PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 4, Fasc. 2 (1984), p. 197-207

ROBUST ESTIMATION AND FINITE POPULATION

BY

JOSEMAR RODRIGUES (BERKELEY, CALIFORNIA)

Abstract. The main problem in this paper* is to examine the robust estimator of a population total in the context of Royall and Herson [3] under multiple regression superpopulation models. The condition on the sample that protects the estimator against bias is studied for polynomial regression models.

1. Introduction. In this paper we are interested in estimating the population total

$$t = \sum_{k=1}^{N} y_k$$

under the superpopulation model in which y_k $(k = 1, \ldots, N)$ are values of a random variable Y_k such that $Y_k = \beta_0 + e_k \sqrt{f(x_k)}$, $k = 1, \ldots, N$, where e_k are independent random variables with mean zero and variance σ^2. The parameter β_0 is unknown and $f(x)$ is a known function of x. The population is finite with units labelled $1, \ldots, N$. For each element of the population we observe the pairs (x_k, y_k), $k = 1, \ldots, N$. If we adopt the above model, we will obtain a condition such that the linear unbiased estimator under this model turns out to be linear unbiased under the multiple regression models. We use the notation $\xi(\delta_0, \ldots, \delta_J: f^0)$, introduced by Royall and Herson [3], to denote the multiple regression model

$$Y_k = \sum_{j=0}^{J} \delta_j \beta_j x_{kj} + e_k \sqrt{f_k^0},$$

* This paper was prepared with support of the University of São Paulo and Fundação de Amparo à Pesquisa do Estado de São Paulo.

where $x_{k0} = 1$, $x_{k1}, \ldots, x_{kJ}$ are known number for $k = 1, \ldots, N$, δ_j's are zeros or ones, $f_k^0 = f(x_k)$ if $j = 0$ or $f_k^0 = f^0(\delta_0, \delta_1 x_{k1}, \ldots, \delta_J x_{kJ})$ if $j \geqslant 1$ and $f(x)$ is a known function. If $\delta_j = 1$, the term $\beta_j x_{kj}$ appears in the multiple regression model, and if $\delta_j = 0$, then this term is absent in the model. The random variables $e_1, \ldots, e_N$ are uncorrelated with mean zero and variance σ^2. The main contribution of this paper is to extend Royall and Herson's results by using a general variance function $f(x)$ in our superpopulation model.

2. Best linear unbiased estimator. Royall and Herson [3] introduced the following definition:

Definition 2.1. For a given sample s and a model ξ, and estimator $\hat{T}$ is *unbiased* for $T = \sum_{k=1}^{N} Y_k$ if

$$\mathrm{E}_\xi[\hat{T} - T] = 0,$$

where the subscript indicates that the expectation is taken with respect to probability distribution of the model ξ.

By the generalized Gauss-Markov theorem ([2], p. 230) and Section 3.1 in [3], the best linear unbiased estimator (B.L.U.E.) of T under the model $\xi(1:f(x))$ is

$$\hat{T}(1:f(x)) = \sum_{k \in s} Y_k + \frac{\sum_{k \in s} Y_k (N-n)/f(x_k)}{\sum_{k \in s} 1/f(x_k)},$$

where $\sum_{k \in s}$ denotes the sum over all units in the sample s.

Remarks. (1) If $f(x) = 1$, then we have the estimator

$$\hat{T}(1:1) = \frac{N}{n} \sum_{k \in s} Y_k,$$

which is the expansion estimator when the simple random sampling is used (see [3]).

(2) The estimator $\hat{T}(1:1)$ is biased under the model $\xi(0, 1:f^0(x))$ for any function $f^0(x)$ unless with $\beta_1 = 0$ or $\bar{x}_{1s} = \bar{x}_1$. For

$$\begin{aligned} \mathrm{E}_\xi[\hat{T}(1:1) - T] &= \frac{N}{n} \sum_{k \in s} \beta_1 x_{k1} - \sum_{k=1}^{N} \beta_1 x_{k1} \\ &= N\beta_1 \bar{x}_{1s} - \beta_1 N \bar{x}_1 = N\beta_1 (\bar{x}_{1s} - \bar{x}_1), \end{aligned}$$

where

$$\bar{x}_{1s} = \frac{\sum_{k \in s} x_{k1}}{n} \quad \text{and} \quad \bar{x}_1 = \frac{\sum_{k=1}^{N} x_{k1}}{N}.$$

3. Robustness for multiple regression models.

Definition 3.1. Let $s^*(J)$ be any sample such that

$$\frac{\sum_{k\in\bar{s}} x_{kj}}{N-n} = \frac{\sum_{k\in s} x_{kj}/f(x_k)}{\sum_{k\in s} 1/f(x_k)}, \qquad j = 1,\ldots,J, \tag{3.1}$$

where $\bar{s} = \{1,\ldots,N\}-s$.

Remark. If $x_{kj} = x_k^j$ and $f(x) = 1$, it turns out to be the condition of a balanced sample introduced by Royall and Herson [3]. Suppose the model $\xi(1{:}f(x))$ is wrong and the correct model is $\xi(\delta_0, \delta_1, \ldots, \delta_J{:}f^0)$. Then we have the following

LEMMA 3.1. *If $s = s^*(J)$, then $\hat{T}(1{:}f(x))$ is unbiased under the multiple regression model $(\delta_0, \delta_1, \ldots, \delta_J{:}f^0)$ for any function f^0.*

Proof. We have

$$\begin{aligned}
&\mathrm{E}_\xi[\hat{T}(1{:}f(x))-T] \\
&= \sum_{k\in s}\Big(\sum_{j=0}^{J} \delta_j\beta_j x_{kj}\Big) + (N-n)\frac{\sum_{k\in s}\sum_{j=0}^{J}\delta_j\beta_j x_{kj}/f(x_k)}{\sum_{k\in s} 1/f(x_k)} - \sum_{k=1}^{N}\Big(\sum_{j=0}^{J}\delta_j\beta_j x_{kj}\Big) \\
&= \sum_{j=0}^{J}\delta_j\beta_j\left(\sum_{k\in s} x_{kj} + \frac{(N-n)\sum_{k\in s} x_{kj}/f(x_k)}{\sum_{k\in s} 1/f(x_k)} - \sum_{k=1}^{N} x_{kj}\right) \\
&= \sum_{j=0}^{J}\delta_j\beta_j\left(\frac{(N-n)\sum_{k\in s} x_{kj}/f(x_k)}{\sum_{k\in s} 1/f(x_k)} - \sum_{k\in\bar{s}} x_{kj}\right) = 0 \qquad \text{if } s = s^*(J).
\end{aligned}$$

Remark. If we choose a sample $s = s^*(J)$, then the estimator $\hat{T}(1{:}f(x))$ is robust in the sense that the bias is eliminated under any multiple regression models. The following theorem states the estimator $\hat{T}(1{:}f(x))$ is B.L.U.E. in a special class of models. The technique which is used to prove the theorem below is the same as that introduced by Scott et al. [5].

THEOREM 3.1. *The estimator $\hat{T}(1{:}f(x))$ is B.L.U.E. under the model*

$$\xi(\delta_0, \delta_1, \ldots, \delta_J{:}\, f^*(x)),$$

where

$$f^*(x) = f(x)\sum_{j=0}^{J} a_j\delta_j x_{kj}, \qquad k = 1,\ldots,N,$$

$$\boldsymbol{x} = (x, \delta_0, \delta_1 x_{k1}, \ldots, \delta_J x_{kJ}), \qquad a_j > 0,\ j = 0, 1, \ldots, J,$$

if $s = s^*(J)$.

Proof. Let $\hat{T}_j(0, 0, \ldots, \delta_j = 1, 0, \ldots, 0: f(x)\, x_{kj})$ be the B.L.U.E. under the model $\xi(0, 0, \ldots, \delta_j = 1, 0, \ldots, 0: f(x)\, x_{kj})$. Then by [5] we have the B.L.U.E.

$$\hat{T}_j(0, 0, \ldots, 0, \delta_j = 1, 0, \ldots, 0: f(x)\, x_{kj}) = \sum_{k\in s} Y_k + \Big(\sum_{k\in\bar{s}} x_{kj}\Big) \frac{\sum_{k\in s} Y_k/f(x_k)}{\sum_{k\in s} x_{kj}/f(x_k)}.$$

If $s = s^*(J)$, then

$$\hat{T}_j(0, 0, \ldots, 0, \delta_j = 1, 0, \ldots, 0: f(x)\, x_{kj}) = \hat{T}(1: f(x)), \quad j = 0, 1, \ldots, J.$$

Thus $\hat{T}(1: f(x))$ is B.L.U.E. under the model $\xi(0, 0, \ldots, 0, \delta_j = 1, 0, \ldots, 0: f(x)\, x_{kj})$, $j = 0, 1, \ldots, J$. We now consider the model $\xi_j(\delta_0, \ldots, \delta_j = 1, \ldots, \delta_j: f(x)\, x_{kj})$. Since the expression

$$\mathrm{E}_{\xi_j}[\hat{T}(1: f(x) - T]^2 = \mathrm{Var}_{\xi_j}\Big(\hat{T}(1: f(x)) - \sum_{k\in s} Y_k\Big) + \mathrm{Var}_{\xi_j}\Big(\sum_{k\in\bar{s}} Y_k\Big)$$

(see [3], p. 882) depends only on the function $f(x)\, x_{kj}$ and the estimator $\hat{T}(1: f(x))$ is unbiased under the model ξ_j, we conclude that $\hat{T}(1: f(x))$ is B.L.U.E. under the model ξ_j, $j = 0, 1, \ldots, J$. But we have

$$\mathrm{E}_{\xi}[\hat{T}(1: f(x)) - T]^2 = \sum_{j=0}^{J} \delta_j a_j \mathrm{E}_{\xi_j}[\hat{T}(1: f(x)) - T]^2$$

and $\hat{T}(1: f(x))$ is unbiased under the model $\xi(\delta_0, \delta_1, \ldots, \delta_J: f^*(x))$. Then $\hat{T}(1: f(x))$ is B.L.U.E. under the model ξ.

4. Polynomial regression models. Suppose that $x_{kj} = x_k^j$, $k = 1, \ldots, N$ and $j = 0, 1, \ldots, J$. Then this particular model $\xi(\delta_0, \delta_1, \ldots, \delta_j: f^0(x))$ is known as the *polynomial regression model*. Using (3.1) for $f(x) = x$ we see that $s^*(J)$ is a sample such that

$$\frac{\sum_{k\in\bar{s}} x_k^j}{N-n} = \frac{\sum_{k\in s} x_k^{j-1}}{\sum_{k\in s} 1/x_k}, \quad j = 0, 1, \ldots, J. \tag{4.1}$$

Under the model $\xi(\delta_0, \delta_1, \ldots, \delta_j: x)$ the mean squared error (M.S.E.) of the estimator $\hat{T}(1:x)$ is of the form ([3], p. 882)

$$\mathrm{E}_{\xi}[\hat{T}(1:x) - T]^2 = \sigma^2 \left[\frac{(N-n)^2}{\sum_{k\in s} 1/x_k} + \sum_{k\in\bar{s}} x_k\right] \tag{4.2}$$

for any s. If we choose the sample $s = s^*(J)$, the M.S.E. of $\hat{T}(1:x)$ turns out to be

$$\frac{\sigma^2 N(N-n)\bar{x}_{\bar{s}}}{n}, \quad \text{where } \bar{x}_{\bar{s}} = \frac{\sum_{k\in\bar{s}} x_k}{N-n}.$$

For a small sampling fraction, $\bar{x}_{\bar{s}}$ is approximately equal to

$$\bar{x} = \frac{\sum_{k=1}^{N} x_k}{N}.$$

Consequently,

$$\text{M.S.E.} \simeq \frac{\sigma^2 N(N-n)\bar{x}}{n}. \tag{4.3}$$

Remarks. (1) Note that (4.3) is the same expression as (3.1) in [3] under the condition (3.1) for $x_{kj} = x_k^j$ and $f(x) = 1$.

(2) Suppose we adopt the model $\xi(1{:}x)$ and the sampling fraction is small. Then (4.2) is approximately equal to

$$\sigma^2 \left[\frac{(N-n)^2}{\sum_{k \in s} 1/x_k} + (N-n)\bar{x} \right],$$

which is minimized if we choose the sample such that $\sum_{k \in s} 1/x_k$ is the maximum (optimal sample). Condition (4.1) provides protection against the bias under the general polynomial regression model, but some efficiency is lost with respect to the optimal sample under the model $\xi(1{:}x)$ (see [3]).

(3) By [5], the polynomial regression model with $f^*(x) = \sigma_1^2 x + \sigma_2^2 x^2$ is often a realistic model. Next we will compare the expansion estimator with a balanced sample, i.e., $f(x) = 1$, and the estimator $\hat{T}(1{:}x)$ with $s = s^*(J)$, both under the polynomial regression model with $f^*(x) = \sigma_1^2 x + \sigma_2^2 x^2$. It is interesting to note that these estimators are B.L.U.E. under this model with their respective samples.

THEOREM 4.1. *The estimator $\hat{T}(1{:}x)$ with $s = s^*(J)$ is more efficient than the expansion estimator with a balanced sample, both under the polynomial regression model $\xi(1, 1{:}f^*(x))$, i.e.,*

$$E_\xi[\hat{T}(1{:}x) - T]^2 \leqslant E_\xi[\hat{T}(1{:}1) - T]^2, \quad \textit{where } f^*(x) = \sigma_1^2 x + \sigma_2^2 x^2.$$

Proof. If $f(x) = 1$, we obtain from (3.1) under general regression models the condition of balanced sample $\bar{x}_s^{(j)} = \bar{x}^{(j)}$, where

$$\bar{x}_s^{(j)} = \sum_{k \in s} \frac{x_k^j}{n} \quad \text{and} \quad \bar{x}^{(j)} = \sum_{k=1}^{N} \frac{x_k^j}{N}, \quad j = 0, \ldots, J.$$

The M.S.E. of the estimator $\hat{T}(1{:}1)$ under the polynomial regression model $\xi(1, 1{:}f^*(x))$ with balanced sampling, by [5] and [3], is

$$E_\xi[\hat{T}(1{:}1) - T]^2 = \frac{\sigma^2 N(N-n)}{n} [\sigma_1^2 \bar{x} + \sigma_2^2 \bar{x}^{(2)}]. \tag{4.4}$$

If $s = s^*(J)$, it follows from (4.1) that the M.S.E. of $\hat{T}(1{:}x)$ under the model $\xi(1, 1{:}f^*(x))$ is

$$\text{(4.5)} \qquad E_\xi[\hat{T}(1{:}x)-T]^2 = \sigma^2\left[\sigma_1^2\,\bar{x}_{\bar{s}}+\sigma_2^2\,\bar{x}_{\bar{s}}^2\left(1-\frac{n}{N}\right)+\sigma_2^2\frac{n}{N}\,\bar{x}_{\bar{s}}^{(2)}\right],$$

where

$$\bar{x}_{\bar{s}}^{(j)} = \frac{\sum\limits_{k\in\bar{s}} x_k^j}{N-n}, \qquad j = 0, 1, \ldots, J.$$

It follows from (4.1) for $j = 1$ and $j = 2$ (by the Cauchy-Schwarz inequality and Jensen's inequality) that $\bar{x}_{\bar{s}} \leqslant \bar{x}$ and $\bar{x}_{\bar{s}}^{(2)} \leqslant \bar{x}^{(2)}$, respectively. We conclude from (4.4) and (4.5) that

$$E_\xi[\hat{T}(1{:}x)-T]^2 \leqslant E_\xi[\hat{T}(1{:}1)-T]^2.$$

5. The meaning of condition (4.1). It may be difficult to obtain a sample which exactly satisfies (4.1). On the other hand, if we consider a special sampling design, it is possible to obtain a sample which approximately satisfies condition (4.1).

Definition 5.1. The function $P(s)$ such that $P(s) \geqslant 0$ for all $s \in S$, where S is the set of all samples, and $\sum_{s\in S} P(s) = 1$ is called the *sampling design.*

Definition 5.2. The *inclusion probability* π_k *of unit* k is the probability of selecting that unit, i.e.,

$$\pi_k = \sum_{s\supset\{k\}} P(s),$$

where the summation extends over all samples s such that $k \in s$. By [1], p. 11, we have

$$\sum_{k=1}^{N} \pi_k = n,$$

where n is the sample size.

THEOREM 5.1. *If*

$$\pi_k = 1-\frac{(N-n)/x_k}{\sum\limits_{k=1}^{N} 1/x_k} \geqslant 0,$$

then

$$E_p\left[\frac{\sum\limits_{k\in\bar{s}} x_k^j}{N-n}\right] = \sum_{k=1}^{N}\frac{x_k^{j-1}}{\sum\limits_{k=1}^{N} 1/x_k}, \qquad j = 0, 1, \ldots, J,$$

where E_p *denotes the expectation with respect to* $P(s)$.

Proof. We have

$$\mathrm{E}_p\left[\frac{\sum_{k\in\bar{s}} x_k^j}{N-n}\right] = \sum_{\bar{s}} \frac{\sum_{k\in\bar{s}} x_k^j}{N-n} P(\bar{s}) = \sum_{k=1}^{N} \frac{x_k^j}{N-n} \sum_{\bar{s}\supset\{k\}} P(\bar{s}) = \sum_{k=1}^{N} \frac{x_k^j}{N-n}(1-\pi_k)$$
$$= \sum_{k=1}^{N} \frac{x_k^{j-1}}{\sum_{k=1}^{N} 1/x_k}.$$

Remark. We conclude from Theorem 5.1 that for sufficiently large n and with a large sampling fraction the condition (4.1) is approximately satisfied if we choose a sampling design with the inclusion probability

$$1-\frac{(N-n)/x_k}{\sum_{k=1}^{N} 1/x_k} \geqslant 0, \quad k = 1, \ldots, N.$$

6. Stratified random sampling. The purpose of this section is to prove that the stratified sample and condition (4.1) together imply a higher efficiency than is achieved by a sample s which satisfies only condition (4.1). By [4], the population is divided into N strata as follows: N_1 units with the smallest x values form stratum 1, the next N_2 units form stratum 2, etc. A sample s_h of size n_h is selected from the N_h units in the h-th stratum.

Remarks. (1) A natural estimator for t under the model $\xi(1:x)$ is

$$\hat{T}_{st}(1:x) = \sum_{h=1}^{H} \hat{T}_h(1:x),$$

where

$$\hat{T}_h(1:x) = \sum_{k\in s_h} Y_{kh} + \frac{\sum_{k\in s_h} Y_{kh}/x_{kh}}{\sum_{k\in s_h} 1/x_{kh}}(N_h-n_h), \quad h = 1, \ldots, H,$$

and

$$t = \sum_{h=1}^{H}\sum_{k=1}^{N_h} y_{kh} = \sum_{h=1}^{H} t_h, \quad t_h = \sum_{k=1}^{N_h} y_{kh}.$$

(2) If

$$\frac{\sum_{k\in\bar{s}_h} x_{kh}^j}{N_h-n_h} = \frac{\sum_{k\in s_h} x_{kh}^{j-1}}{\sum_{k\in s_h} 1/x_{kh}}, \quad h = 1, \ldots, H \tag{6.1}$$

(one-step balanced), the estimator $\hat{T}_h(1{:}x)$ is protected against the bias with respect to T_h under the superpopulation model

$$y_{kh} = \sum_{j=0}^{J} \delta_j \beta_{jh} x_{kh}^j + e_{kh}\sqrt{f_{kh}^0}, \qquad h = 1, \ldots, H.$$

(3) We have

$$E_{\xi_{st}}[\hat{T}_{st} - T]^2 = \sum_{h=1}^{H} E_{\xi_{st}}[\hat{T}_h(1{:}x) - T_h]^2 = \sigma^2 \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} \bar{x}_{\bar{s}_h}$$

if condition (6.1) is satisfied, where ξ_{st} indicates the above model for $h = 1, \ldots, H$.

(4) *Optimal allocation sample.* Suppose the cost of sampling is given by a fixed c_0 plus the cost c_h for each unit sampled in stratum h. Let the total cost be

$$C = c_0 + \sum_{h=1}^{H} c_h n_h.$$

If (6.1) is satisfied, then $\bar{x}_{\bar{s}_h} \leqslant \bar{x}_h$ and the M.S.E. of the estimator $\hat{T}_{st}$ is less than or equal to

$$\sigma^2 \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h} \bar{x}_h. \tag{6.2}$$

By [4] the expression (6.2) is minimized when $n_h \alpha N_h \bar{x}_h^{1/2}$, $h = 1, \ldots, H$, under the condition that C is fixed and c_h is constant in each stratum h (optimal allocation).

THEOREM 6.1. *If the sampling fraction is small and* (4.1) *is satisfied, then*

$$E_\xi[\hat{T}(1{:}x) - T]^2 - E_{\xi_{st}}[\hat{T}_{st}(1{:}x) - T]^2$$

under a one-step balanced sampling in each stratum h and $n_h \alpha N_h \bar{x}_h^{1/2}$.

Proof. From (4.3) and (6.2) we have

$$\begin{aligned}
&E_\xi[\hat{T}(1{:}x) - T]^2 - E_{\xi_{st}}[\hat{T}_{st}(1{:}x) - T]^2 \\
&\simeq \sigma^2 \frac{N(N-n)\bar{x}}{n} - \sigma^2 \sum_{h=1}^{H} \frac{N_h}{n_h}(N_h - n_h)\bar{x}_{\bar{s}_h} \\
&\geqslant \sigma^2 \frac{N(N-n)\bar{x}}{n} - \sigma^2 \sum_{h=1}^{H} \frac{N_h(N_h - n_h)}{n_h}\bar{x}_h \\
&= \sigma^2 \frac{N(N-n)\bar{x}}{n} - \sigma^2 \{[\sum_{h=1}^{N} N_h \bar{x}_h^{1/2}]^2/n - N\bar{x}\} \\
&= \frac{\sigma^2}{n}[N^2\bar{x} - (\sum_{h=1}^{H} N_n \bar{x}_h^{1/2})^2]
\end{aligned}$$

$$= \frac{\sigma^2}{n} \{\bar{N}^2 \bar{x} - [\sum_{h=1}^{N} N_h^{1/2} (N_h \bar{x}_h)^{1/2}]^2\}$$

$$\geqslant 0,$$

where the last inequality holds by the Cauchy-Schwartz inequality.

Acknowledgement. The author is very grateful to the referees for their comments and suggestions.

References

[1] C. M. Cassel, C. E. Sarndal and J. H. Wretman, *Foundations of inference in survey sampling*, J. Wiley, New York 1977.

[2] C. R. Rao, *Statistical inference and its applications*, 2nd ed., J. Wiley, New York 1973.

[3] R. M. Royall and J. Herson, *Robust estimation in finite population*, J. Amer. Statist. Assoc. 68 (1973), p. 880-889.

[4] — *Robust estimation in finite population, II. Stratification on a size variable*, ibidem 68 (1973), p. 890-893.

[5] A. J. Scott, K. R. W. Brewer and E. W. Ho, *Finite population sampling and robust estimation*, ibidem 73 (1978), p. 362.

Instituto de Mathemática e Estatistica
Universidade de São Paulo
São Paulo, Brasil

Received on 24. 7. 1980;
revised version on 24. 5. 1982