Optimal variance estimation for generalized regression predictor

Arijit Chaudhuri[a,*], Debesh Roy[b]

[a] *Computer Science Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India*

[b] *Department of Statistics, Presidency College, Calcutta 700073, India*

Received 25 April 1994

Abstract

The generalized regression (greg) predictor for the finite population total of a real variable is often employed when values of an auxiliary variable are available. Several variance estimators for it do well in large samples though bearing no optimality properties. We find a variance estimator which, under a restrictive model, has an optimality property under 'exact' as well as 'asymptotic' analysis. But this involves model parameters. Under a further restriction on the model, two model-parameter-free variance estimators are derived sharing the same 'asymptotic' optimality. Numerical illustrations through simulation are presented to demonstrate marginal improvements in using them rather than their predecessors. Two of the latter, though not optimal, are simpler, intuitively appealing, compete well in large samples, generally applicable and should be persisted with in practice.

AMS subject classification: 62D05

Keywords: Asymptotic analysis; Confidence interval; Generalized regression predictor; Quadratic estimator; Super-polution modelling; Survey population; Variance estimation

1. Introduction

We consider estimating the survey population total Y of a real variable y when positive values of another variable x well correlated with y are at hand. From the works of Cassel et al. (1976) and Särndal (1980) we know that, if a linear regression, through the origin, of y on x may be modelled as plausible, then what they call a 'generalized regression' (greg) estimator, rather predictor for Y because Y is then a random variable, is popularly employed. With a more general linear model postulation of course appropriate alternative estimators or predictors are available from the above sources and also from those of Särndal (1982) and Särndal et al. (1992). For simplicity, we shall no more refer to these alternatives in this paper. Irrespective of the validity of any model, the greg predictor above is known to have a property of being 'asymptotically design unbiased' (ADU) for Y if one applies Brewer's (1979) asymptotic approach. This property, renders it 'robustness' as is generally recognized in case one uses large samples. In order that one may construct for Y a confidence interval (CI) using the greg predictor denoted by t_g one needs an appropriate estimator, say, v_g of its mean square error (MSE), which we shall refer to as a variance estimator. Särndal (1982), Kott (1990) and Särndal et al. 1992) have given several variance estimators for t_g. They are known to perform well in large samples. But no optimality property is known about them. In this article, after giving certain preliminaries in Section 2 we present a variance estimator in Section 3, studded with a certain optimality property under the linear regression model with a zero intercept but with general model variances σ_i^2 for the regression errors which are assumed to be independent across population members labelled $i = 1, \ldots, N$. But this involves the regression slope parameter β and σ_i^2 which cannot be known in practice. So, instead of the 'exact' optimality above an 'asymptotic' optimality following Brewer's (1979) approach is aimed at. The same variance estimator turns out 'asymptotically' optimal too. With a further restriction on the model that $\sigma_i^2 = \sigma^2 f_i (i = 1, \ldots, N$, with $\sigma > 0$ unknown but f_i known), substituting suitable estimators for β and σ^2 two alternative variance estimators are derived from the above optimal variance estimator. Fortunately, both turn out optimal 'asymptotically' and both are applicable, provided f_i is known as we assume to be the case. This same restriction applies to two variance estimators of Kott (1990) but not to the two given by Särndal (1982), each of which fails to share the optimality property we are looking for. To have a theoretical comparison of relative efficiencies of the variance estimators is not easy. So, we resort to a numerical exercise through simulation study, reported in Section 4, to compare, using several criteria, the efficacies of the confidence intervals for Y using t_g and the above-noted six variance estimators. We find them to be quite competitive when data are generated to fit the model that yields optimality for our variance estimators, which naturally in this tailormade situation have a slight edge over the four others. But Särndal's (1982) variance estimators besides being generally applicable continue as viable competitors even though they lack optimality. Allowing variations in f_i we find, however, our variance estimators to display robustness in this respect. So, if one needs optimality only then it is worthwhile to go for the new estimators which are more complicated but otherwise it is safe to persist with Särndal's easier estimators. A message of interest is that Särndal's estimators do not lag behind the optimal ones which are available in rarer situations.

2. Notation and preliminaries

We consider a survey population $U = (1, \ldots, i, \ldots, N)$. On it are defined two real variables x and y taking values x_i (>0, known) and y_i with population totals X and Y, respectively. The problem is to estimate Y on ascertaining the values of y_i for the units in a sample s of a given size n ($< N$). We require every unit in s to be distinct. The selection probability of s is $p(s)$ and the design p is supposed to be so restricted that the inclusion probabilities π_i of units i and π_{ij} of pairs of units (i, j) are positive. Though extremely restrictive, a popular model M, say, is postulated to relate $\underline{Y} = (y_1, \ldots, y_i, \ldots, y_N)$ with $\underline{X} = (x_1, \ldots, x_i, \ldots, x_N)$ for which one may validly write

$$y_i = \beta x_i + \varepsilon_i, \quad i \in U. \tag{2.1}$$

Here β is an unknown constant, ε_i's are independently distributed random variables with means $E_m(\varepsilon_i) = 0$ and variances $V_m(\varepsilon_i) = \sigma_i^2$, ($\sigma_i > 0$, unknown). If this model is tenable, then the well-known greg predictor for Y is

$$t_g = \sum \frac{y_i}{\pi_i} I_{si} + \hat{\beta}_Q \left(X - \sum \frac{x_i}{\pi_i} I_{si}\right);$$

here $I_{si} = 1$ if $i \in s$, and is 0 otherwise; $\sum$ is sum over i in U; $\hat{\beta}_Q = (\sum Q_i x_i y_i I_{si})/(\sum Q_i x_i^2 I_{si})$; $Q_i (> 0)$ is an assignable constant. One may note that though there is no intercept term in (2.1), t_g does not have a ratio form. If there was a nonzero intercept term in (2.1) an appropriate corresponding greg predictor for Y would be different from t_g. But in this paper we treat only t_g. For subsequent use let us write

$$e_i = y_i - \hat{\beta}_Q x_i, \qquad g_{si} = 1 + \left(X - \sum \frac{x_i}{\pi_i} I_{si}\right) \frac{Q_i \pi_i x_i}{\sum Q_i x_i^2 I_{si}},$$

$$t_g = \hat{\beta}_Q X + \sum \frac{e_i}{\pi_i} I_{si} = \sum \frac{y_i}{\pi_i} g_{si} I_{si}.$$

For the model M we further assume that

$$\delta_i = E_m(y_i^4) < +\infty.$$

The particular case of M will be denoted by $M(f)$ when $\sigma_i^2 = \sigma^2 f_i$, $i \in U$; here $\sigma (> 0)$ is unknown but f_i is known. By E_p, V_p, we shall denote expectation, variance with respect to p. We shall write $\sum\sum$ for sum over $i, j \, (i \neq j)$ in U and $A_{ij} = (\pi_i \pi_j - \pi_{ij})/2\pi_{ij}$. Further let

$$B_Q = (\sum Q_i x_i y_i \pi_i)/(\sum Q_i x_i^2 \pi_i), \qquad E_i = y_i - B_Q x_i, \quad i \in U$$

Särndal (1982) approximates $V_p(t_g)$ by

$$V = \sum\sum \left(\frac{E_i}{\pi_i} - \frac{E_j}{\pi_j}\right)^2 A_{ij} \pi_{ij} \tag{2.2}$$

and gives two estimators for it, to be called variance estimators for t_g, as, on writing $I_{sij} = I_{si}I_{sj}$,

$$v_1 = \sum\sum\left(\frac{e_i}{\pi_i} - \frac{e_j}{\pi_j}\right)^2 \Delta_{ij} I_{sij} \quad \text{and} \quad v_2 = \sum\sum\left(\frac{g_{si}e_i}{\pi_i} - \frac{g_{sj}e_j}{\pi_j}\right)^2 \Delta_{ij} I_{sij}.$$

At this stage let us observe a crucial fact that though the model M motivates the predictor t_g the latter has the following model-free property as an important merit. Brewer (1979) introduced an interesting asympotic approach which assumes hypothetical reappearance of U, say, T (> 1) times. On each appearance, a sample of n distinct units as before is drawn from it adopting the same design p. All samples so drawn independently are pooled together and the statistics of the form t_g are calculated from the pooled sample. Allowing T to tend to infinity a limit of the design-based expectation of t_g may be taken. This limiting value may be denoted by $\lim E_p(t_g)$. Similarly one may define $\lim E_p(g_{si})$. It is well-known that

$$\lim E_p(g_{si}) = 1 \tag{2.3}$$

so that for large samples g_{si} is close to unity and that

$$\lim E_p(t_g) = Y. \tag{2.4}$$

Eq. (2.4) implies that t_g is 'asymptotically design unbiased' for Y. For t_g to the ADU no model is required. The variance estimators v_1 and v_2 also do not need any model. But Kott (1990) recommended two variance estimators for t_g as

$$k_j = \frac{v_j}{E_m(v_j)} E_m(t_g - Y)^2, \quad j = 1, 2,$$

which are practicable when the model $M(f)$ is valid. So to compare v_j with k_j, ($j = 1, 2$), we require $M(f)$. If we postulate $M(f)$, then it is of interest to investigate availability of an appropriate optimal variance estimator under $M(f)$. This investigation is undertaken in Section 3.

3. Optimum variance estimation

For simplicity let us write V of (2.2) in the form

$$V = \sum \alpha_i y_i^2 + \sum\sum \alpha_{ij} y_i y_j,$$

with

$$\alpha_i = \left(\frac{1}{\pi_i} - 1\right) + (Q_i^2 x_i^2 \pi_i^2) V_p\left(\sum \frac{x_i}{\pi_i} I_{si}\right) \Big/ \left(\sum Q_i x_i^2 \pi_i\right)^2$$

$$- 2Q_i x_i \pi_i \left[\sum_{k=1}^{N} x_k\left(\frac{\pi_{ik}}{\pi_i \pi_k} - 1\right)\right] \Big/ \left(\sum Q_i x_i^2 \pi_i\right),$$

$$\alpha_{ij} = \left(\frac{\pi_{ij}}{\pi_i\pi_j} - 1\right) + Q_iQ_j\pi_i\pi_j x_i x_j V_p\left(\sum \frac{x_i}{\pi_i} I_{si}\right) \Big/ \left(\sum Q_i x_i^2 \pi_i\right)^2$$

$$- Q_i x_j \pi_j \sum_{k=1}^{N} \left(\frac{\pi_{ik}}{\pi_i\pi_k} - 1\right) x_k \Big/ \left(\sum Q_i x_i^2 \pi_i\right)$$

$$- Q_j x_i \pi_i \sum_{k=1}^{N} \left(\frac{\pi_{jk}}{\pi_j\pi_k} - 1\right) x_k \Big/ \sum Q_i x_i^2 \pi_i.$$

Let us restrict to a class of nonhomogeneous quadratic estimators of V of the form

$$v = v(s, \underline{Y}) = a_s + \sum b_{si} y_i^2 I_{si} + \sum\sum b_{sij} y_i y_j I_{sij}. \tag{3.1}$$

Let the constants a_s, b_{si}, b_{sij}, free of $\underline{Y}$, be subject to

$$E_p(a_s) = 0, \quad E_p(b_{si} I_{si}) = \alpha_i \quad \text{for } i \in U \quad \text{and} \quad E_p(b_{sij} I_{sij}) = \alpha_{ij} \quad \text{for } i, j\ (i \neq j) \text{ in } U \tag{3.2}$$

It follows that $E_p(v) = V$. An optimal estimator of V within the class of estimators (3.1) satisfying (3.2) is derived in the following theorem.

Theorem 1. *Under $\underline{M}$, for all estimators v of the form* (3.1) *satisfying* (3.2),

$$M(v) = E_m E_p(v - V)^2 \geq \sum \alpha_i^2 \left(\frac{1}{\pi_i} - 1\right)\eta_i^2 + \sum\sum \alpha_{ij}^2 \left(\frac{1}{\pi_{ij}} - 1\right)\eta_{ij},$$

where $\eta_i^2 = \delta_i - (\sigma_i^2 + \mu_i^2)^2$ and $\eta_{ij} = (\sigma_i^2 + \mu_i^2)(\sigma_j^2 + \mu_j^2) - \mu_i^2\mu_j^2$, and $\mu_i = \beta x_i$, $i \in U$.

Equality is attained in the above if v equals

$$v_0 = \sum \alpha_i (y_i^2 - \sigma_i^2 - \mu_i^2) \frac{I_{si}}{\pi_i} + \sum\sum \alpha_{ij} (y_i y_j - \mu_i\mu_j) \frac{I_{sij}}{\pi_{ij}} + \sum \alpha_i (\sigma_i^2 + \mu_i^2)$$

$$+ \sum\sum \alpha_{ij} \mu_i \mu_j.$$

Proof (*sketch*). Following Godambe and Thompson (1977), we get, for the anticipated variance of v and $\underline{M}$,

$$M(v) = E_m E_p(v - V)^2 = E_p E_m(v) + E_p\{E_m(v - V)\}^2 - V_m(V)$$

$$= E_p E_m(v) + E_p\Bigg[a_s + \sum (\sigma_i^2 + \mu_i^2) b_{si} I_{si} + \sum\sum \mu_i\mu_j b_{sij} I_{sij}$$

$$- \sum \alpha_i(\sigma_i^2 + \mu_i^2) - \sum\sum \alpha_{ij}\mu_i\mu_j \Bigg]^2 - V_m(V).$$

To get rid of the second term in $M(e)$ we take

$$a_s = \sum (\alpha_i - b_{si} I_{si})(\sigma_i^2 + \mu_i^2) + \sum\sum (\alpha_{ij} - b_{sij} I_{sij})\mu_i\mu_j.$$

Then writing $b_{si}^* = \alpha_i/\pi_i$, $b_{sij}^* = \alpha_{ij}/\pi_{ij}$, denoting by $\sum_s$ the sum over all samples and using (3.2), it is possible to show, following Cassel et al. (1976) that

$$M(e) \geqslant \sum_s \sum (b_{si} - b_{si}^*)^2 \eta_i^2 I_{si} p(s) + \sum_s \sum \sum (b_{sij} - b_{sij}^*)^2 \eta_{ij} I_{sij} p(s)$$

$$+ \sum \alpha_i^2 \left(\frac{1}{\pi_i} - 1\right)\eta_i^2 + \sum\sum \alpha_{ij}^2 \left(\frac{1}{\pi_{ij}} - 1\right)\eta_{ij}$$

$$\geqslant \sum \alpha_i^2 \left(\frac{1}{\pi_i} - 1\right)\eta_i^2 + \sum\sum \alpha_{ij}^2 \left(\frac{1}{\pi_{ij}} - 1\right)\eta_{ij} = M_0, \text{ say.}$$

This reduces to equality when $b_{si} = b_{si}^*$ and $b_{sij} = b_{sij}^*$.

Thus, $M(e) \geqslant M_0 = M(e_0)$.

But this e_0 involves unknown β and σ_i^2 and hence is not usable. So, this 'exact' optimality is not fruitful. So, we adopt Brewer's (1979) asymptotic approach as an alternative. Let us restrict to a class of estimators or rather predictors of V, say, $w = w(s, \underline{Y})$, which are 'asymptotically design unbiased' for V, i.e. let

$$\lim E_p(w) = V. \tag{3.3}$$

Among all such w, one, say, $e^* = e^*(s, \underline{Y})$ will be regarded as 'asymptotically optimum' if

$$E_m \lim E_p(w - V)^2 \geqslant E_m \lim E_p(e^* - V)^2. \tag{3.4}$$

Asymptotic optimality of e_0 among ADU estimators of V is established in Theorem 2 below.

Theorem 2. *Under $\underline{M}$, for w subject to (3.3), we have*

$$E_m \lim E_p(w - V)^2 \geqslant M_0 = E_m \lim E_p(e_0 - V)^2.$$

Proof (*sketch*). Let $h = h(s, \underline{Y})$ satisfy $\lim E_p(h) = 0$; further let

$$\bar{e} = \sum \alpha_i y_i^2 \frac{I_{si}}{\pi_i} - \sum\sum \alpha_{ij} y_i y_j \frac{I_{sij}}{\pi_{ij}} \quad \text{and} \quad \Delta = E_m(w - V).$$

Following and extending Godambe–Joshi (1965) and Godambe–Thompson (1977) type analysis we have

$$E_m \lim E_p(w - V)^2 \geqslant \lim E_p V_m(\bar{e}) + \lim E_p V_m(h) + \lim E_p \Delta_m^2 - V_m(V) \geqslant M_0. \tag{3.5}$$

Equality holds on choosing

$$h = -\left[\sum \alpha_i(\sigma_i^2 + \mu_i^2)\left(\frac{I_{si}}{\pi_i} - 1\right) - \sum\sum \alpha_{ij}\mu_i\mu_j\left(\frac{I_{sij}}{\pi_{ij}} - 1\right)\right] = h_0, \text{ say.}$$

It follows that $v_0 = \bar{v} + h_0$ satisfies $\lim E_p(v_0 - V) = 0$, $E_m(v_0 - V) = 0$ and the lower bound in (3.5) is attained for $v = v_0$. Thus v_0 is the desired 'asymptotically optimum' estimator for V and the Theorem 2 follows.

As v_0 is not usable we restrict $\underline{M}$ *to* $\underline{M}(f)$. Under validity of $\underline{M}(f)$, we note that v_0 reduces to

$$v_{0f} = \sum \alpha_i(y_i^2 - \beta^2 x_i^2 - \sigma^2 f_i)(I_{si}/\pi_i) + \sigma^2 \sum \alpha_i f_i + \beta^2 \sum \alpha_i x_i^2$$
$$+ \sum\sum \alpha_{ij}(y_i y_j - \beta^2 x_i x_j)I_{sij}/\pi_{ij} + \beta^2 \sum\sum \alpha_{ij} x_i x_j.$$

Let $\theta = \beta^2$ and $\phi = \sigma^2$ be, respectively, estimated by estimators $\hat{\theta}$ *and* $\hat{\phi}$ *which are model unbiased satisfying* $E_m(\hat{\theta}) = \theta$ and $E_m(\hat{\phi}) = \phi$, given by

$$\hat{\theta} = \left(\frac{\sum x_i y_i (I_{si}/f_i)}{\sum x_i^2 (I_{si}/f_i)}\right)^2 - \frac{1}{(n-1)}\left[\left(\frac{\sum y_i^2 (I_{si}/f_i)}{\sum x_i^2 (I_{si}/f_i)}\right) - \left(\frac{\sum x_i y_i (I_{si}/f_i)}{\sum x_i^2 (I_{si}/f_i)}\right)^2\right],$$

$$\hat{\phi} = \frac{1}{(n-1)}\left[\sum y_i^2 \frac{I_{si}}{f_i} - \left(\sum x_i y_i \frac{I_{si}}{f_i}\right)^2 \Big/ \left(\sum x_i^2 \frac{I_{si}}{f_i}\right)\right].$$

We propose now the two following practicable estimators v_{01} and v_{02} for V given below.

$$v_{01} = \sum \alpha_i\left(y_i^2 - \hat{\theta} x_i^2 - \hat{\phi} f_i\right)\frac{I_{si}}{\pi_i} + \hat{\phi}\sum \alpha_i f_i + \hat{\theta}\sum \alpha_i x_i^2$$
$$+ \sum\sum \alpha_{ij}\left(y_i y_j - \hat{\theta} x_i x_j\right)\frac{I_{sij}}{\pi_{ij}} + \hat{\theta}\sum\sum \alpha_{ij} x_i x_j,$$

$$v_{02} = \left(\sum \alpha_i x_i^2\right)\left[\left(\frac{\sum \alpha_i y_i^2 (I_{si}/\pi_i)}{\sum \alpha_i x_i (I_{si}/\pi_i)}\right) - \hat{\phi}\left(\frac{\sum \alpha_i f_i (I_{si}/\pi_i)}{\sum \alpha_i x_i^2 (I_{si}/\pi_i)}\right)\right]$$
$$+ \left(\sum\sum \alpha_{ij} x_i x_j\right)\left(\frac{\sum\sum \alpha_{ij} y_i y_j I_{sij}/\pi_{ij}}{\sum\sum \alpha_{ij} x_i x_j I_{sij}/\pi_{ij}}\right) + \hat{\phi}\sum \alpha_i f_i.$$

It can be seen from the following theorem that both v_{01} and v_{02} are ADU for V and share with v_0 the same 'asymptotic optimality' (AO).

Theorem 3. *Under* $\underline{M}(f)$,

(i) $\lim E_p(v_{0j}) = V$ *and*
(ii) $E_m \lim E_p(v_{0j} - V)^2 = M_0$, $j = 1, 2$.

Proof. Adopting Brewer's (1979) asymptotic approach and applying Slutzky's well-known limit theorem with various sequences as required, (i) follows promptly. To prove (ii), we just note that (i) combined with $E_m(v_{oj} - V) = 0$ for $j = 1, 2$, implies that

$$E_m \lim E_p(v_{oj} - V)^2 = \lim E_p V_m(v_{oj}) - V_m(V) = M_O, \quad j = 1, 2.$$

Remark 1. An estimator for V can be 'AO' only if it is 'both (a) ADU and (b) model unbiased simultaneously'. Hence one may check that v_j, k_j ($j = 1, 2$) cannot be 'AO' for V.

In order to check how v_{oj} may fare vis-a-vis v_j and k_j ($j = 1, 2$) in practice when values of n and N should be moderately large we consider numerical exercises to examine how they may fare respectively in combination with t_g in producing appropriate confidence intervals (CI) for Y. For this we generate $\underline{Y}, \underline{X}$ subject to $\underline{M}$ so that use of t_g may be appropriate and in fact take the special case $\underline{M}(f)$ only for applicability of k_j and v_{oj} ($j = 1, 2$). Since f_i cannot be known in practice, to study robustness of the procedures we generate $\underline{Y}, \underline{X}$ with a given f_i but apply alternative choices of f_i in the formulae of k_j, v_{oj} ($j = 1, 2$) and see variation in performances of corresponding CI's. As the main emphasis is not on the model but on the design-based aspect of sampling through its hypothetically repeated realizability we take several replicates of the samples drawn by the same procedure and calculate various performance criteria for the CI's. To construct a CI for Y with t_g as a point estimator and v_g as a variance estimator for it as usual we regard $(t_g - Y)/\sqrt{v_g}$ as a variable distributed for large n approximately as a standardized normal deviate τ. Then for a given γ in (0, 1), writing τ_γ for the 100γ% point on the right tail area of the normal distribution, $t_g \pm \tau_{\alpha/2}\sqrt{v_g}$ gives the $100(1 - \alpha)$% CI for Y, with α in (0, 1). In Section 4 we describe how we implement the simulation to evaluate the performances of the resulting CI's.

4. Simulation study

We illustrate three sets of generation of $\underline{Y}, \underline{X}$ subject to the model $\underline{M}(f)$ allowing variation in (N, n) and other respects discussed below.

First we take (1) $N = 160$, $n = 30$; to keep x_i positive we take $x_i = 10 + \theta_i$ generating θ_i as random samples from the negative exponential density $f(\zeta) = \lambda \exp(-\lambda\zeta)$, $\lambda > 0$, $\zeta > 0$, taking $\lambda = 0.6$. Next, following the very well known convention, justified empirically, we take $f_i = x_i^{c_0}$; for illustration we choose $c_0 = 1.1$ only; drawing τ_i ($i = 1, \ldots, N$) as random samples from N(0, 1), taking $\varepsilon_i = \sigma\tau_i\sqrt{f_i}$ and choosing (β, σ) differently we generate $\underline{Y}$'s. To draw samples from U we apply Midzuno's (1952) scheme of selection drawing first a unit from U with a probability proportional to size measures ξ_i ($i = 1, \ldots, N$) taken as the highest integer not exceeding $8.2 + 0.65x_i^h$ with $h = 0.78$; the remaining $(n - 1)$ units are chosen at random without replacement from the $(N-1)$ units left in U after the first draw. We take the number of replicates R of the samples as 1000. By $\sum_r$ we denote sum over the replicates. Using the generated Y, $\underline{X}$, $Z = (\xi_1, \ldots, \xi_i, \ldots, \xi_N)$, and drawing the sample each time we construct the 95% CI for Y given by $t_g \pm 1.96\sqrt{v_g}$ taking $\alpha = 0.05$ and v_g as v_j, k_j and v_{aj} $(j = 1, 2)$.

Secondly, to investigate robustness we take f_i as x_i^c in the formulae for k_j and v_{aj} $(j = 1, 2)$ with c permitted to both agree with and differ from c_0. We keep everything else in tact essentially as in (1) but as this is only for another illustration of relative efficacies of the above CI's we take (2) $N = 60$, $n = 20$; $\lambda = 0.4$, $\beta = 1$, $\sigma = 1.8$, $c_0 = 1.5$ and $R = 100$.

Thirdly, to avoid skewness in the distribution of $\underline{X}$, we draw x_i from the uniform distribution $U(10, 20)$ but take (3) $N = 68$, $n = 16$; $c_0 = 1.5$, $\beta = 3$, $\sigma = 1$ and keeping everything else in tact generate $\underline{Y}$, $\underline{Z}$ but take $R = 100$ only. To investigate robustness of the CI's we take f_i as x_i^c with c same as or different from c_0 as in (2). The findings for these situations (1)–(3) are presented, respectively, in Tables 1–3 below.

In order to compare the CI's based on different choices of v_g we consider the following three criteria:

(I) ACP (actual coverage percentage) = The percentage of replicates for which the CI covers Y – the closer it is to 95, other things remaining in tact, the better.

(II) ARB (absolute pseudo-relative bias) = $(1/R)\sum_r |v_g - V|/V$.

(III) ACV (average coefficient of variation) = The average over the replicates of the values of $\sqrt{v_g}/t_g$. This reflects the length of CI relative to t_g. The lesser the values of II and III the better the CI.

To check the departure of the distribution of $d = (t_g - Y)/\sqrt{v_g}$ from N(0, 1) we further consider the following measures of skewness and kurtosis, respectively, as:

(IV) SK = root beta one = $(1/R)\sum_r [(d - \bar{d})/s_d]^3$; $\bar{d} = (1/R)\sum_r d$; $s_d^2 = (1/R)\sum_r (d - \bar{d})^2$.

(V) KU = beta two = $(1/R)\sum_r [(d - \bar{d})/s_d]^4$.

The closer SK and (KU − 3) to zero the less the departure from N(0, 1). In t_g we tried three choices of Q_i as $(1/\pi_i x_i)$, $(1-\pi_i)/(\pi_i x_i)$ and $(1/x_i^c)$, $i \in U$. But as the values of criteria I–V in case (1) differed negligibly for these choices, we present the results only for $1/\pi_i x_i$ in Table 1. For the cases (2) and (3) we tried Q_i only as $1/\pi_i x_i$, $i \in U$.

In Table 1 we present values of (ACP, 10^4 ACV and 10^3 ARB) for respective v_g presented column-wise in four rows marked (i)–(iv) for four respective choices of (β, σ) as (i) (3, 0.6), (ii) (2, 0.8), (iii) (3, 1.5) and (iv) (2, 1.5). As SK equals 0.02 and KU turns out 3.1 for every parametric combination, the values of (IV) and (V) are not shown in Table 1.

In Tables 2 and 3 ACP, 10^4 ACV, 10^3 ARB, SK and KU-3 are presented as sets of five values each along 6 rows for 6 different choices of v_g and along 5 columns for 5 different choices of c as 1.1, 1.3, 1.5, 1.7 and 1.9.

We repeated the same exercise with $N = 80$ and $n = 15$ keeping everything else in tact but the results turn out similar to the above and so are not shown.

Table 1
Performances of confidence intervals in terms of values of (ACP, 10^4 ACV, 10^3 ARB) for 6 choices of v_g

(β, σ)	v_1	v_2	k_1	k_2	v_{01}	v_{02}
(i)	(94.7, 191, 151)	(94.6, 193, 151)	(94.5, 194, 151)	(94.5, 194, 151)	(94.5, 189, 151)	(94.7, 190, 131)
(ii)	(94.7, 191, 227)	(94.6, 193, 227)	(94.5, 194, 227)	(94.5, 194, 227)	(94.6, 190, 226)	(94.7, 190, 226)
(iii)	(94.7, 191, 284)	(94.7, 193, 284)	(94.5, 194, 284)	(94.7, 194, 284)	(90.5, 190, 283)	(94.6, 191, 283)
(iv)	(94.7, 191, 427)	(94.6, 193, 427)	(94.5, 194, 427)	(94.5, 194, 427)	(94.6, 190, 426)	(94.6, 191, 426)

Table 2
Performances of confidence intervals in terms of values of (ACP, 10^4 ACV, 10^2 ARB, SK and KU − 3)

e_k	c				
	1.1	1.3	1.5	1.7	1.9
e_1	(95, 387, 205, 0.13, 0.27)	(95, 387, 205, 0.13, 0.27)	(95, 387, 205, 0.13, 0.27)	(95, 387, 205, 0.13, 0.27)	(95, 387, 205, 0.13, 0.27)
e_2	(95, 389, 207, 0.09, 0.27)	(95, 389, 207, 0.09, 0.27)	(95, 389, 207, 0.09, 0.27)	(95, 389, 207, 0.09, 0.27)	(95, 389, 207, 0.09, 0.27)
k_1	(95, 390, 212, 0.08, 0.27)	(95, 391, 214, 0.08, 0.27)	(95, 391, 216, 0.08, 0.27)	(95, 391, 217, 0.08, 0.28)	(95, 392, 220, 0.07, 0.28)
k_2	(95, 390, 212, 0.08, 0.27)	(95, 391, 214, 0.08, 0.27)	(95, 391, 216, 0.08, 0.27)	(95, 391, 217, 0.08, 0.28)	(95, 392, 220, 0.07, 0.28)
e_{01}	(95, 381, 197, 0.14, 0.35)	(95, 382, 197, 0.13, 0.36)	(95, 382, 196, 0.12, 0.38)	(95, 382, 195, 0.12, 0.39)	(94, 383, 195, 0.13, 0.43)
e_{02}	(95, 382, 198, 0.14, 0.33)	(95, 382, 198, 0.13, 0.33)	(95, 383, 199, 0.13, 0.33)	(95, 383, 201, 0.13, 0.34)	(94, 383, 203, 0.17, 0.35)

Table 3
Performances of confidence intervals in terms of values of (ACP, 10^4 ACV, 10^3 ARB, SK, KU − 3)

v_k	c				
	1.1	1.3	1.5	1.7	1.9
v_1	(95, 167, 262, 0.00, 0.87)	(95, 167, 262, 0.00, 0.87)	(95, 167, 262, 0.00, 0.87)	(95, 167, 262, 0.00, 0.87)	(95, 167, 262, 0.00, 0.87)
v_2	(95, 165, 262, 0.00, 0.85)	(95, 165, 262, 0.00, 0.85)	(95, 165, 262, 0.00, 0.85)	(95, 165, 262, 0.00, 0.85)	(95, 168, 262, 0.00, 0.85)
k_1	(97, 165, 264, 0.00, 0.85)	(97, 165, 264, 0.00, 0.85)	(97, 165, 264, 0.01, 0.85)	(97, 169, 264, 0.01, 0.85)	(97, 169, 264, 0.01, 0.85)
k_2	(97, 165, 264, 0.00, 0.85)	(97, 165, 264, 0.00, 0.85)	(97, 165, 264, 0.01, 0.85)	(97, 169, 265, 0.01, 0.85)	(97, 169, 265, 0.01, 0.85)
v_{01}	(94, 160, 247, 0.00, 0.86)	(94, 161, 246, 0.00, 0.86)	(94, 161, 247, 0.00, 0.90)	(95, 161, 250, 0.00, 0.91)	(96, 161, 253, 0.00, 0.93)
v_{02}	(94, 160, 254, 0.00, 0.81)	(94, 160, 255, 0.00, 0.80)	(94, 160, 256, 0.01, 0.81)	(97, 169, 265, 0.01, 0.85)	(94, 161, 259, 0.02, 0.80)

5. Concluding remarks

There is very little to choose among the 6 alternative variance estimators illustrated in their efficacies in yielding confidence intervals in combination with the generalized regression predictor. Each one looks effective in keeping close to the aimed at nominal confidence coefficient of 95%. Variation in the distribution of X does not disturb the relative efficacies for the variance estimators. Deviations in f_i of the type illustrated do not perceptibly jeopardize the relative performances of the alternative procedures. The discrepancies of SK and (KU − 3) from the ideal value of zero do not seem to be significant to raise serious doubts about normality assumption for the pivotal $(t_g - Y)/\sqrt{v_g}$. The crucial messages, in our view, are that even in tailor-made situations Kott's and our asymptotically optimal variance estimators may not outperform Särndal's simpler ones which do not require restrictive model postulations though all of them are close competitors. So, though they lack optimality Särndal's estimators which are motivated rather by intuition and tradition should be put to use in common practice. However, though our optimal estimators are more complicated and restrictive in nature they remain as viable competitors in favourable circumstances in this age of computers. We have illustrated only a very special situation with a single regressor and a linear regression model through the origin, permitting independent error terms with a simple variance structure. What may transpire in more realistic circumstances is of course worth investigation but our present study is limited to this simple situation which we believe to have some interest.

Acknowledgements

The authors are grateful to the referees for their comments which helped improving on an earlier draft.

References

Brewer, K.R.W. (1979). A class of robust sampling designs for large-scale surveys. *J. Amer. Statist. Assoc.* **74**, 911–915.

Cassel, C.M., C.E. Särndal and J.H. Wretman (1976). Some results on generalized difference estimator and generalized regression estimator for finite populations. *Biometrika* **63**, 615–620.

Godambe, V.P. and V.M. Joshi (1965). Admissibility and Bayes estimation in sampling finite populations. *Ann. Math. Statist.* **36**, 1707–1722.

Godambe, V.P. and M.E. Thompson (1977). Robust near optimal estimation in survey practice. *Bull. Internat. Statist. Inst.* **47** (3), 129–146.

Kott, P.S. (1990). Estimating the conditional variances of a design consistent regression estimator. *J. Statist. Plann. Inference* **24** (3), 287–296.

Midzuno, H. (1952). On the sampling system with probabilities proportional to sum of sizes. *Ann. Inst. Statist. Math.* **3**, 99–107.

Särndal, C.E. (1980). On π-inverse weighting versus best linear weighting in probability sampling. *Biometrika* **67**, 639–650.

Särndal, C.E. (1982). Implications of survey designs for generalized regression estimators of linear functions. *J. Statist. Plann. Inference* **7**, 155–170.

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