm{R}_{diff}}\right] \doteq \lambda_n C_0^2\bar{Y}^2\left\{1-\frac{\rho_{01}^2+\rho_{02}^2-2\rho_{12}\rho_{01}\rho_{02}}{1-\rho_{12}^2}\right\}. \tag{19}$$

Here $\bar{y}_{\mathrm{R}_{diff}}=\bar{y}+k_1\delta_{x_1}+k_2\delta_{x_2}$ denotes the difference estimator where the computed mse is such that k_1 and k_2 minimize the mse depending on these parameters. Hence at the first order, according to this result, the mean squared error does not depend on the parameters a_j and b_j for this combination of functions, and the mse remains identical for any function $f_{\theta_1}(.;.)$ and $f_{\theta_2}(.;.)$. This is a complement to the similar result known for the product of two functions in a ratio adjustement with just quotients (Abu-Dayyeh et al., 2003).

4.2 Approximated mse of the combined ratio and regression estimators

Case $k_0 = 1$

When for $\bar{y}_{\mathrm{R}_f^{c2}}$, $k_0 = 1$, while k_1, k_2 are free parameters or when for $\bar{y}_{\mathrm{R}_f^{c3}}$, $k_0 = 1$, α_1, α_2 are free parameters while k_1 and k_2 are the regression coefficients (assimilated to the population ones (Kadilar and Cingi, 2005)), or for $\bar{y}_{\mathrm{R}_f^{c4}}$ the free parameters are k_1, k_2 with $k_0 = 1$, then the mses of the four estimators are written from the expectation of the squared of:

$$\begin{array}{rcl} \bar{y}_{\mathrm{R}_f^{c}} - \bar{Y} & \doteq & \delta_y \bar{Y} + k_1 \delta_{x_1} + \bar{Y} a \delta_{x_2} \\ \bar{y}_{\mathrm{R}_f^{c2}} - \bar{Y} & \doteq & \delta_y \bar{Y} + (k_1 + a_1 \bar{Y}) \delta_{x_1} + (k_2 + a_2 \bar{Y}) \delta_{x_2} \\ \bar{y}_{\mathrm{R}_f^{c3}} - \bar{Y} & \doteq & \delta_y \bar{Y} + (k_1 + \alpha_1 a_1 \bar{Y}) \delta_{x_1} + (k_2 + \alpha_2 a_2 \bar{Y}) \delta_{x_2} \\ \bar{y}_{\mathrm{R}_f^{c4}} - \bar{Y} & \doteq & \delta_y \bar{Y} + k_1 \delta_{x_1} + k_2 \delta_{x_2} \, . \end{array} \tag{20}$$

Thus it is recognized the difference estimator such that at the first order approximation, the estimators have same mse for optimal values of k_1, k_2, α_1, α_2, or a respectively such that:

$$\mathrm{mse}_{(\min)}^{*(k_0=1)} \left[\bar{y}_{\mathrm{R}_f^{c}} \right] \doteq \mathrm{mse}_{(\min)}^{(k_0=1)} \left[\bar{y}_{\mathrm{R}_f^{c2}} \right] \doteq \mathrm{mse}_{(\min)}^{(k_0=1)} \left[\bar{y}_{\mathrm{R}_f^{c3}} \right] \doteq \mathrm{mse}_{(\min)}^{(k_0=1)} \left[\bar{y}_{\mathrm{R}_f^{c4}} \right] \doteq \mathrm{mse} \left[\bar{y}_{\mathrm{R}_{diff}} \right] . \tag{21}$$

Case $k_0 \neq 1$

The corresponding terms as defined from the generalized expression proposed in (11) are presented in table 3 for each estimator.

Table 3: Parameterization for four combined estimators and the Rao estimator.

	$\bar{y}_{\mathrm{R}_f^{c}}$	$\bar{y}_{\mathrm{R}_f^{c2}}$	$\bar{y}_{\mathrm{R}_f^{c3}}$	$\bar{y}_{\mathrm{R}_f^{c4}}$	$\bar{y}_{\mathrm{R}_{rao91}}$
T	$(k_0-1)\bar{Y}$	$(k_0-1)\bar{Y}$	$(k_0-1)\bar{Y}$	$(k_0-1)\bar{Y}$	$(k_0-1)\bar{Y}$
U_0	$k_0\bar{Y}$	$k_0\bar{Y}$	$k_0\bar{Y}$	$k_0\bar{Y}$	$k_0\bar{Y}$
U_1	k_1	$k_1 + a_1 k_0 \bar{Y}$	$k_1 + a_1\alpha_1 k_0 \bar{Y}$	k_1	k_1
U_2	$ak_0\bar{Y}$	$k_2 + a_2 k_0 \bar{Y}$	$k_2 + a_2\alpha_1 k_0 \bar{Y}$	k_2	k_2
V_{11}	0	$a_1 k_1 + b_1 k_0 \bar{Y}$	$b_1 k_0 \bar{Y}$	0	0
V_{22}	$bk_0\bar{Y}$	$a_2 k_2 + b_2 k_0 \bar{Y}$	$b_2 k_0 \bar{Y}$	0	0
V_{01}	0	$a_1 k_0 \bar{Y}$	$a_1 k_0 \bar{Y}$	0	0
V_{02}	$ak_0\bar{Y}$	$a_2 k_0 \bar{Y}$	$a_2 k_0 \bar{Y}$	0	0
V_{12}	ak_1	$a_1 k_2 + a_2 k_1 + a_1 a_2 k_0 \bar{Y}$	$a_1 a_2 k_0 \bar{Y}$	$k_1 a_2 + k_2 a_1$	0

In comparaison to the Rao estimator (Rao, 1991; Lu, 2017), the combined estimatord adds new terms in the general expression (11), hence may be able to improve the mse for some populations. As an example, the estimator $\bar{y}_{\mathrm{R}_f^{c}}$ and $\bar{y}_{\mathrm{R}_f^{c4}}$ are considered nextafter in the generalized cases while the other combined estimators (regression and additive) are left for perspective.

Approximated mse of $\bar{y}_{\mathbf{R}_f^{c}}$

- The mse for the parametric bivariate combined ratio and regression is written as a function of k_0 and k_1, a and b by taking the expectation of the squared of:

$$\bar{y}_{\mathrm{R}_f^{c}} - \bar{Y} \doteq k_0\bar{Y} + k_0\delta_y\bar{Y} + k_1\delta_{x_1} + k_0\bar{Y}a\delta_{x_2} + k_0\bar{Y}a\delta_y\delta_{x_2} + k_1 a\delta_{x_1}\delta_{x_2} + k_0\bar{Y}b\delta_{x_2}^2 - \bar{Y} \, . \tag{22}$$

Thus,

$$\begin{aligned}\mathrm{mse}_{(k_0,k_1)}\left[\bar{y}_{\mathrm{R}_f^c}\right] &\doteq \lambda_n C_0^2 k_0^2 \bar{Y}^2 + 2b\lambda_n C_2^2 k_0^2 \bar{Y}^2 + a^2\lambda_n C_2^2 k_0^2 \bar{Y}^2 + k_0^2\bar{Y}^2 - 2a\lambda_n C_{02} k_0 \bar{Y}^2 - 2k_0\bar{Y}^2 \\ &\quad -2b\lambda_n C_2^2 k_0 \bar{Y}^2 + \bar{Y}^2 + 2\lambda_n k_0 k_1 \bar{Y}(C_{01} + 2aC_{12}) - 2a\lambda_n C_{12} k_1 \bar{Y} + \lambda_n C_1^2 k_1^2 \\ &\doteq A k_0^2 + B k_1^2 + 2C k_0 k_1 - 2D_0 k_0 - 2D_1 k_1 + E\,.\end{aligned} \tag{23}$$

Where,

$$\begin{aligned} A &= \bar{Y}^2 + \bar{Y}^2\lambda_n\left(C_0^2 + (a^2+2b)C_2^2 + 4aC_{02}\right) \\ B &= \lambda_n C_1^2 \\ C &= 2\bar{Y}\lambda_n(aC_{12} + 0.5C_{01}) \\ D_0 &= \bar{Y}^2 + \bar{Y}^2\lambda_n\left(bC_2^2 + aC_{02}\right) \\ D_1 &= a\bar{Y}\lambda_n C_{12} \\ E &= \bar{Y}^2\,. \end{aligned} \tag{24}$$

- There are two cases to solve for identifying the optimal values of k_0 and k_1.

  - If $k_0 \neq 1$, in order to find the optimal values of k_0 and k_1 by minimizing the mse, the system coming from the derivatives is solved. The solution if its exists is given by:

$$\begin{aligned} k_{0;\mathrm{opt}} &= \{BD_0 - CD_1\}\left\{AB - C^2\right\}^{-1} \\ k_{1;\mathrm{opt}} &= \{AD_1 - CD_0\}\left\{AB - C^2\right\}^{-1}\,. \end{aligned} \tag{25}$$

    Thus, the minimum value for the mse is obtained when these two optimal values $k_{0;\mathrm{opt}}$ and $k_{1;\mathrm{opt}}$ are inserted in the amse (23). The resulting minimal mse is a quotient of two polynomials of the two variables a and b.

  - If $k_0 = 1$, the mean squared error is rewritten as follows, $\mathrm{mse}_{(1,k_1)}\left[\bar{y}_{\mathrm{R}_f^c}\right] \doteq Bk_1^2 + 2(C - D_1)k_1 + A - 2D_0 + E$. When $k_0 = 1$ and constant, the minimum is reached at the new optimal solution given by $k_{1;\mathrm{opt}} = (D_1 - C)B^{-1}$. The resulting mse is rewritten $(A - 2D_0 + E) - (D_1 - C)^2 B^{-1}$, and is higher than when k_0 is a free parameter. Note that on the contrary to the previous expression when k_0 is a free parameter, this expression of the mse does not depend on b because the term depending on b cancels out by difference in $A - 2D_0$.
    When X_1 and X_2 are not perfectly correlated ($\rho_{12} \neq 0$), it can be shown that it exists a value of the scalar a where its mean squared error is minimum at $a = a_{\mathrm{opt}}$,

$$a_{\mathrm{opt}} = \left\{C_{01}C_{12} - C_1^2 C_{02}\right\}\left\{C_1^2 C_2^2 - C_{12}^2\right\}^{-1}\,. \tag{26}$$

    The corresponding minimum is found equal to:

$$\mathrm{mse}_{(\min)}^{*(k_0=1)}\left[\bar{y}_{\mathrm{R}_f^c}\right] \doteq \mathrm{mse}\left[\bar{y}_{\mathrm{R}_{diff}}\right]\,. \tag{27}$$

    Note that this mse does not depend on a or b anymore, hence it corresponds to all the functions $f_\theta(.;.)$ with same values $a = a_{\mathrm{opt}}$ as higher order derivatives do not enter the mse at the first order.

At the first order approximation, the approximated mse of the generalized estimator related to (Hanif et al., 2009; Muneer et al., 2017) can be smaller than several usual estimators when the scalars a and b are well chosen.

- When the linearizing amse is considered, after adding the terms coming from bmse_L in (14), the new coefficients to enter the linear system may be written as follows, $\breve{D}_0 = D_0$, $\breve{D}_1 = D_1$, $\breve{E} = E$, and, $\breve{A} = A + \bar{Y}^2\lambda_n^2(bC_2^2 + aC_{02})^2$, $\breve{B} = B + a^2\lambda_n^2 C_{12}^2$, $\breve{C} = C + a\bar{Y}\lambda_n^2(bC_2^2 + aC_{02})C_{12}$. As expected only terms with λ_n^2 add up to the new expressions. Similarly than for the usual form of the mean squared error, the coefficients k_0 and k_1 are found by solving a linear program. Thus, the new minimum value for the linearizing amse is obtained when these two new optimal values $\breve{k}_{0;\text{opt}}$ and $\breve{k}_{1;\text{opt}}$ are inserted in the mse.

Approximated mse of $\bar{y}_{\mathbf{R}_f^{c4}}$

- When $k_0 \neq 1$, the mse of the combined estimator $\bar{y}_{\mathrm{R}_f^{c4}}$ is written from the following difference which is squared for computing the mse:

$$\bar{y}_{\mathrm{R}_f^{c4}} - \bar{Y} \doteq (k_0 - 1)\bar{Y} + k_0\bar{Y}\delta_y + k_1\delta_{x_1} + k_2\delta_{x_2} + (a_1k_2 + a_2k_1)\delta_{x_1}\delta_{x_2}\,. \tag{28}$$

The mean squared error at the first order is written as proposed in the formula (11) for instance. For its minimization, the mean squared error is rewritten via a quadratic function as follows:

$$\begin{aligned}\text{mse}_{(k_0,k_1,k_2)}\left[\bar{y}_{\mathrm{R}_f^{c4}}\right] \doteq\ & Ak_0^2 + Bk_1^2 + Ck_2^2 + 2Dk_0k_1 + 2Ek_0k_2 + 2Fk_1k_2 \\ & -2G_0k_0 - 2G_1k_1 - 2G_2k_2 + H\,.\end{aligned} \tag{29}$$

With coefficients:

$$\begin{aligned} A &= \bar{Y}^2(1 + \lambda_n C_0^2) \\ B &= \lambda_n C_1^2 \\ C &= \lambda_n C_2^2 \\ D &= \bar{Y}\lambda_n(a_2C_{12} + C_{01}) \\ E &= \bar{Y}\lambda_n(a_1C_{12} + C_{02}) \\ F &= \lambda_n C_{12} \\ G_0 &= \bar{Y}^2 \\ G_1 &= a_2\lambda_n C_{12}\bar{Y} \\ G_2 &= a_1\lambda_n C_{12}\bar{Y} \\ H &= \bar{Y}^2\,. \end{aligned} \tag{30}$$

The solutions for the values of k_0, k_1 and k_2 for this estimator are found in closed-form as follows. Let denote the symmetric inverted matrix with cell values $A' = BC - F^2$, $B' = AC - E^2$, $C' = AB - D^2$, $D' = EF - CD$, $E' = DF - BE$, and $F' = DE - AF$, and the inverted matricial determinant $\Delta'_- = \{AA' + DD' + EE'\}^{-1}$ from the linear problem. The optimal solution minimizing the amse is given by:

$$\begin{aligned} k_{0;\text{opt}} &= \{A'G_0 + D'G_1 + E'G_2\}\Delta'_- \\ k_{1;\text{opt}} &= \{D'G_0 + B'G_1 + F'G_2\}\Delta'_- \\ k_{2;\text{opt}} &= \{E'G_0 + F'G_1 + C'G_2\}\Delta'_-\,. \end{aligned} \tag{31}$$

The minimum value for the mse is obtained when these three optimal values $k_{0;\text{opt}}$, $k_{1;\text{opt}}$ and $k_{2;\text{opt}}$ are inserted in the amse (29). The resulting mean squared error is a function of a_1, a_2, b_1 and b_2.

- When the linearizing amse is considered, after adding the terms coming from bmse_L in (14), the new coefficients to enter the linear system may be written as follows, $\breve{A} = A$, $\breve{D} = D$, $\breve{E} = E$, $\breve{G}_0 = G_0$, $\breve{G}_1 = G_1$, $\breve{G}_2 = G_2$, $\breve{H} = H$, and, $\breve{B} = B + a_2^2\lambda_n^2 C_{12}^2$, $\breve{C} = C + a_1^2\lambda_n^2 C_{12}^2$, $\breve{F} = F + a_1a_2\lambda_n^2 C_{12}^2$. Similarly than for the usual form of the mean squared error, the coefficients k_0, k_1 and k_2 are found by solving a linear program. Thus, the new minimum value for the linearizing amse is obtained when these two new optimal values $\breve{k}_{0;\text{opt}}$, $\breve{k}_{1;\text{opt}}$ and $\breve{k}_{2;\text{opt}}$ are inserted in the amse (29).

4.3 Approximated mse of $\bar{y}_{\mathbf{R}_{est}}$

For the minimization of the amse, the usual way from the literature needs to identify the quantities A, B, ... which becomes more tedious when the ratio model increases in complexity. We propose a new approach more generic via matrix algebra which provides a numerical solution such as finding an analytical expression via matricial products becomes facultative for visualization purpose. For our proposed generic approach, let denote $\Psi = (T, U_0, U_1, U_2, V_{11}, V_{22}, V_{01}, V_{02}, V_{12})^T$, and:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & \lambda_n C_1^2 & \lambda_n C_2^2 & \lambda_n C_{01} & \lambda_n C_{02} & \lambda_n C_{12} \\ 0 & \lambda_n C_0^2 & \lambda_n C_{01} & \lambda_n C_{02} & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_n C_{01} & \lambda_n C_1^2 & \lambda_n C_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_n C_{02} & \lambda_n C_{12} & \lambda_n C_2^2 & 0 & 0 & 0 & 0 & 0 \\ \lambda_n C_1^2 & 0 & 0 & 0 & \lambda_n^2 C_1^4 & \lambda_n^2 C_1^2 C_2^2 & \lambda_n^2 C_{01} C_1^2 & \lambda_n^2 C_{02} C_1^2 & \lambda_n^2 C_1^2 C_{12} \\ \lambda_n C_2^2 & 0 & 0 & 0 & \lambda_n^2 C_1^2 C_2^2 & \lambda_n^2 C_2^4 & \lambda_n^2 C_{01} C_2^2 & \lambda_n^2 C_{02} C_2^2 & \lambda_n^2 C_{12} C_2^2 \\ \lambda_n C_{01} & 0 & 0 & 0 & \lambda_n^2 C_{01} C_1^2 & \lambda_n^2 C_{01} C_2^2 & \lambda_n^2 C_{01}^2 & \lambda_n^2 C_{01} C_{02} & \lambda_n^2 C_{01} C_{12} \\ \lambda_n C_{02} & 0 & 0 & 0 & \lambda_n^2 C_{02} C_1^2 & \lambda_n^2 C_{02} C_2^2 & \lambda_n^2 C_{01} C_{02} & \lambda_n^2 C_{02}^2 & \lambda_n^2 C_{02} C_{12} \\ \lambda_n C_{12} & 0 & 0 & 0 & \lambda_n^2 C_1^2 C_{12} & \lambda_n^2 C_{12} C_2^2 & \lambda_n^2 C_{01} C_{12} & \lambda_n^2 C_{02} C_{12} & \lambda_n^2 C_{12}^2 \end{pmatrix}. \quad (32)$$

Let denote $\tilde{K} = (1, K^T)^T$ where $K = (k_0, k_1, k_2)^T$ or any other vector of coefficients $K = (k_0, k_1)^T$ for instance. Let also denote the matrix Φ which maps the vector $\tilde{K}$ to the parameters Ψ, hence $\Psi = \Phi\tilde{K}$ and it is defined by two blocks $\Phi = [\xi_0 | \xi_K]$. For instance, for $\bar{y}_{\mathrm{R}_f^{c2}}$, $\bar{y}_{\mathrm{R}_f^{c4}}$, $\bar{y}_{\mathrm{R}_f^{c}}$ and $\bar{y}_{\mathrm{R}_f^{a}}$, it is obtained respectively $\Phi = \Phi_{\mathrm{Rc2}}$, $\Phi = \Phi_{\mathrm{Rc4}}$, $\Phi = \Phi_{\mathrm{Rc}}$ and $\Phi = \Phi_{\mathrm{Ra}}$ where:

$$\Phi_{\mathrm{Rc2}} = \begin{pmatrix} -\bar{Y} & \bar{Y} & 0 & 0 \\ 0 & \bar{Y} & 0 & 0 \\ 0 & a_1\bar{Y} & 1 & 0 \\ 0 & a_2\bar{Y} & 0 & 1 \\ 0 & b_1\bar{Y} & a_1 & 0 \\ 0 & b_2\bar{Y} & a_2 & 0 \\ 0 & a_1\bar{Y} & 0 & 0 \\ 0 & a_2\bar{Y} & 0 & 0 \\ 0 & a_1a_2\bar{Y} & a_2 & a_1 \end{pmatrix}, \Phi_{\mathrm{Rc4}} = \begin{pmatrix} -\bar{Y} & \bar{Y} & 0 & 0 \\ 0 & \bar{Y} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & a_2 & a_1 \end{pmatrix},$$
$$\Phi_{\mathrm{Rc}} = \begin{pmatrix} -\bar{Y} & \bar{Y} & 0 \\ 0 & \bar{Y} & 0 \\ 0 & 0 & 1 \\ 0 & a\bar{Y} & 0 \\ 0 & 0 & 0 \\ 0 & b\bar{Y} & 0 \\ 0 & 0 & 0 \\ 0 & a\bar{Y} & 0 \\ 0 & 0 & a \end{pmatrix}, \Phi_{\mathrm{Ra}} = \begin{pmatrix} -\bar{Y} & \bar{Y} & \bar{Y} \\ 0 & \bar{Y} & \bar{Y} \\ 0 & a_1\bar{Y} & 0 \\ 0 & 0 & a_2\bar{Y} \\ 0 & b_1\bar{Y} & 0 \\ 0 & 0 & b_2\bar{Y} \\ 0 & a_1\bar{Y} & 0 \\ 0 & 0 & a_2\bar{Y} \\ 0 & 0 & 0 \end{pmatrix}. \quad (33)$$

Then,

$$\begin{aligned} \mathrm{amse}_L[\bar{y}_{\mathrm{R}_{est}}] &= \Psi^T Q \Psi \\ &= \xi_0^T Q \xi_0 + 2\xi_0^T Q \xi_K K + K^T \xi_K^T Q \xi_K K. \end{aligned} \quad (34)$$

Thus, after derivation, the system to solve for identifying the optimal value of K is as follows:

$$\xi_K^T Q \xi_0 + \xi_K^T Q \xi_K K = 0. \quad (35)$$

When three parameters are involved, this leads to retrieve the notation in the paragraph just before:

$$\begin{pmatrix} G_0 \\ G_1 \\ G_2 \end{pmatrix} = -\xi_K^T Q \xi_0 \quad \text{and} \quad \begin{pmatrix} A & D & E \\ D & B & F \\ E & F & C \end{pmatrix} = \xi_K^T Q \xi_K. \quad (36)$$

When the solution K_{opt} is inserted in the expression of the approximated mse, the amse is minimum. This may be relevant for one variable or more than two variables by updating Q as a perspective.

5 Experiments

In summary to the previous section, the generalized expressions of the amse of the considered estimators with the functions $f_\theta(.;.)$ lead to complementary results. To demonstrate the interest of using the parameterization for the comparison of estimators, it is proposed to perform an empirical analysis of ratio estimators in this section.

5.1 Empirical settings

The empirical results allow to check the behavior of several estimators with diverse datasets when the function $f_\theta(.;.)$ is in stake. They are:

- $\bar{y}_1$ for the estimator denoted $\bar{y}_{\text{R}_{rao91}}$ in the previous sections as a competitive baseline,
- $\bar{y}_2$ for the additive estimator $\bar{y}_{\text{R}_f^a}$ with $a_1 = a_2 = a$ varying and $\alpha_1 + \alpha_2 = 1$,
- $\bar{y}_3$ for the difference estimator $\bar{y}_{\text{R}_{diff}}$ which does not depend on the quantity a,
- $\bar{y}_4$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with $a = -0.5$ and $b = 0.375$ (see subsection 3.1),
- $\bar{y}_5$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with a varying, $b = 0.375$, and with no constraint,
- $\bar{y}_6$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with $a = a(\gamma)$ and $b = b(\gamma)$ (see subsection 2.2),
- $\bar{y}_7$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with a (and b) varying but with the constraint $k_0 = 1$,
- $\bar{y}_8$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with $a = a_{\text{opt}}$ and $b = 0.375$,
- $\bar{\bar{y}}_8$ for combined estimator $\bar{y}_{\text{R}_f^c}$ with $a = a_{\text{opt}}$ and $b = b_{\text{opt}}$,
- $\bar{y}_9$ for the combined estimator $\bar{y}_{\text{R}_f^c}$ with a varying, $b \in \{0.05, 0.15, 0.25, 0.35, 0.45, 0.55\}$.

The experiments are based on several datasets for comparing the considered ratio estimators with real data. The other settings are the choice for the functions in the ratio estimator and the choice of an indicator for summarizing the numerical informations.

Parametric functions The functions for the ratio adjustement part are as follows.

- In (Diana and Perri, 2010) a particular function is defined $f_\theta(\bar{x};\bar{X}) = \frac{\bar{X}+\gamma}{\bar{x}+\gamma}$ with $\theta = \gamma$. From the second-order polynomial approximation, the corresponding coefficients a and b from the serie approximation are $a(\gamma)$ and $b(\gamma)$ given at subsection 2.2. Note that the range for γ is chosen in order to have the range of a equal to $[-1.2; +1.2]$ while removing the values around 0.
- In the generic case, $f_\theta(\bar{x};\bar{X}) \doteq 1 + a\delta_x + b\delta_x^2$, even if a and b are allowed to change jointly, b is generally fixed to $b = 3/8$ in the proposed experiments.

Indicators The indicator computed in the experiments for an estimator $\bar{y}_k$ where $1 \leq k \leq 8$ is defined as follows, $PRE = \lambda_n \bar{Y}^2 C_0^2 / \text{mse}[\bar{y}_k]$, and similarly for $\bar{\bar{y}}_8$. This is the gain when the variance of the usual mean estimator is compared with the mean squared error of a given mean estimator.

Real data The four selected datasets are $D1$ (Muneer et al., 2017) (Data 1, page 2185), $D2$ (Muneer et al., 2017) (Data 3, page 2185) and $D3$ (Abu-Dayyeh et al., 2003) (page 296) and an other dataset from the literature. Their corresponding statistics entering in the expression of the estimators are in the table 4 below. Note that the value $n = 8$ is small but seems suitable for the computation of the mse according to the experiments.

Table 4: Statistics for the considered populations.

	N	n	$\bar{Y}$	$\bar{X}_1$	$\bar{X}_2$	C_0^2	C_1^2	C_2^2	C_{12}	ρ_{01}	ρ_{02}	ρ_{12}
$D1$	100	29	2.364	2.925	5.2390	2.5582	0.0661	0.0461	0.0047	0.1602	0.0829	0.0846
$D2$	97	30	3135.619	3050.278	2743.9587	4.8674	5.4812	6.2422	3.5810	0.8072	0.8501	0.6122
$D3$	332	80	1093.100	181.570	143.3100	0.7626	0.7684	0.7616	0.6441	0.9730	0.8620	0.8420
$D4$	18	8	13.797	2.444	38.4440	0.1864	12.5025	0.1184	0.4950	0.8210	0.5903	0.4069

5.2 Empirical results with the usual amse

- As a can vary in the generalized expressions of the mse, it is observed graphically the shape of the indicator named *PRE* when b is fixed to $3/8$ while a varies in $[-1.2;+1.2]$ in order to observe the variations of the different mse and compare their values. On the figure 1, it is shown from the graphical output of RStudio in (R Core Team, 2017), the indicators *PRE* for most of the estimators considered, as a curve depending on the scalar a. The value of the indicator *PRE* for the estimator $\bar{y}_4$ is at the intersection of the curve for $\bar{y}_5$ and the vertical line passing through $a = -0.5$.

Table 5: Empirical numerical results with the indicator PRE for the four datasets and height estimators for the first height columns. (*) The written solutions for $\bar{y}_5$, $\bar{y}_6$ and $\bar{y}_7$ are numerical optima: they are not algebrical hence may be not directly useful for practical interest because generally an explicit solution is required for the estimates. At the six last columns, empirical numerical results with the indicator PRE for the four datasets with the estimator $\bar{\bar{y}}_8$ found by optimizing the approximation of the mse, while the corresponding optimal values for a, b, k_0 and k_1 are also presented.

	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_5$ (*)	$\bar{y}_6$ (*)	$\bar{y}_7$ (*)	$\bar{y}_4$	$\bar{y}_8$	$\bar{\bar{y}}_8$	$k_{0;\text{opt}}$	$k_{1;\text{opt}}$	a_{opt}	b_{opt}
D1	1.0941	1.0287	1.0315	1.0944	1.0941	1.0943	1.0944	1.0943	1.2159	0.910	-2.073	-0.5203	37.4471
D2	6.9710	6.6723	6.8589	7.4174	6.9753	7.3578	7.3658	7.3579	9.7233	1.076	-1379.910	-0.5027	-0.3041
D3	21.2856	21.2583	21.2783	21.8135	21.2856	21.8127	12.0856	21.8126	21.2871	0.999	-924.634	-0.1469	0.0265
D4	4.0559	2.4561	4.0430	4.0755	4.0560	4.0754	3.9588	4.0754	4.0481	0.998	-1.1706	-0.3853	0.0048

- The numerical results corresponding to the figure 1 with the indicator *PRE* are in Table 5. The optimization of the parameter a in $\bar{y}_5$ is able to decrease the mse of the estimator $\bar{y}_4$ for at least one dataset out of four when a_{opt} (see table 5) is very different from -0.5. According to the figure 1, the mse of the estimators $\bar{y}_{\text{R}_f^a}$ and $\bar{y}_{\text{R}_f^m}$ with $\alpha_2 = 1 - \alpha_1$ and $a_1 = a_2$ is minimum for a particular value of a depending on the population and for which it reaches the one of the difference estimator: this result may ask to solve for a quartic equation in order find the optimal analytical value of a. Visually, the graphics allow to compare the estimators and empirically validate the expressions and bounds of the mse. The additional term bmse in (14) divived by the amse for each dataset is respectively equal to $2.6\,10^{-7}$, $2.1\,10^{-2}$, $2.7\,10^{-3}$ and $6.1\,10^{-4}$, thus these numbers associated to quadratic shapes of the curves in figure 1 may justify the computation of the usual amse instead of the linearizing one for the considered populations.

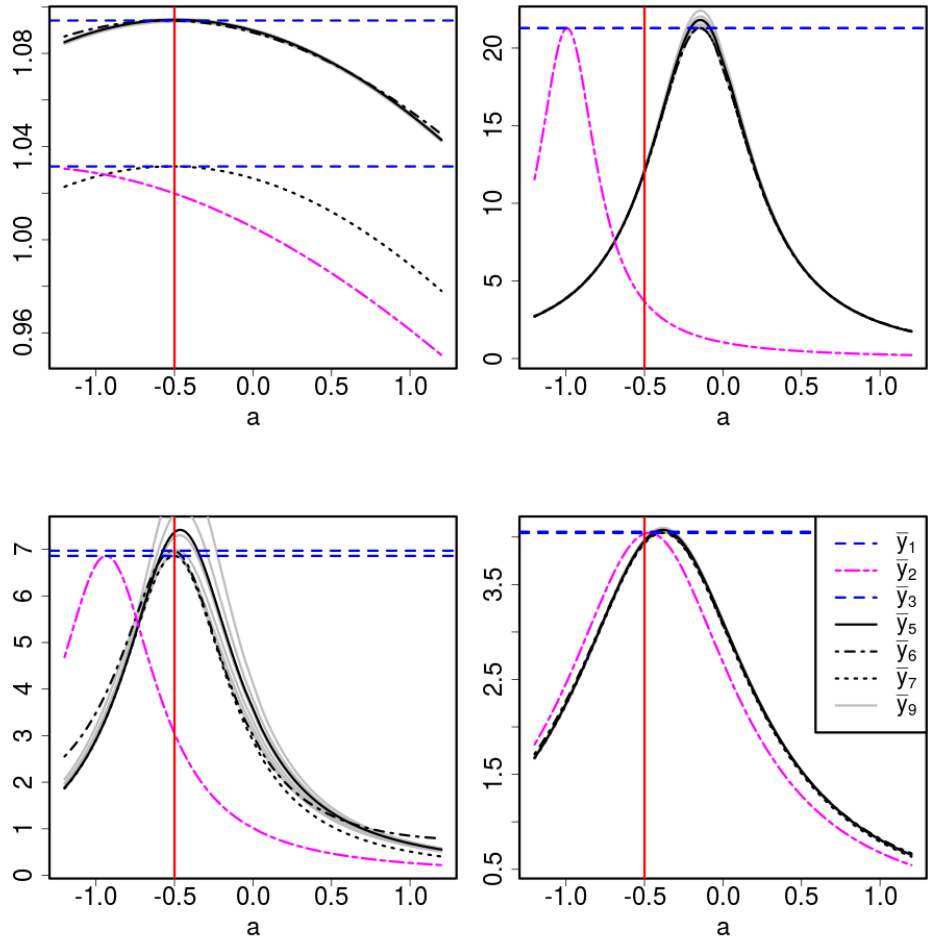

Figure 1: Curves for the populations $D1$ (top) and $D2$ (bottom) in the left column $D3$ (top) and $D4$ (bottom) in the right column, of the indicator *PRE* as a function of the varying quantity a.

- Considering the smaller mean squared error of $\bar{y}_8$ for some populations, it is optimized analytically the amse of $\bar{y}_{\mathrm{R}_f^c}$ with respect to a and b despite the nonlinearities. A solution leads to the same result for a equal to a_{opt} also and for b a new analytical value with the following expression,

$$b_{\text{opt}} = \frac{C_{12}^2\left(C_{12}^2C_0^2+C_1^2{C_{02}}^2-2C_{01}C_{12}C_{02}\right)-2C_1^2C_2^2\left(C_1^2{C_{02}}^2+C_{12}^2C_0^2-2C_{01}C_{12}C_{02}\right)+\left(C_1^2C_0^2-{C_{01}}^2\right)C_1^2{C_2}^4}{\left({C_1}^4{C_2}^4-2C_1^2C_{12}^2C_2^2+{C_{12}}^4\right)C_2^2}. \tag{37}$$

Changing the value of b while keeping the same value for a implies that the estimator denoted $\bar{y}_8$ may be improved with this new value of b. But due to the nonlinearities of the mse w.r.t. the parameters a and b, the value of b_{opt} looks less reliable than a_{opt} which is obtained from a quadratic equation: there is no insurance that the resulting solution is a minimum of the mse according to the conducted numerical results. When denoting the corresponding estimator $\bar{\bar{y}}_8$, these results are summarized in table 5 at the six last columns.

- The case when a and b are both free to vary jointly may be presented for each population in a three dimensional space with dimensions a, b, and *PRE*. This graphics leads to a visual comparison of the estimator for different values of a and b, each point of the surface is a different estimator such as it is

possible to check if it is enough near the optimum given the chosen intervals. In figure 1, it is shown several sections of the surface denoted as $\bar{y}_9$ for different values $b \in \{0.05, 0.15, 0.25, 0.35, 0.45, 0.55\}$. For two datasets, the surface does not depend on b for the range of values considered while mostly for $D2$ and slightly for $D3$ it does.

6 Conclusion and perspectives

Herein, we propose to review and analyze graphically several existing ratio estimators from the literature via generic models when two auxiliary variables are available. The main difference with previous serie approximations is to consider the relative differences for the means, the derivatives as fully variables and the visualization when the derivatives take their values inside intervals. This brings a complementary view of their behaviors as it becomes possible to compare visually their efficiency (and eventually their bias) while checking the validity of their mse approximation in the vicinity of a chosen function for modeling the ratios. The main perspective remains a further study of the behaviour of the mses plus the bias w.r.t. the function $f_\theta(.;.)$ with eventually higher orders in the approximations.

References

Abu-Dayyeh, W., M. Ahmed, R. Ahmed, and H. Muttlak (2003, 7). Some estimators of a finite population mean using auxiliary information. *Applied Mathematics and Computation 139*(2-3), 287–298.

Adichwal, N. K., P. Sharma, and R. Singh (2017, Jun). Generalized class of estimators for population variance using information on two auxiliary variables. *International Journal of Applied and Computational Mathematics 3*(2), 651–661.

Allen, J., H. P. Singh, and F. Smarandache (2003). A family of estimators of population mean using multi-auxiliary information in presence of measurement errors. *International Journal of Social Economics 30*, 837–849.

Bahl, S. and R. Tuteja (1991). Ratio and product type exponential estimators. *Journal of Information and Optimization Sciences 12*(1), 159–164.

Bhushan, S. and C. Kumari (2018). A new log type estimator for estimating the population variance. *International Journal of Computational and Applied Mathematics (IJCAM) 13*(1), 43–54.

Cochran, W. G. (1940). The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science 30*(2), 262–275.

Diana, G., M. Giordan, and P. F. Perri (2011, Jun). An improved class of estimators for the population mean. *Statistical Methods & Applications 20*(2), 123–140.

Diana, G. and P. F. Perri (2007). Estimation of finite population mean using multi-auxiliary information. *Metron LXV*(1), 99–112.

Diana, G. and P. F. Perri (2010). In *Using auxiliary information under a generic sampling design, Compstat*.

Gupta, S. and J. Shabbir (2008). On improvement in estimating the population mean in simple random sampling. *Journal of Applied Statistics 35*(5), 559–566.

Hanif, M., N. Hamad, and M. Q. Shahbaz (2009). A modified regression type estimator in survey sampling. *World Applied Sciences Journal 12*(7), 1559–1561.

Haq, A. and J. Shabbir (2013). Improved family of ratio estimators in simple and stratified random sampling. *Communications in Statistics - Theory and Methods 42*(5), 782–799.

John, S. (1969). On multivariate ratio and product estimators. *Biometrika 56*(3), 533–536.

Kadilar, C. and H. Cingi (2004, April). Ratio estimators in simple random sampling. *Appl. Math. Comput. 151*(3), 893–902.

Kadilar, C. and H. Cingi (2005). A new estimator using two auxiliary variables. *Applied Mathematics and Computation 162*(2), 901 – 908.

Khoshnevisan, M., R. Singh, P. Chauhan, N. Sawan, and F. Smarandache (2007). A general family of estimators for estimating population mean using known value of some population parameter(s). *Far East Journal of Theoretical Statistics 22*(2), 181–191.

Kumar, M. and G. K. Vishwakarma (2017, Dec). Estimation of mean in double sampling using exponential technique on multi-auxiliary variates. *Communications in Mathematics and Statistics 5*(4), 429–445.

Lu, J. (2017). Efficient estimator of a finite population mean using two auxiliary variables and numerical application in agricultural,biomedical, and power engineering. *Hindawi, Mathematical Problems in Engineering*.

Muneer, S., A. Khalil, J. Shabbir, and G. Narjis (2018). A new improved ratio-product type exponential estimator of finite population variance using auxiliary information. *Journal of Statistical Computation and Simulation 88*(16), 3179–3192.

Muneer, S., J. Shabbir, and A. Khalil (2017). Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. *Communications in Statistics - Theory and Methods 46*(5), 2181–2192.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika 45*(1/2), 154–165.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, P. S. R. S. and G. S. Mudholkar (1967). Generalized multivariate estimator for the mean of finite populations. *Journal of the American Statistical Association 62*(319), 1009–1012.

Rao, T. J. (1991). On certain methods of improving ratio and regression estimators. *Communications in Statistics - Theory and Methods 20*(10), 3325–3340.

Shabbir, J., S. Gupta, and S. Ahmed (2018). A generalized class of estimators under two-phase stratified sampling for non response. *Communications in Statistics - Theory and Methods 0*(0), 1–17.

Shahzad, U., M. Hanif, and N. Koyuncu (2018, Sep). A new estimator for mean under stratified random sampling. *Mathematical Sciences 12*(3), 163–169.

Shukla, G. K. (1966). An alternative multivariate ratio estimate for finite population. *Calcutta Statistical Association Bulletin 15*(2-3), 127–134.

Singh, H. P. and M. R. Espejo (2003). On linear regression and ratio-product estimation of a finite population mean. *Journal of the Royal Statistical Society. Series D (The Statistician) 52*(1), 59–67.

Singh, H. P. and A. Yadav (2018). Improved generalized family of estimators of population mean using information on transformed auxiliary variables. *Pakistan Journal of Statistics and Operation Research*, 913–934.

Singh, H. P., A. Yadav, and S. K. Pal (2018). An efficient use of two auxiliary variables in stratified random sampling. *Int. J. Agricult. Stat. Sci. 14*(2), 739–747.

Singh, M. P. (1965). On the estimation of ratio and product of the population parameters. *Sankhya B 27*, 231–328.

Singh, M. P. (1967, Dec). Ratio cum product method of estimation. *Metrika 12*(1), 34–42.

Solanki, R. S. and H. P. Singh (2015, Feb). Efficient classes of estimators in stratified random sampling. *Statistical Papers 56*(1), 83–103.

Srivastava, S. K. (1971). A generalized estimator for the mean of a finite population using multi-auxiliary information. *Journal of the American Statistical Association 66*(334), 404–407.

Srivastava, S. K. and H. S. Jhajj (1981). A class of estimators of the population mean in survey sampling using auxiliary information. *Biometrika 68*(1), 341–343.

Srivastava, S. K. and H. S. Jhajj (1983). A class of estimators of the population mean using multi auxiliary information. *Calcutta Statistical Association Bulletin 32*(1-2), 47–56.

Upadhyaya, L., H. Singh, and J. Vos (1985). On the estimation of population means and ratios using supplementary information. *Statistica Neerlandica 39*(3), 309–318.

Vishwakarma, G. K. and M. Kumar (2015, Dec). An efficient class of estimators for the mean of a finite population in two-phase sampling using multi-auxiliary variates. *Communications in Mathematics and Statistics 3*(4), 477–489.

Yadav, S. K. and C. Kadilar (2013). Improved class of ratio and product estimators. *Applied Mathematics and Computation 219*(22), 10726–10731.

Yasmeen, U., M. N. Amin, and M. Hanif (2015). Generalized exponential estimators of finite population mean using transformed auxiliary variate. *International Journal of computational Mathematics 1*(2), 1–10.

Complementary contents to 'Visualization of generalized mean estimators using auxiliary information in survey sampling'

Appendix 1: Approximated mse from a quadratic function

From the definion of the matrices Q and Φ and the resulting matricial expression of the amse, the optimal vector K which minimizes the amse is defined as follows:

$$\begin{aligned} K_{\text{opt}} &= argmin_K \, \text{amse}_L\left[\bar{y}_{\text{R}_{est}}\right] \\ &= -(\xi_K^T Q \xi_K)^{-1} \xi_K^T Q \xi_0 \,. \end{aligned} \tag{38}$$

Note that for the constraints such as the sum of the components of K is one, we need to add a Lagrangian which leads to update this unconstrained solution K_{opt} as in the usual linear regression with constraints. The constraint is typically the sum to one for an additive estimator. More generally let suppose $\Omega K = \omega$, for instance $\omega = 1$ while Ω is a vector of 1 in order to insure $k_0 + k_1 = 1$. The constrained solution is written:

$$K_{\text{opt}}^c = K_{\text{opt}} + (\xi_K^T Q \xi_K)^{-1} \Omega^T [\Omega (\xi_K^T Q \xi_K)^{-1} \Omega^T]^{-1} (\omega - \Omega K_{\text{opt}}) \,. \tag{39}$$

When the unconstrained solution K_{opt} is inserted in the matricial expression of the amse, the amse is minimum:

$$\text{amse}_{L(min)}\left[\bar{y}_{\text{R}_{est}}\right] = \xi_0^T Q \xi_0 - \xi_0^T Q \xi_K (\xi_K^T Q \xi_K)^{-1} \xi_K^T Q \xi_0 \,. \tag{40}$$

The corresponding bias is denoted $\text{bias}_{(opt)}\left[\bar{y}_{\text{R}_{est}}\right]$, it is written afterwards when the quantities T, U_0, U_1, U_2, V_{11}, V_{22}, V_{01}, V_{02}, and V_{12} from the optimal vector $\Psi_{(opt)} = \Phi(1, \tilde{K}_{(opt)}^T)^T$ are replaced in the expectation of $[\bar{y}_{\text{R}_{est}} - \bar{Y}]$.

Note that a common requirement for a mean squared error is that it becomes small (Diana et al., 2011) with λ_n and cancels out when the sample equals the whole population. When $n = N$, this induces that $\lambda_n = 0$ hence the matrix $(\xi_K^T Q \xi_K)$ is not invertible because Q reduces to a zero matrix except that its first cell (top-left) is equal to 1, thus this singularity may be addressed as a perspective.

Appendix 2: Approximated mse of estimators via sampling

The resampling procedures are usually used after a first step such as an estimation: this is a way to find the variability of the estimates. For instance, boostrapping and Jackknifing lead to the variance and the bias wanted for constructing an approximate mse from a sample. This supposes the existence of a sample which is the case when one wants to estimate a population mean from a sample mean. This is an alternative to the subsection on a direct linearization in order to avoid any analytical approximation. The corresponding numerical approximation of the mean squared error is thus as follows:

$$\begin{aligned} \text{mse}\left[\bar{y}_{\text{R}_{est}}\right] &\doteq \widehat{\text{amse}}_{BJ}\left[\bar{y}_{\text{R}_{est}}\right] \\ \widehat{\text{amse}}_{BJ}\left[\bar{y}_{\text{R}_{est}}\right] &= \widehat{E}_s\left(\widehat{\text{Var}}_{BJ}\left[\bar{y}_{\text{R}_{est}}\right]\right) + \widehat{E}_s\left(\widehat{\text{bias}}_{BJ}\left[\bar{y}_{\text{R}_{est}}\right]\right)^2 . \end{aligned} \tag{41}$$

In the case of the bootstrap, the resampling statistics are written for instance as follows:

$$\begin{aligned} \widehat{\text{Var}}_B\left[\bar{y}_{\text{R}_{est}}\right] &= \tfrac{N-n}{N} \tfrac{1}{L} \textstyle\sum_{j=1}^L (\bar{y}_{bj} - \bar{y})^2 \\ \widehat{\text{bias}}_B\left[\bar{y}_{\text{R}_{est}}\right] &= \tfrac{1}{L} \textstyle\sum_{j=1}^L \bar{y}_{bj} - \bar{y} \,. \end{aligned} \tag{42}$$

The positivity of the variance and the squared lead to the positivity of this amse. Note that other expressions from sampling for the bias, variance and mean squared error are available such as from the Jackknife framework. A nice property of the resampling approach is to avoid any serie expansion while bringing an objective function for any hidden parameter such as the regression coefficients. If some exponentiation is involved, it is required here an optimisation via a numerical or an expansion approach.