Survey Methodology, December 1992
Vol. 18, No. 2, pp. 219–224
Statistics Canada

An Estimating Function Approach to Finite Population Estimation

HAROLD J. MANTEL[1]

ABSTRACT

Godambe and Thompson (1986) define and develop simultaneous optimal estimation of superpopulation and finite population parameters based on a superpopulation model and a survey sampling design. Their theory defines the finite population parameter, θ_N, as the solution of the optimal estimating equation for the superpopulation parameter θ; however, some other finite population parameter, ϕ, may be of interest. We propose to extend the superpopulation model in such a way that the parameter of interest, ϕ, is a known function of θ_N, say $\phi = f(\theta_N)$. Then ϕ is optimally estimated by $f(\theta_s)$, where θ_s is the optimal estimator of θ_N, as given by Godambe and Thompson (1986), based on the sample s and the sampling design.

KEY WORDS: Estimating functions; Generalized linear estimator; Finite population parameter.

1. ESTIMATION OF A MEAN

The problem discussed in this paper is the estimation of a finite population parameter such as the mean based on a sample survey. There is also a hypothesized superpopulation regression model relating the variable of interest to some known covariables. The objective is an estimation procedure which has good properties with respect to both the sampling design and the hypothesized model. The approach here is based on the work of Godambe and Thompson (1986).

We suppose that we have a finite population of labeled individuals $\boldsymbol{P} = \{i{:}i = 1, \ldots, N\}$. With each individual i is associated an unknown variable y_i and a vector of covariables, $\boldsymbol{x}_i$. The vector $\boldsymbol{x}_i$ may be known for all $i \in \boldsymbol{P}$ or only for i in the sample and the population mean $\bar{\boldsymbol{x}}_N$ would be known. Letting E_m denote expectation with respect to the superpopulation model, the model assumptions are:

(i) y_i and y_j are independent for $i \neq j$

(ii) $E_m(y_i) = \boldsymbol{x}_i^T\beta$ for some unknown real vector β

(iii) $E_m(y_i - \boldsymbol{x}_i^T\beta)^2 = \sigma^2 v_i$, $i = 1, \ldots, N$, for known v_i and some unknown σ^2.

Following Godambe and Thompson (1986) we define a finite population parameter $\hat{\beta}_N$ as the solution of the linearly optimal estimating equation

$$g^* = \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i^T\beta)\boldsymbol{x}_i/v_i = 0, \tag{1}$$

that is,

$$\hat{\beta}_N = (X_N^T V_N^{-1} X_N)^{-1} X_N^T V_N^{-1} y_N, \tag{2}$$

[1] H.J. Mantel, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

where $y_N^T = (y_1, \ldots, y_N)$, V_N is a diagonal matrix with entries $v_1, \ldots, v_N$, and X_N is a matrix with N rows, the ith row being $\boldsymbol{x}_i^T$.

Now $\hat{\beta}_N$ is unknown. Godambe and Thompson (1986) defined and developed simultaneous optimal estimation of β and $\hat{\beta}_N$ based on the model and the sampling design. We will denote the data from a sample survey by $\chi_s = \{(i, y_i), i \in s\}$.

For simultaneous estimation of β and $\hat{\beta}_N$ we consider estimating functions $\boldsymbol{h}(\chi_s, \beta)$ such that $E_p(\boldsymbol{h}) = \boldsymbol{g}^*$ in (1), where E_p denotes expectation with respect to the sampling design. A function $\boldsymbol{h}^*$ in this class is called optimal if for all other $\boldsymbol{h}$ in the class $E_m E_p\{\boldsymbol{h}\boldsymbol{h}^T\} - E_m E_p\{\boldsymbol{h}^*\boldsymbol{h}^{*T}\}$ is non-negative definite. Theorem 1 of Godambe and Thompson (1986) shows that the optimal function $\boldsymbol{h}^*$ is given by

$$\boldsymbol{h}^*(\chi_s, \beta) = \sum_{i \in s} (y_i - \boldsymbol{x}_i^T\beta)\boldsymbol{x}_i/\pi_i v_i, \tag{3}$$

where π_i is the probability under the sampling design that individual i is included in the sample s. We will denote the root of this function by $\hat{\beta}_s$, that is,

$$\hat{\beta}_s = (X_s^T \Pi_s^{-1} V_s^{-1} X_s)^{-1} X_s^T \Pi_s^{-1} V_s^{-1} y_s, \tag{4}$$

where y_s is the vector of y_is for $i \in s$, Π_s and V_s are diagonal matrices with entries π_i and v_i respectively, $i \in s$, and X_s is the matrix with rows $\boldsymbol{x}_i^T$, $i \in s$.

So far we have discussed only estimation of β or $\hat{\beta}_N$. Our problem was to estimate $\bar{y}_N$, the population mean of the y_is. One possibility is to use a generalized regression estimator,

$$\bar{y}_{\text{GREG}} = \bar{x}_N^T \hat{\beta}_s + \mathbf{1}_s^T \Pi_s^{-1}(y_s - X_s\hat{\beta}_s)/N, \tag{5}$$

where $\mathbf{1}_s$ is a vector of 1's whose length is the size of the sample s. This estimator is discussed, for example, by Särndal, Swensson and Wretman (1992). The first part of the estimator gives good model properties while the second part gives good design properties. However, the model and design justifications of $\bar{y}_{\text{GREG}}$ in (5) do not depend on the particular form of $\hat{\beta}_s$, and there is no immediately apparent reason why $\hat{\beta}_s$ in (5) could not be replaced by a purely model based estimator of β. The design optimality of $\hat{\beta}_s$ is apparently irrelevant.

The estimator we will propose here more closely integrates the hypothesized model with the finite population parameter $\bar{y}_N$. Since $\hat{\beta}_N$ in (2) is optimally estimated by $\hat{\beta}_s$ in (4), functions of $\hat{\beta}_N$ are optimally estimated by the same function of $\hat{\beta}_s$. If $\bar{y}_N = \boldsymbol{u}^T\hat{\beta}_N$ for some vector $\boldsymbol{u}$ then we would estimate $\bar{y}_N$ by $\boldsymbol{u}^T\hat{\beta}_s$. Such a $\boldsymbol{u}$ exists if and only if $V_N \mathbf{1}_N$ is in the column space of X_N, in which case, with $V_N\mathbf{1}_N = X_N\boldsymbol{a}$, we may take $\boldsymbol{u} = X_N^T V_N^{-1} X_N \boldsymbol{a}/N = \bar{x}_N$. The idea then is that if $V_N\mathbf{1}_N$ is not in the column space of X_N, we will add it. In doing so we lose something of model efficiency, though the augmented model remains valid in light of the original model. We relax model efficiency to gain some sort of finite population relevance. As an interesting special case we note that when the model variances do not depend on i our approach leads to including an arbitrary constant term in the regression model.

The approach taken here seems quite similar to that of Little (1983) who suggests model based estimation restricting attention to models that yield asymptotically design consistent estimators. Alternatively, Isaki and Fuller (1982) suggest restricting to designs for which the model based estimator is asymptotically design consistent.

2. COMPARISON TO THE GENERALIZED REGRESSION ESTIMATOR

Let W_N be the design matrix for the augmented model, that is

$$W_N = (V_N \mathbf{1}_N, X_N). \tag{6}$$

For the discussion of this section we assume that $V_N \mathbf{1}_N$ is not in the column space of X_N. Similarly, let W_s be the augmented form of X_s, and γ, $\hat{\gamma}_N$, and $\hat{\gamma}_s$ be the augmented forms of β, $\hat{\beta}_N$, and $\hat{\beta}_s$ respectively.

For convenience, we will refer to our estimator of the population mean as the augmented regression estimator,

$$\bar{y}_{\text{AREG}} = \bar{w}_N^T \hat{\gamma}_s. \tag{7}$$

We first show that $\bar{y}_{\text{AREG}}$ is also a type of generalized difference estimator. From (6), if $\boldsymbol{u}$ is a vector of appropriate length with the first entry equal to one and the rest zeros then $W_N \boldsymbol{u} = V_N \mathbf{1}_N$ and $W_s \boldsymbol{u} = V_s \mathbf{1}_s$. Then

$$\mathbf{1}_s^T \Pi_s^{-1} W_s \hat{\gamma}_s = \boldsymbol{u}^T W_s^T V_s^{-1} \Pi_s^{-1} W_s \hat{\gamma}_s = \boldsymbol{u}^T W_s^T V_s^{-1} \Pi_s^{-1} \boldsymbol{y}_s = \mathbf{1}_s^T \Pi_s^{-1} \boldsymbol{y}_s$$

and it follows that the second part of the generalized regression estimator in (5) with $\hat{\beta}_s$ replaced by $\hat{\gamma}_s$ is equal to 0.

Secondly, let us compare $\bar{y}_{\text{AREG}}$ in (7) to $\bar{y}_{\text{GREG}}$ in (5). A few tedious calculations give us that

$$\bar{y}_{\text{AREG}} = \bar{x}_N \hat{\beta}_s + (c_1/c_2) \mathbf{1}_s^T \Pi_s^{-1} (y_s - X_s \hat{\beta}_s)/N,$$

where

$$c_1 = \mathbf{1}_N^T (V_N \mathbf{1}_N - X_N (X_s^T V_s^{-1} \Pi_s^{-1} X_s)^{-1} X_s^T \Pi_s^{-1} \mathbf{1}_s)$$

and

$$c_2 = \mathbf{1}_s^T \Pi_s^{-1} (V_s \mathbf{1}_s - X_s (X_s^T V_s^{-1} \Pi_s^{-1} X_s)^{-1} X_s^T \Pi_s^{-1} \mathbf{1}_s).$$

Written in this way $\bar{y}_{\text{AREG}}$ appears very similar to $\bar{y}_{\text{GREG}}$ except for an adjusted weight for the second part. It does not seem possible to give an heuristic explanation of the weight (c_1/c_2). However, we note that c_1 is just the population sum of the residuals from a weighted regression of the v_i's onto the x_i's based on the sample s, and c_2 looks something like a Horvitz-Thompson estimator of c_1, except that the residuals also depend on the sample s. For large samples from large populations we would expect (c_1/c_2) to be close to 1.

In comparing $\bar{y}_{\text{AREG}}$ with $\bar{y}_{\text{GREG}}$ we may say that $\bar{y}_{\text{AREG}}$ is more design based and $\bar{y}_{\text{GREG}}$ is more model based. Of course, $\bar{y}_{\text{GREG}}$ is design consistent, but $\bar{y}_{\text{AREG}}$ has also a finite sample design justification in that $\hat{\gamma}_s$ is the solution of an estimating equation which is design unbiased for the parameter defining equation of $\hat{\beta}_N$. Parameter defining equations are discussed by Godambe and Thompson (1984, 1986).

3. VARIANCE ESTIMATION AND CONFIDENCE INTERVALS

A method of confidence interval construction which would be consistent with the general philosophy of estimating functions would be to construct an asymptotically multivariate normal pivotal based on $\boldsymbol{h}^*$ and an estimator of its variance. Approximate confidence regions for $\hat{\gamma}_N$ would then correspond to probability regions of the estimated multivariate normal distribution of this approximate pivotal. However, we are not interested in $\hat{\gamma}_N$ but in a non-injective function of $\hat{\gamma}_N$. We will adopt the more straight-forward approach of estimating the variance of $\bar{y}_{\mathrm{AREG}}$ directly.

Särndal, Swensson, and Wretman (1989) have investigated variance estimation for $\bar{y}_{\mathrm{GREG}}$ in (5) for the case that the second part is zero. As we have seen in section 2, our estimator $\bar{y}_{\mathrm{AREG}}$ is precisely of that type. Their variance estimator may be written as

$$\hat{V}_g = \sum_{i\in s} \sum_{j\in s} \tilde{\Delta}_{ij} g_{is} \tilde{e}_{is} g_{js} \tilde{e}_{js}, \tag{8}$$

where $\tilde{\Delta}_{ij} = (\pi_{ij} - \pi_i\pi_j)/\pi_{ij}$, π_{ij} is the design probability that both individuals i and j are included in the sample s, g_{is} is the ith element of the row vector $\bar{w}_N^T(W_s^T V_s^{-1} \Pi_s^{-1} W_s)^{-1} W_s^T V_s^{-1}$, and $\tilde{e}_{is} = (y_i - x_i^T\hat{\gamma}_s)/\pi_i$. See Särndal, Swensson and Wretman (1989) for a detailed discussion of the model and design properties of $\hat{V}_g$ in (8). Note that $\bar{y}_{\mathrm{AREG}}$ in (7) may be written as $\bar{y}_{\mathrm{AREG}} = \sum_{i\in s} g_{is}\, y_i/\pi_i$ and

$$\bar{y}_{\mathrm{AREG}} - \bar{y}_N = \sum_{i\in s} g_{is}\tilde{e}_{iN} = \bar{w}_N^T(\hat{\gamma}_s - \hat{\gamma}_N),$$

where $\tilde{e}_{iN} = (y_i - w_i^T\hat{\gamma}_N)/\pi_i$. Now, with $V_N\mathbf{1}_N = W_N\boldsymbol{a}$, we have $\bar{w}_N^T = \mathbf{1}_N^T V_N V_N^{-1} W_N/N = \boldsymbol{a}^T W_N^T V_N^{-1} W_N/N$, so that for large samples g_{is} will be near $1/N$ for $i\in s$. The design variance of $\bar{y}_{\mathrm{AREG}}$ is then approximately equal to

$$\sum_{i\in P} \sum_{j\in P} \Delta_{ij}\tilde{e}_{iN}\tilde{e}_{jN}/N^2,$$

where $\Delta_{ij} = (\pi_{ij} - \pi_i\pi_j)$, and this may be estimated by

$$\hat{V}_1 = \sum_{i\in s} \sum_{i\in s} \tilde{\Delta}_{ij}\tilde{e}_{is}\tilde{e}_{js}/N^2. \tag{9}$$

$\hat{V}_1$ in (9) was considered in early work on the general regression estimator, for example, Särndal (1981, 1982). Now $\hat{V}_g$ in (8) may be thought of as a version of $\hat{V}_1$ in (9) adjusted for the realized values of g_{is}, $i\in s$. Särndal, Swensson and Wretman (1989) show that $\hat{V}_g$ in (8), as well as being design consistent for the design variance of $\bar{y}_{\mathrm{AREG}}$, is often model unbiased or nearly model unbiased for the model mean squared error of $\bar{y}_{\mathrm{AREG}}$.

Now approximate confidence intervals for $\bar{y}_N$ could be constructed based on a standard normal approximation to the distribution of $(\bar{y}_{\mathrm{AREG}} - \bar{y}_N)/\{\hat{V}_g\}^{1/2}$. The justification of this procedure, from both a design and a model point of view, is asymptotic and the question of its appropriateness for particular finite samples must be addressed. One possibility is to compare a set of confidence intervals obtained by this procedure to a set of purely model based intervals based on a further assumption of normality of errors and a t-statistic. If the two sets of intervals are wildly different there may be reason to doubt the validity of the jointly model and design based intervals, but more work is needed before this question can be answered satisfactorily.

An alternative approach to variance estimation in this framework is given by Binder (1983). The design variance of $\boldsymbol{h}^*$ as an estimator of $\boldsymbol{g}^*$ at $\hat{\gamma}_N$ could be estimated using standard design based techniques substituting $\hat{\gamma}_s$ for $\hat{\gamma}_N$, and then the variance of $\hat{\gamma}_s$ as an estimator of $\hat{\gamma}_N$ would be derived from a Taylor linearization of $\boldsymbol{h}^*$ about $\hat{\gamma}_N$. Taylor linearization could again be used to derive an estimator of the variance of a function of $\hat{\gamma}_s$ as an estimator of the same function of $\hat{\gamma}_N$.

4. AREAS FOR FURTHER RESEARCH

We have seen how the approach described here could be used for the estimation of finite population means or, more generally, for functions of linear regression parameters. It is natural to wonder whether and how the approach may be adapted to the estimation of other types of finite population parameters such as distribution functions and quantiles or to estimation for small areas.

Consider the special case of estimation of a distribution function at one point. There are two possible approaches to incorporate covariate information into a model. The first is to model the probability explicitly as a function of the covariates, an example is the logistic model. A second approach, which is common in the context of estimating a distribution function, as in Chambers and Dunstan (1986), Rao, Kovar and Mantel (1990), and others, is to model the residuals from a regression of the observed variable onto the covariables as being independent and identically distributed from some unknown distribution. The present approach requires that the parameter of interest be a function of the finite population parameter. Can this approach be adapted for the estimation of distribution functions or quantiles?

Another important problem in survey sampling is small area estimation, that is estimation of totals, means or proportions for subsets of the finite population. A good review is given in Platek, Rao, Särndal and Singh (1987). An obvious adaptation of the approach of Section 1 is to apply it separately within each domain of interest, what might be described as post-stratified generalized regression estimation. Note that this approach would require the totals of the covariates for each domain of interest. A very common approach in small area estimation is to borrow strength across areas via a model relating small areas to each other and to some covariates. A good review is given in Singh, Mantel and Thomas (1991). A very fruitful approach has been the empirical Bayes estimation based on random effects models which was introduced by Fay and Herriot (1979). Liang and Waclawiw (1990) discuss estimating functions for empirical Bayes models. Can the idea of modelling to borrow strength across small areas be formulated in such a way that the parameters of interest become functions of a population parameter?

ACKNOWLEDGEMENT

I am grateful to a referee and to the editor for helpful comments and suggestions.

REFERENCES

BINDER, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.

FAY, R.E., and HERRIOT, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

GODAMBE, V.P., and THOMPSON, M.E. (1984). Robust estimation through estimating equations. *Biometrika*, 71, 115-125.

GODAMBE, V.P., and THOMPSON, M.E. (1986). Parameters of superpopulation and survey population: their relationships and estimation. *International Statistical Review*, 54, 127-138.

ISAKI, C.T., and FULLER, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

LIANG, K.-Y., and WACLAWIW, M.A. (1990). Extension of the Stein Estimating Procedure through the use of estimating functions. *Journal of the American Statistical Association*, 85, 435-440.

LITTLE, R.J.A. (1983). Estimating a finite population mean from unequal probability samples. *Journal of the American Statistical Association*, 78, 596-604.

PLATEK, R., RAO, J.N.K., SÄRNDAL, C.-E., and SINGH, M.P. (Eds.) (1987). *Small Area Statistics An International Symposium*. New York: Wiley.

SÄRNDAL, C.-E. (1981). Frameworks for inference in survey sampling with applications to small area estimation and adjustment for nonresponse. *Bulletin of the International Statistical Institute*, 49, 494-513.

SÄRNDAL, C.-E. (1982). Implications of survey design for generalized regression estimation of linear functions. *Journal of Statistical Planning 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. *Biometrika*, 76, 527-537.

SÄRNDAL, C.-E., SWENSSON, B., and WRETMAN, J.H. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

SINGH, A.C., MANTEL, H.J., and THOMAS, B.W. (1991). Time series generalizations of Fay-Herriot estimation for small areas. *Proceedings of the Section on Survey Research Methods, American Statistical Association*.