SSC Annual Meeting, June 1996
Proceedings of the Survey Methods Section

ESTIMATING THE VARIANCE OF THE REGRESSION PARAMETER ESTIMATOR IN STATISTICS CANADA'S GENERALIZED ESTIMATION SYSTEM

A. Théberge[1]

ABSTRACT

Statistics Canada's Generalized Estimation System (GES) calculates survey estimates using auxiliary information. The possibility of providing certain diagnostic capabilities to the software is now being examined. As part of those diagnostics, it is required to estimate the variance of the regression parameter estimator. Some of the work done on this subject will be presented here. An overview of the calibration method used by GES will be given. The asymptotic variance of the regression parameter estimator will be derived. Two known approaches for estimating this asymptotic variance will be given, as well as a third new approach. Comparisons of the three estimators under Bernoulli sampling and under stratified random sampling will be done. Finally, some simulation results will be presented.

KEY WORDS: Asymptotic variance; Bernoulli sampling; Calibration; Generalized Estimation System; Regression parameter.

RÉSUMÉ

Le Système généralisé d'estimation (SGE) de Statistique Canada calcule des estimés d'une enquête en utilisant de l'information auxiliaire. La possibilité d'inclure certaines fonctions de diagnostics dans le logiciel est actuellement sous étude. Parmi ces diagnostics, on requiert l'estimation de la variance des estimés des paramètres de régression. Une partie du travail réalisé sur ce sujet sera présentée dans cet exposé. Un survol de la méthode d'étalonnage utilisée par le SGE sera donné. La variance asymptotique des estimés des paramètres de régression sera calculée. Deux approches connues pour estimer cette variance asymptotique seront exposées, ainsi qu'une troisième approche nouvelle. Nous comparerons les trois estimateurs pour un plan d'échantillonnage de Bernoulli et pour un plan d'échantillonnage aléatoire stratifié. Finalement, quelques résultats de simultation seront présentés.

MOTS CLÉS: Variance asymptotique; échantillonnage de Bernoulli; étalonnage; Système généralisé d'estimation; paramètre de régression.

1. INTRODUCTION

The Generalized Estimation System (GES) is one of many generalized survey tools that Statistics Canada has developed and maintains. Some of those other tools are the Generalized Sampling software (GSAM), Data Collection and Capture software (DC2), and the Generalized Edit and Imputation System (GEIS). The Generalized Estimation System can be used to compute estimates and variance estimates using auxiliary information. A more detailed description of the software can be found in Estevao, Hidiroglou and Särndal (1995).

Among the improvements we wish to make to GES, it is our intention to add diagnostic capabilities. In particular, we would like to estimate the variance of the regression parameter estimator in order to build confidence intervals or perform statistical tests. This paper will describe how these variance estimates could be derived.

The family of estimators used by GES can be viewed as calibration estimators. This calibration method is explained in Deville and Särndal (1992) and will be reviewed in Section 2. The asymptotic variance of the regression parameter estimator is derived in Section 3. In Section 4, three possible estimators of the asymptotic variance will be given and developed in more detail for the cases of Bernoulli sampling and of stratified random sampling. The results of Monte Carlo simulations comparing the variance estimates will be presented in Section 5. Finally, concluding remarks are made in Section 6.

[1] Alain Théberge, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6.

2. CALIBRATION

For a population of size N, let Y and C be two vectors in $\mathbb{R}^N$ with C assumed to be known. We wish to estimate $Y'C$ by $Y_s' W_s$ where Y_s is the sub-vector of Y that corresponds to the sample s of size n, and $W_s \in \mathbb{R}^n$ is a vector of weights for the sampled units. Let A be the diagonal matrix of sampling weights equal to the inverse of the units' probabilities of being included in the sample. We want to minimize the distance between W_s and $A_s C_s$ under the constraint $X_s' W_s = X'C$, where $X \in \mathbb{R}^{N\times p}$ is a matrix of auxiliary information, and the subscript s denotes the sub-matrix or sub-vector corresponding to the sample s. We will assume that X_s is of full rank. This formulation of the problem is slightly more general than the one given in Deville and Särndal (1992) where C is assumed to be a vector of ones. Let $\| \alpha \|_{U_s}^2 = \alpha' U_s \alpha$ be the square of the norm of $\alpha \in \mathbb{R}^n$, where $U_s \in \mathbb{R}^{n\times n}$ is a positive diagonal matrix. Using this norm as a distance, if Q is a positive diagonal matrix, we can look for a vector of weights W_s which minimizes $\| W_s - A_s C_s \|_{A_s^{-1} Q_s^{-1}}$ while satisfying the calibration equation $X_s'(W_s - A_s C_s) = (X'C - X_s' A_s C_s)$. The solution, which can be found through methods given for example, in Ben-Israel and Greville (1980), is

$$W_s = A_s C_s + A_s Q_s X_s (X_s' A_s Q_s X_s)^{-1} (X'C - X_s' A_s C_s), \tag{1}$$

where it is assumed that X_s is of full rank. Thus,

$$Y_s' W_s = \hat{Y}' C + (Y - \hat{Y})' A \Delta C \tag{2}$$

where

$$\begin{aligned} \hat{Y} &= X (X_s' A_s Q_s X_s)^{-1} X_s' A_s Q_s Y_s \\ &= X \hat{\beta}_s, \end{aligned} \tag{3}$$

and $\Delta = \operatorname{diag}(\delta_k)$, with δ_k equal to 1 if unit k is in the sample, and 0 otherwise. The estimator obtained in (2) through calibration is the generalized regression estimator described in Särndal (1982).

3. ASYMPTOTIC VARIANCE OF $\hat{\beta}$

Let $\hat{F}_s = X_s' A_s Q_s Y_s$ and $\hat{G}_s = X_s' A_s Q_s X_s$ we have

$$\hat{\beta}_s = \hat{G}_s^{-1} \hat{F}_s. \tag{4}$$

If we approximate $\hat{G}_s^{-1} \hat{F}_s$ using the differential at the point $(F,G) = (X'QY, X'QX)$ we get

$$\begin{aligned} \hat{G}_s^{-1} \hat{F}_s &\doteq G^{-1}F + G^{-1} d\hat{F}_s + d(\hat{G}_s^{-1})F \\ &= G^{-1}F + G^{-1} d\hat{F}_s - G^{-1}(d\hat{G}_s)G^{-1}F \\ &= G^{-1}(F + ((\hat{F}_s - F) - (\hat{G}_s - G)G^{-1}F)) \\ &= G^{-1}F + G^{-1}(\hat{F}_s - \hat{G}_s G^{-1}F) \end{aligned} \tag{5}$$

Thus, if we note $\beta = G^{-1}F$ and $Y^* = X\beta$, we have

$$\begin{aligned} \hat{\beta}_s &\doteq \beta + (X'QX)^{-1}(X_s' A_s Q_s Y_s - X_s' A_s Q_s X_s \beta) \\ &= \beta + (X'QX)^{-1} X_s' A_s Q_s (Y_s - Y_s^*) \\ &= \beta + (X'QX)^{-1} X' Q \operatorname{diag}(Y - Y^*) A \Delta \mathbf{1}_{N\times 1}, \end{aligned} \tag{6}$$

where $\mathbf{1}_{a\times b} \in \mathbb{R}^{a\times b}$ is a matrix of 1's. If we define $T = \operatorname{diag}(Y - Y^*) Q X (X'QX)^{-1}$ and if we denote $\mathrm{V}(\hat{\beta}_s)$ the asymptotic variance of $\hat{\beta}_s$, then from (6)

$$\begin{aligned} \mathrm{V}(\hat{\beta}_s) &= \mathrm{V}(T' A \Delta \mathbf{1}_{N\times 1}) \\ &= T'(\mathrm{V}(A \Delta \mathbf{1}_{N\times 1}))T \\ &= T'(A \Pi A - \mathbf{1}_{N\times N})T, \end{aligned} \tag{7}$$

where $\Pi = (\pi_{kl})$ is the matrix of probabilities that the pair of units (k,l) is in the sample.

4. ESTIMATORS OF THE ASYMPTOTIC VARIANCE OF $\hat{\beta}$

If we suppose that $\pi_{kl} > 0$ for all pairs of units, then the asymptotic variance matrix (7) can be estimated by

$$\hat{\mathrm{V}}_{\mathrm{HT}}(\hat{\beta}_s) = \hat{T}_s'(A_s \mathbf{1}_{n\times n} A_s - \Pi_s^{(-1)}) \hat{T}_s, \tag{8}$$

where $\hat{T}_s = \operatorname{diag}(Y_s - \hat{Y}_s) Q_s X_s (X'QX)^{-1}$ and $\Pi_s^{(-1)} = (\pi_{kl}^{-1})$ for $k, l \in s$. This estimator will be referred to as the Horvitz-Thompson estimator of the asymptotic variance, since apart from the replacement of the unknown β by $\hat{\beta}_s$ we are replacing the double summation over the population in (7) by a double summation of the same terms over the sample, weighted by the inverse of the probabilities of inclusion of the sampled pairs of units.

Using methods given in Binder (1996), we would obtain an estimator $\hat{\mathrm{V}}_{\mathrm{B}}(\hat{\beta}_s)$, of the form given by (8), but with $\hat{T}_s$ replaced by the matrix $\tilde{T}_s = \operatorname{diag}(Y_s - \hat{Y}_s) Q_s X_s (X_s' A_s Q_s X_s)^{-1}$.

Under stratified sampling with a simple random sample s_h, of size n_h, taken within each stratum of size N_h, the asymptotic variance (7) becomes

$$\mathrm{V}(\hat{\beta}_s)=\sum_h \frac{(1-f_h)N_h^2}{n_h(N_h-1)}(\boldsymbol{T}_h'\boldsymbol{T}_h-N_h\bar{\boldsymbol{T}}_h'\bar{\boldsymbol{T}}_h)\ , \qquad (9)$$

where $f_h = n_h/N_h$, $\boldsymbol{T}_h$ is the sub-matrix of $\boldsymbol{T}$ that corresponds to stratum h, and $\bar{\boldsymbol{T}}_h = \mathbf{1}_{1\times N_h}\boldsymbol{T}_h/N_h$.

With this stratified sampling plan, the variance estimator (8) becomes

$$\hat{\mathrm{V}}_{\mathrm{HT}}(\hat{\beta}_s)=\sum_h \frac{(1-f_h)N_h^2}{n_h(n_h-1)}(\hat{\boldsymbol{T}}_{s_h}'\hat{\boldsymbol{T}}_{s_h}-n_h\overline{\hat{\boldsymbol{T}}}_{s_h}'\overline{\hat{\boldsymbol{T}}}_{s_h}), \qquad (10)$$

where $\hat{\boldsymbol{T}}_{s_h}$ is the sub-matrix of $\hat{\boldsymbol{T}}_s$ for stratum h, and $\overline{\hat{\boldsymbol{T}}}_{s_h} = \mathbf{1}_{1\times n_h}\hat{\boldsymbol{T}}_{s_h}/n_h$.

The estimator obtained with the Binder method, $\hat{\mathrm{V}}_{\mathrm{B}}(\hat{\beta}_s)$, is of the same form as that given in (10) except that $\hat{\boldsymbol{T}}_{s_h}$ and $\overline{\hat{\boldsymbol{T}}}_{s_h}$ are replaced with $\tilde{\boldsymbol{T}}_{s_h}$ and $\overline{\tilde{\boldsymbol{T}}}_{s_h}$ respectively.

Let the sample s of size n be selected with Bernoulli sampling, where each unit would have a probability $f = \nu/N$ of being in the sample, with ν being the expected sample size. Then the asymptotic variance (7) is

$$\mathrm{V}(\hat{\beta}_s)=\frac{(1-f)N^2}{\nu}\frac{\boldsymbol{T}'\boldsymbol{T}}{N}. \qquad (11)$$

With Bernoulli sampling, the estimator (8) becomes

$$\hat{\mathrm{V}}_{\mathrm{HT}}(\hat{\beta}_s)=\frac{(1-f)N^2}{\nu}\frac{\hat{\boldsymbol{T}}_s'\hat{\boldsymbol{T}}_s}{\nu}. \qquad (12)$$

For $\boldsymbol{X} = \mathbf{1}_{N\times 1}$ and $\boldsymbol{Q} = \sigma^2\boldsymbol{I}$, the estimator obtained with the Binder method for Bernoulli sampling is

$$\hat{\mathrm{V}}_{\mathrm{B}}(\hat{\beta}_s) = \frac{(1-f)N^2}{n}\frac{\hat{\boldsymbol{T}}_s'\hat{\boldsymbol{T}}_s}{n}, \qquad (13)$$

where n is the effective sample size.

A third estimator of the asymptotic variance (11) comes naturally to mind, and it is

$$\hat{\mathrm{V}}_{\mathrm{CAL}}(\hat{\beta}_s) = \frac{(1-f)N^2}{\nu}\frac{\hat{\boldsymbol{T}}_s'\hat{\boldsymbol{T}}_s}{n}. \qquad (14)$$

This estimator will be preferred to the one given by (12) because of the randomness of the sample size n under Bernoulli sampling. The estimator (13) seems to overcorrect for this randomness. This third estimator is simply a calibrated estimator of the asymptotic variance (11) where the auxiliary information matrix $\boldsymbol{X}_T = \mathbf{1}_{N\times 1}$ was used with $\boldsymbol{Q}_T = \sigma^2\boldsymbol{I}$. More generally, other matrices could be used for $\boldsymbol{X}_T$ and $\boldsymbol{Q}_T$ and the next section will present the results of simulations done with less trivial $\boldsymbol{X}_T$ or $\boldsymbol{Q}_T$ matrices. The matrices $\boldsymbol{X}_T$ and $\boldsymbol{Q}_T$ need not be the same as the matrices $\boldsymbol{X}$ and $\boldsymbol{Q}$ used to define β.

Calibration can also be used in the case of stratified sampling. Equation (9) can be written as the difference of two terms as follows:

$$\mathrm{V}(\hat{\beta}_s)=\sum_h \frac{(1-f_h)N_h^2}{n_h(N_h-1)}\boldsymbol{T}_h'\boldsymbol{T}_h - \sum_h \frac{(1-f_h)N_h^3}{n_h(N_h-1)}\bar{\boldsymbol{T}}_h'\bar{\boldsymbol{T}}_h. \qquad (15)$$

The first term which is a population total, that is a single sum over the population, can be estimated using calibration, while the second term which is a double sum over the population can be estimated with the Horvitz-Thompson method.

5. MONTE CARLO STUDY

A Monte Carlo simulation was done using a population of 284 Swedish municipalities found in Särndal, Swensson and Wretman (1992). The variable of interest is the 1985 population, the auxiliary variables are the 1975 population and a variable which takes the value one for each municipality. The parameter β is the slope in a linear regression with intercept of the 1985 population over the 1975 population. Using Bernoulli sampling, 1000 samples of expected size 200 were drawn, and the bias and mean square error over the 1000 samples of the three variance estimators were compared. Then the same thing was done with samples of expected size 100 and 50, and with simple random samples of size 200, 100 and 50. To compute $\hat{\mathrm{V}}_{\mathrm{CAL}}(\hat{\beta})$, the vector of the 1975 population was used as the auxiliary information matrix $\boldsymbol{X}_T$, along with $\boldsymbol{Q}_T = (\mathrm{diag}(\boldsymbol{X}_T))^{-1}$.

Note that for both Bernoulli sampling and simple random sampling, the asymptotic variance we are estimating is 2.5×10^{-5} for n=200, 1.1×10^{-4} for n=100, and 2.8×10^{-4} for n=50. The bias of the three variance estimators under Bernoulli sampling and simple random sampling are shown in Table 1 and Table 2 respectively. For both sampling plans, with a sample size of 200 the bias of $\hat{\mathrm{V}}_{\mathrm{B}}(\hat{\beta})$ is much higher than that of the other two estimators although it has the smaller bias with a sample size of 50.

Table 1. Bias under Bernoulli sampling of all 284 municipalities

Estimator	Expected sample size		
	n=200	n=100	n=50
$\hat{V}_{HT}(\hat{\beta})$	4.2×10^{-8}	-2.7×10^{-5}	-1.6×10^{-4}
$\hat{V}_{B}(\hat{\beta})$	2.2×10^{-5}	7.8×10^{-5}	6.7×10^{-5}
$\hat{V}_{CAL}(\hat{\beta})$	-1.1×10^{-7}	-3.3×10^{-5}	-1.8×10^{-4}

Table 2. Bias under simple random sampling of all 284 municipalities

Estimator	Sample size		
	n=200	n=100	n=50
$\hat{V}_{HT}(\hat{\beta})$	1.1×10^{-7}	-2.2×10^{-5}	-1.6×10^{-4}
$\hat{V}_{B}(\hat{\beta})$	2.1×10^{-5}	8.8×10^{-5}	8.9×10^{-5}
$\hat{V}_{CAL}(\hat{\beta})$	-4.8×10^{-8}	-2.9×10^{-5}	-1.8×10^{-4}

Tables 3 and 4 show the mean square error of the 3 variance estimates under Bernoulli sampling and simple random sampling respectively. We can see that $\hat{V}_{B}(\hat{\beta})$ has a much higher mean square error than the other two estimators, and this both for Bernoulli sampling and simple random sampling. The difference is measured in orders of magnitude. The estimator $\hat{V}_{CAL}(\hat{\beta})$ is always the one with the smallest mean square error. Its advantage over $\hat{V}_{HT}(\hat{\beta})$ increases slightly as the sample size decreases. That advantage is also slightly higher under Bernoulli sampling than under simple random sampling. It seems that although with simple random sampling, calibration is used to estimate only the first term of (15), the gains due to calibration are almost as good as those under Bernoulli sampling where calibration is used to estimate all of the asymptotic variance.

A similar simulation was done after removing the three largest municipalities from the population. Two calibration variables were used to compute $\hat{V}_{CAL}(\hat{\beta})$: the 1975 population, and a variable which takes the value 1 for every municipality. The matrix Q_T was the identity matrix. The results are given in Tables 5 and 6. Note that the asymptotic variance we are estimating is 2.8×10^{-4} for both Bernoulli sampling and simple random sampling. With this more homogeneous population, the mean square error of the estimators $\hat{V}_{B}(\hat{\beta})$ and $\hat{V}_{CAL}(\hat{\beta})$ is the same, both for Bernoulli sampling and for simple random sampling. Their mean square error is close to half that of $\hat{V}_{HT}(\hat{\beta})$. The estimator $\hat{V}_{CAL}(\hat{\beta})$ though, has the advantage of also performing well with the population of all 284 municipalities, including the three large ones.

Table 3. MSE under Bernoulli sampling of all 284 municipalities

Estimator	Expected sample size		
	n=200	n=100	n=50
$\hat{V}_{HT}(\hat{\beta})$	5.7×10^{-11}	5.7×10^{-9}	6.6×10^{-8}
$\hat{V}_{B}(\hat{\beta})$	3.9×10^{-9}	5.5×10^{-8}	1.3×10^{-7}
$\hat{V}_{CAL}(\hat{\beta})$	5.5×10^{-11}	4.7×10^{-9}	5.4×10^{-8}

Table 4. MSE under simple random sampling of all 284 municipalities

Estimator	Sample size		
	n=200	n=100	n=50
$\hat{V}_{HT}(\hat{\beta})$	5.9×10^{-11}	5.5×10^{-9}	6.4×10^{-8}
$\hat{V}_{B}(\hat{\beta})$	3.9×10^{-9}	6.3×10^{-8}	1.6×10^{-7}
$\hat{V}_{CAL}(\hat{\beta})$	5.7×10^{-11}	4.7×10^{-9}	5.4×10^{-8}

Table 5. Bias when sampling with n=100 from the 281 smallest municipalities

Estimator	Sampling method	
	Bernoulli	Simple random sampling
$\hat{V}_{HT}(\hat{\beta})$	-3.3×10^{-5}	-3.5×10^{-5}
$\hat{V}_{B}(\hat{\beta})$	-5.1×10^{-5}	-4.7×10^{-5}
$\hat{V}_{CAL}(\hat{\beta})$	-6.3×10^{-5}	-6.0×10^{-5}

Table 6. MSE when sampling with n=100 from the 281 smallest municipalities

Estimator	Sampling method	
	Bernoulli	Simple random sampling
$\hat{V}_{HT}(\hat{\beta})$	2.9×10^{-8}	3.1×10^{-8}
$\hat{V}_{B}(\hat{\beta})$	1.6×10^{-8}	1.8×10^{-8}
$\hat{V}_{CAL}(\hat{\beta})$	1.6×10^{-8}	1.8×10^{-8}

6. CONCLUSIONS

To conclude, it should be noted that all three approaches can be used for variance estimation in general, not just for estimating the variance of the regression parameter estimator. When estimating the variance of a total, the Binder method coincides with the *g*-weighted method described in Särndal, Swensson and Wretman (1989).

All three variance estimators can be used with more complex sampling plans than Bernoulli sampling or simple random sampling. All we need to know are the second order probabilities of inclusion which are assumed positive.

As we have seen, very different results can be obtained for different populations, and more comparisons should be made with different populations. Studies could also be made on the optimum choice of calibration variables. Using calibration to estimate variances, however, seems to be a promising method.

REFERENCES

Ben-Israel, A., and Greville, T.N.E. (1980). *Generalized Inverses: Theory and Applications*, Huntington, New York: Robert E. Krieger Publishing Company.

Binder, D. (1996). "Linearization methods for single phase and two-phase samples: a cookbook approach" , *Survey Methodology*, 22, 17-22.

Deville, J.-C., and Särndal, C.-E. (1992). "Calibration estimators in survey sampling", *Journal of the American Statistical Association*, 87, 376-382.

Estevao, V., Hidiroglou, M. A., and Särndal, C.-E. (1995). "Methodological principles for a generalized estimation system at Statistics Canada", *Journal of Official Statistics*, 11, 181-204.

Särndal, C.-E. (1982). "Implications of survey design for generalized regression estimation of linear functions", *Journal of Statistical Planning and Inference*, 7, 155-170.

Särndal, C.-E., Swensson, B., and Wretman, J.H. (1989). "The weighted residual technique for estimating the variance of the general regression estimator of the finite population total", *Biometrika*, 76, 527-537.

Särndal, C.-E., Swensson, B., and Wretman, J.H. (1992). *Model Assisted Survey Sampling*, New York: Springer-Verlag.