The quasi-regression form of calibration estimates

László Mihályffy,
Retired senior statistical adviser of the HCSO
E-mail: laszlo.mihalyffy@ksh.hu

For an arbitrary calibration estimator, an alternative expression called quasi-regression form is can be used in variance computations. In the case of simple random sampling it yields an explicit expression for the difference between the estimated variance of the arbitrary calibrated estimate and that of the generalized regression estimate.

KEYWORDS:
Estimations.

In the literature on calibration methods, the *Deville–Särndal* paper [1992] is a key reference. It is shown in that paper that under some mild conditions any calibration estimator is asymptotically equivalent to the generalized regression estimator (called henceforth GREG), and therefore the variance and the estimated variance of the latter may be used for the former. A small Monte Carlo study with simple random samples of size $n = 200$ from a population consisting of $N = 2000$ units has yielded practically the same variance for the most common calibration estimators in use.

In this paper a method is given to assign approximate variance and sample estimate of the variance – different from those of the GREG – to an arbitrary calibration estimator. By the Deville–Särndal principle, these variances will be quite close to their counterparts corresponding to the GREG, yet in some cases the difference may be interesting, and the extra computing needed is not substantial. The idea of our method is to re-write a given calibration estimator in a form similar to that of the GREG, and then the variance and the variance estimate can be determined in a similar way as in the case of the latter. The GREG in this paper plays the role of the baseline, therefore we begin with a brief review on that estimator.

Provided we are given a sample $\{1,\ 2,\ ...,\ n\}$ from a finite universe of size *N*, and the design enables the use of the Horvitz–Thompson estimator, consider the following problem referred to as (P1) in the subsequent considerations. Find the calibrated weights $w_1,\ w_2,\ ...,\ w_n$ by minimising the distance function

$$\sum_{j=1}^{n}\left(w_j - d_j\right)^2 / d_j \ , \qquad /1/$$

subject to the calibration constraints

$$\sum_{j=1}^{n} x_{ji} w_j = X_i \ , \qquad i = 1,\ 2,\ ...,\ m \ . \qquad /2/$$

In equations /1/ and /2/, $d_1,\ d_2,\ ...,\ d_n$ stand for the design weights, $x_{j1},\ x_{j2},\ ...,\ x_{jm}$ are the values of the auxiliary variables observed on sample unit *j*, and $X_1,\ X_2,\ ...,\ X_m$ are the population totals of the auxiliary variables. The unique solution of the problem (P1) for w_j can be given explicitly, and the calibrated total of some study variable y_j can be written as

$$\hat{Y}^{\text{reg}} = \hat{Y} + \sum_{i=1}^{m} b_i \left(X_i - \hat{X}_i\right). \qquad /3/$$

$\hat{Y}^{\text{reg}}$ is called generalized regression estimate of the population total Y; $\hat{Y}$, $\hat{X}_1$, ... and $\hat{X}_m$ are Horvitz–Thompson estimates based on the design weights d_j, and b_1, b_2, ..., b_m are generalized regression coefficients estimated from the sample.

To emphasize the baseline function of the GREG in this paper, the results coming from the problem (P1) will be denoted with symbols having a superscript $(.)^{\text{o}}$; thus e.g. w_1^{o}, w_2^{o}, ..., w_n^{o} will stand for the calibrated weights and /3/ will be re-written as

$$\hat{Y}^{\text{reg}} = \hat{Y} + \sum_{i=1}^{m} b_i^{\text{o}} \left(X_i - \hat{X}_i \right). \qquad /3a/$$

Matrix algebra will often be used in this paper hence we need matrix-vector notations, too. Some of the most important of those are as follows. The superscript $(.)^T$ denotes transpose of matrices or vectors;

$$\mathbf{d} = \left(d_1, d_2, ..., d_n\right)^T,$$

$$\mathbf{w}^{\text{o}} = \left(w_1^{\text{o}}, w_2^{\text{o}}, ..., w_n^{\text{o}}\right)^T,$$

$$\mathbf{y} = \left(y_1, y_2, ..., y_n\right)^T,$$

$$\mathbf{x} = \left(x_{ji}\right), \quad j = 1, 2, ..., n, \quad i = 1, 2, ..., m,$$

$$\mathbf{b}^{\text{o}} = \left(b_1^{\text{o}}, b_2^{\text{o}}, ..., b_m^{\text{o}}\right)^T,$$

$\mathbf{\Omega}$ is the diagonal matrix with entries d_1, d_2, ..., d_n in the main diagonal.

Note that

$$\sum_{j=1}^{n} d_j y_j = \mathbf{d}^T \mathbf{y} = \hat{Y};$$

by analogy we have

$$\mathbf{d}^T \mathbf{x} = \left(\hat{X}_1, \hat{X}_2, ..., \hat{X}_m\right).$$

Further notations:

$$X = \left(X_1, X_2, ..., X_n\right)^T,$$

$$\hat{X} = \left(\hat{X}_1, \hat{X}_2, ..., \hat{X}_m\right)^T.$$

Except for the last two symbols, matrices and vectors are denoted by bold-face letters that may be capital, lower case or even Greek characters. Note also that with these notations the vector $\mathbf{b}^{\mathrm{o}}$ of regression coefficients can be written as follows:

$$\mathbf{b}^{\mathrm{o}} = \left(\mathbf{x}^T \boldsymbol{\Omega} \mathbf{x}\right)^{-1} \mathbf{x}^T \boldsymbol{\Omega} \mathbf{y} \ .$$

In some cases a generalized version of the problem (P1) is considered where the distance function /1/ has the following form:

$$\sum_{j=1}^{n} \left(w_j - d_j\right)^2 / q_j d_j \ , \qquad /1a/$$

and $q_1, q_2, \ldots, q_n$ are positive weights chosen properly. For any unit *j* in the sample or in the population, q_j can always be identified with the reciprocal of the variance σ_j^2 of the random variable Y_j in the super-population model, $j = 1, 2, \ldots, N$; see e.g. *Särndal*, *Swensson* and *Wretman* ([1992] p. 225–229.). However, the option of using weights q_j other than unity would have no impact on our conclusions therefore we assume throughout that $q_j = 1$ for all *j*. In any case, it is interesting to note that the estimator /3/ – or /3a/ – can be derived in two different ways: either by solving the calibration problem (P1) or by means of the super-population principle.

1. The general calibration estimator and its quasi-regression form

With the same assumptions on sample and universe as in the introductory section, consider the following calibration problem (P2). Find the calibrated weights $w_1, w_2, \ldots, w_n$ by minimising the distance function

$$F = F\left(w_1, w_2, \ldots, w_n, d_1, d_2, \ldots, d_n\right) \ , \qquad /4/$$

subject to the calibration constraints

$$\sum_{j=1}^{n} x_{ji} w_j = X_i \ , \qquad i = 1, 2, \ldots, m \ . \qquad /2/$$

and the individual bounds on the calibrated weights

$$L \le w_j / d_j \le U \ . \qquad /5/$$

The distance function F is supposed to be strictly convex and continuously differentiable at least twice. In the majority of cases it is also assumed that F is separable which means that it is of the form

$$F = \sum_{j=1}^{n} G\left(w_j, d_j\right),$$

where G is strictly convex and continuously differentiable at least twice; term j in this representation depends only on w_j and d_j.

Denote $\mathbf{w} = \left(w_1,\ w_2,\ ...,\ w_n\right)^T$ the unique solution of (P2) – distinguishing it in this way from the solution of (P1) – and denote $\hat{Y}^{\text{cal}}$ the calibrated estimate of Y with these weights. We point out the following.

Result 1. $\hat{Y}^{\text{cal}}$ can be written in form as follows:

$$\hat{Y}^{\text{cal}} = \hat{Y} + \sum_{i=1}^{m} b_i \left(X_i - \hat{X}_i\right) = \hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b}\ , \qquad /6/$$

where $\mathbf{b} = \mathbf{b}^{\text{o}} + \mathbf{b}'$, and

$$\begin{aligned}\mathbf{b}' = \frac{\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}}{\left(X - \hat{X}\right)^T \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right)} \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right) \stackrel{\text{def}}{=} \\ \stackrel{\text{def}}{=} C_{\text{o}} \left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1} \left(X - \hat{X}\right)\end{aligned}.$$

Note that $\mathbf{b}$ depends on the problem (P2) only through the expression $\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}$, and that $\hat{X}$ depends on the sample and the design weights d_j.

Proof. Starting with the right-hand side of /6/, we have

$$\begin{aligned}\hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b} = \hat{Y} + \left(X - \hat{X}\right)^T \mathbf{b}^{\text{o}} + \left(X - \hat{X}\right)^T \mathbf{b}' = \\ = \hat{Y}^{\text{reg}} + \left(X - \hat{X}\right)^T \mathbf{b}' = \hat{Y}^{\text{reg}} + \left(\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}\right),\end{aligned}$$

as was to be shown.

While Result 1 is almost trivial, expression /6/ is useful in examining the estimated variance of $\hat{Y}^{\text{cal}}$. It is easy to see that the existence of $\left(\mathbf{x}^T \mathbf{\Omega} \mathbf{x}\right)^{-1}$ is sufficient for that of $\hat{Y}^{\text{reg}}$ and also for the "quasi-regression" representation /6/, thus the term "quasi-regression form of calibrated estimates" is justified.

2. Linearization and variance expressions

With the quasi-regression forms introduced in the preceding section, one should proceed in the same way as in the case of "ordinary" regression estimates.
To this end:

– first the quasi-regression estimate should be linearized, then
– the linearized expression can be treated as the Horvitz–Thompson estimate of a total, and
– expressions for the variance and the sample estimate of the variance should be identified, and finally,
– the unknown population values in the variance estimate from the sample should be replaced by the corresponding sample estimates.

Before starting this procedure, the population value of the quasi-regression coefficients **b** should be found. This will be done for the two terms of $\mathbf{b} = \mathbf{b}^{\text{o}} + \mathbf{b}'$ separately. By the principle of the super-population model, the population value of $\mathbf{b}^{\text{o}}$ is $\mathbf{B}^{\text{o}}$, the vector of regression coefficients in the population ($\mathbf{B}^{\text{o}} \neq E\left(\mathbf{b}^{\text{o}}\right)$). As for **b'**, it is straightforward to take the expectation **B'** of **b'** over all samples in the design in consideration as population value. In cases where $(\mathbf{x}^T \mathbf{\Omega} \mathbf{x})^{-1}$ does not exist we take $\mathbf{b}^{\text{o}} = \mathbf{b}' = \mathbf{b} = 0$. The population value of **b** is then defined as $\mathbf{B} = \mathbf{B}^{\text{o}} + \mathbf{B}'$, its components will be denoted by $B_1, B_2, \ldots, B_m$.

Now we have to linearize $\hat{Y}^{\text{cal}}$ given by /6/. This estimated total depends on $\hat{Y}$, $\hat{X}_1, \hat{X}_2, \ldots \hat{X}_m$, and a certain number of other sample-depending values determined basically by the distance function F in /4/. Denote $\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_h$ these arguments of $\hat{Y}^{\text{cal}}$; we shall see soon that we need not to have much information on them. Differentiating yields

$$\partial \hat{Y}^{\text{cal}} / \partial \hat{Y} \equiv 1;$$

$$\partial \hat{Y}^{\text{cal}} / \partial \hat{X}_i = \sum_{\text{k}=1}^{\text{m}} \frac{\partial b_k}{\partial \hat{X}_i} \left(X_k - \hat{X}_k \right) - b_i, \qquad i = 1, 2, \ldots, m;$$

$$\partial \hat{Y}^{\text{cal}} / \partial \hat{z}_i = \sum_{k=1}^{m} \frac{\partial b_k}{\partial \hat{z}_i} \left(X_k - \hat{X}_k \right), \qquad i = 1, 2, \ldots, h.$$

Setting the arguments in the last two relations equal to the corresponding population values implies

$$\partial \hat{Y}^{\text{cal}} / \partial \hat{X}_i |_{\hat{X}_i = X_i} = -B_i, \quad i = 1,\ 2,\ \ldots,\ m, \text{ and } \partial \hat{Y}^{\text{cal}} / \partial \hat{z}_i |_{\hat{z}_i = z_i} = 0 .$$ [1]

This suggests that $\hat{Y}^{\text{lin}}$, the linearized version of $\hat{Y}^{\text{cal}}$ can be written as follows:

$$\hat{Y}^{\text{lin}} = Y + \left(\hat{Y} - Y\right) + \sum_{\text{k}=1}^{\text{m}} -B_k \left(\hat{X}_k - X_k\right) = \hat{Y} + \sum_{k=1}^{m} B_k \left(X_k - \hat{X}_k\right), \qquad /7/$$

i.e. the linearization yields that the quasi-regression coefficients b_i are replaced by the corresponding population values. From now on, variance expressions for $\hat{Y}^{\text{cal}}$ are derived in the same way as in the case of the ordinary regression estimator. The approximate variance of $\hat{Y}^{\text{cal}}$ is the variance of $\hat{Y}^{\text{lin}}$, and since $\sum_k B_k X_k$ is constant over all samples, we have

$$AV\left(\hat{Y}^{\text{cal}}\right) = Var\left(\hat{Y}^{\text{lin}}\right) = Var\left(\hat{Y} - \sum_{k=1}^{m} B_k \hat{X}_k\right) = Var\left(\hat{Z}\right),$$

where $\hat{Z}$ is the total of the residuals $z_j = y_j - \sum_{k=1}^{m} B_k x_{jk}$ weighted with the design weights d_j, and $Var\left(\hat{Z}\right)$ is computed with the variance formula of the Horvitz–Thompson estimator. The sample estimate of the variance is also based on the residuals z_j, but the unknown population values B_k should be replaced by the corresponding sample values b_k; moreover, Deville and Särndal advocate the use of calibrated weights w_j in variance estimates rather than that of d_j. It should be emphasized that in this way the estimated variance of $\hat{Y}^{\text{cal}}$ – and not that of $\hat{Y}^{\text{reg}}$ – is determined; and in practice presumably not the Yates–Grundy formula

$$var\left(\hat{Y}^{\text{cal}}\right) \approx var\left(\hat{Z}\right) = \sum_i \sum_{j>i} \frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}} \left(z_i / \pi_i - z_j / \pi_j\right)^2$$

will be used, but e.g. the jackknife method.

In the particular case of simple random sampling an explicit expression can be given for $var\left(\hat{Z}\right)$. We have the following.

Result 2. Assume that the design is simple random sampling without replacement and one of the auxiliary variables assumes the value 1 for each unit of the popula-

[1] The notation is simplified; all arguments in the partial derivatives should set equal to the corresponding population values.

tion.[2] In this case the following relation holds for the sample estimate of the variance of $\hat{Y}^{\text{cal}}$:

$$var\left(\hat{Y}^{\text{cal}}\right) \approx var\left(\hat{Z}\right) = var\left(\hat{Y}^{\text{reg}}\right) + var\left(\hat{X}b'\right) \qquad /8/$$

where $z_j = y_j - \sum_{k=1}^{m} b_k x_{jk}$ and $\hat{Z} = {N}/{n} \sum_{j=1}^{n} z_j$. Furthermore,

$$var\left(\hat{X}\mathbf{b}'\right) < \frac{(1\text{-}f)N^2}{n(n-1)} \frac{\left(\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}\right)^2}{\left(X - \hat{X}\right)^T \left(\mathbf{x}^T\mathbf{x}\right)^{-1} \left(X - \hat{X}\right)}, \qquad /9/$$

where $f = n/N$.

Proof. It is easy to see that the well-known estimated variance for an estimated total under simple random sampling (see *Cochran* [1977] p. 26.) can be re-written in matrix-vector form as follows;

$$var\left(\hat{Z}\right) = \frac{(1-f)N^2}{n(n-1)} \mathbf{z}^T \left(\mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^T\right)\mathbf{z},$$

where $\mathbf{z} = \left(z_1, z_2, ..., z_n\right)^T$, $\mathbf{I}$ is unit matrix of order n and $\mathbf{e}$ is a vector with each component being equal to 1. Thus we have

$$var\left(\hat{Z}\right) = C_1 \left(\mathbf{y}^T - \mathbf{b}^T\mathbf{x}^T\right)\left(\mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^T\right)\left(\mathbf{y} - \mathbf{x}\mathbf{b}\right) \qquad /10/$$

where $C_1 = \dfrac{(1-f)N^2}{n(n-1)}$. Now $\mathbf{b}^{\text{o}}$+$\mathbf{b}$’ should be substituted for **b.** We have to take into account that, owing to simple random sampling, the matrix $\mathbf{\Omega}$ in the expressions of $\mathbf{b}^{\text{o}}$ and **b**’ is now *N*/*n* times the unit matrix. However, the factor *N*/*n* will not occur in the formulae, since it always appears simultaneously in the numerator and in the denominator. Consequently, the factor $\mathbf{y} - \mathbf{x}\mathbf{b}$ becomes

$$\mathbf{y} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} - C_{\text{o}}\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\left(X - \hat{X}\right) =$$
$$= \mathbf{y} - \mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} - C_{\text{o}}\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\left(\mathbf{w}^{\text{o}} - \mathbf{d}\right),$$

or denoting the matrix $\mathbf{x}\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T$ by **P**,

[2] From the viewpoint of regression this means that there is an intercept.

$$\mathbf{y}-\mathbf{xb}=\mathbf{y}-\mathbf{Py}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)=\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right), \qquad /11/$$

where $C_{\mathrm{o}}=\dfrac{\hat{Y}^{\mathrm{cal}}-\hat{Y}^{\mathrm{reg}}}{\left(X-\hat{X}\right)^{T}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right)}$;

note that $\mathbf{\Omega}$ has disappeared from here, too. The matrix **P** is a symmetric projection and, because of the assumption on the auxiliary variable having the value 1 for any unit, the vector **e** is an eigenvector of $\mathbf{P}$: $\mathbf{Pe}=\mathbf{e}$. Substituting the right-hand side of /11/ for **y – xb** in /10/ implies

$$var\left(\hat{Z}\right)=C_1\left(\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{P}\right)-C_{\mathrm{o}}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)^{T}\mathbf{P}\right)\left(\mathbf{I}-\frac{1}{n}\mathbf{ee}^{T}\right)\left(\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}-C_{\mathrm{o}}\mathbf{P}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)\right)=$$
$$=C_1\mathbf{y}^{T}\left(\mathbf{I}-\mathbf{P}\right)\mathbf{y}+C_1C_{\mathrm{o}}^{2}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)^{T}\left(\mathbf{P}-\frac{1}{n}\mathbf{ee}^{T}\right)\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right).$$

Substituting here $\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\mathbf{x}^{T}$ for **P** and making use of the expression for $\mathbf{b}^{\mathrm{o}}$ and the relation $\mathbf{x}^{T}\left(\mathbf{w}^{\mathrm{o}}-\mathbf{d}\right)=X-\hat{X}$ one obtains

$$var\left(\hat{Z}\right)C_1\left(\mathbf{y}-\mathbf{xb}^{\mathrm{o}}\right)^{T}\left(\mathbf{I}-\frac{1}{n}\mathbf{ee}^{T}\right)\left(\mathbf{y}-\mathbf{xb}^{\mathrm{o}}\right)+$$
$$+C_1C_{\mathrm{o}}^{2}\left(\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right)\right)^{T}\left(\mathbf{I}-\frac{1}{n}\mathbf{ee}^{T}\right)\mathbf{x}\left(\mathbf{x}^{T}\mathbf{x}\right)^{-1}\left(X-\hat{X}\right).$$

Using again the argument that an additive constant of the form $\sum_k b_k X_k$ has no impact on the variance, it is easy to see that the right-hand side of the last equality equals $var\left(\hat{Y}^{\mathrm{reg}}\right)+var\left(\hat{X}\mathbf{b}'\right)$ which verifies /8/. Inequality /9/ follows by omitting the matrix $\mathbf{I}-\mathbf{ee}^{T}/n$ from the second term and making use of the fact that its norm equals /1/. The proof is thereby complete.

3. A numerical example

We have considered a universe consisting of $N=2899$ households. In those households, there were $X_1=1076$ individuals aged 15-24 years, $X_2=4239$ individuals aged 25-54 years, $X_3=1382$ individuals aged 55-74 years, $X_4=3193$ males aged 15-74 years, $X_5 = 3504$ females aged 15-74 years, and, finally $Y = 3656$ individuals aged 15-74 who participated in the labour market.

From this universe simple random samples consisting of 25 units were selected, thus the design weight was 116.96 for each unit in the samples. Using X_1, X_2, X_3, X_4, X_5 and $X_6 = N$ as controls,[3] two calibration estimates of Y were computed for each sample. One of them was $\hat{Y}^{\text{reg}}$, the baseline estimate, the other was $\hat{Y}^{\text{cal}}$ obtained with raking, obeying also the individual bounds $40 \le w_j \le 600$ for the final weights.

The following table shows $\hat{Y}^{\text{reg}}$, $\hat{Y}^{\text{cal}}$, $\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}} = \left(X - \hat{X}\right)^T \mathbf{b}'$ and the corresponding standard errors based on /8/ and /9/ for the first six samples.

Estimates and standard errors obtained with two calibration estimators for samples from an artificial population

Number of Sample	$\hat{Y}^{\text{reg}}$		$\hat{Y}^{\text{cal}}$		$\hat{Y}^{\text{cal}} - \hat{Y}^{\text{reg}}$	
	Estimate	S. E.	Estimate	S. E.	Estimate	S. E.
1	2878	308.9	2933	310.5	55	30.9
2	4815	331.4	4797	331.5	–18	7.2
3	3306	393.4	3346	394.1	40	24.1
4	3773	343.1	3739	344.0	–34	19.6
5	2884	253.7	2959	254.7	75	22.8
6	3494	409.4	3575	412.6	81	50.1

It might be surprising that the asymptotic equivalence of calibration estimators is manifest even at such moderate sizes of sample and population as $n = 25$, $N = 2899$.

References

COCHRAN, W. G. [1977]: *Sampling techniques*. John Wiley & Sons. New York.

DEVILLE, J-C. – SÄRNDAL, C-E. [1992]: Calibration estimators in survey sampling. *Journal of the American Statistical Association.* Vol. 37. p. 376–382.

SÄRNDAL, C-E. – SWENSSON, B. – WRETMAN, J. [1992]: *Model assisted survey sampling*. Springer Verlag. New York.

[3] Note that $X_1+X_2+X_3=X_4+X_5$.