A FRESH LOOK AT BIAS-ROBUST ESTIMATION IN A FINITE POPULATION

Phillip S. Kott, Energy Information Administration

Royall and Herson (1973) demonstrate that a non-random, balanced sample coupled with an expansion estimator forms a bias-robust strategy for estimating a finite population total under a polynomial regression model; that is to say, the strategy is unbiased no matter what values the parameters of the model take. Scott, Brewer, and Ho (1978) point out that there are many bias-robust strategies under such a model. They advise concentrating on the best linear unbiased (BLU) estimator given a sample with particular properties.

Cumberland and Royall (1981) observe that a π-balanced mean-of-ratios (mor) strategy is robust under a second order polynomial model (from here on the prefix "bias" will be eliminated from the term "bias-robust"). They note that under a particular error structure this strategy is better than the alternative offered by SBH. We will see that under the same error structure the π-balanced mor strategy is in fact optimal (has the minimum variance) among equally robust, linear estimation strategies. In addition, an intuitive justification of π-balancing conditions is advanced.

Section 1 provides a motivating example. Section 2 develops the basic model and theortical results. Section 3 extends the model to allow additional auxiliary varibles, while section 4 introduces a more general error structure. Section 5 discusses the asymptotic properties of a systematic πps sample drawn from a size ordered list.

1. A MOTIVATING EXAMPLE

Robust mor strategies are developed here with a different type of application in mind from that usually found in the literature. To demonstrate this difference, we contrast the example given in the introduction of Royall and Herson with one of our own. In R-H, a sample of hospitals is chosen to estimate the number of patient-days provided by a population of hospitals. The number a beds in each hospital is known. The authors propose a robust superpopulation model where the number of patient-days in each hospital is a function of the number of beds in that hospital plus a random error term. Eventially the specification of the error terms is assumed to be arbitrary; but in the initial, motivating formulation the errors have variances proportional to the respective number of hospital beds. This is not an unreasonable assumption when the number of patient-days in a particular hospital is a sum of independent random variables -- the number of patient-days in each of the hospital's beds. Our motivating example concerns a quantity-weighted average price for a designated commodity sold by a population of retailers. Each retailer sells every unit of the commodity at the same price, although prices may differ among retailers. If an average commodity price is estimated based on a sample of retailers, a robust model might assume each retailer's price is a function of her (his) quantity plus a random error term. It is not unreasonable to suppose in this case that the error terms are independent and indentically distributed. In the hospital example, an aggregate total, patient-days, is to be estimated not a weighted average. This difference between the two examples is more apparent than real, however. The estimated number of patient-days divided by the known total number of hospital beds results in an estimate of the weighted average of patient-days per bed. The weights in this case are the percentage number of beds in each hospital. These weights are (in principle) multiplied by the (average) numbers of patient-days per bed in the respective hospitals and then summed. The real difference between the two examples is that in the hospital example, the number of patient-days per bed may vary among the beds in the same hospital, but in the price example, each retailer's price applies to her entire quantity sold. As a result, it is reasonable to assume that the variance of the number of patient-days per hospital is directly proportional to the number of beds in that hospital. On the other hand, it is reasonable to assume that the variance of a retailer's revenue (her price times her quantity) is proportional to the <u>square</u> of her quantity.

2. THE BASIC MODELS

2.1 Preliminaries

Let P be a finite population of N units, where each unit i has attached to it a value q_i, unknown before sampling, and a known weight w_i, $\sum^N w_i=1$. The problem is to estimate the weighted average, $A=\sum^N w_i q_i$, with a sample, S, of n units and a linear estimator, $\hat{A}=\sum_S a_i q_i$. The a_i <u>may</u> be functions of S.

Before proceding the reader should be aware that the formulation above differs markedly from the standard one in the literature. The usual problem is to estimate a population total, $t=\sum^N y_i$, when each unit i has attached to it two values y_i and x_i, and only the latter is known for all units. To translate from that notation to ours, one can let $q_i=y_i/x_i$, $w_i=x_i/\sum^N x_i$, and $A=t/\sum^N x_i$. The advantage of the notation used here will be made clear in time.

In classical sampling theory, the units in the sample are randomly selected via a sampling design, D. A sampling design consists of a set of possible samples, $\mathcal{S}_D$, and the probabilities (all positive) of randomly selecting each element of that set. In a nonrandom sampling design, D^*, S^* is the sole element of $\mathcal{S}_{D^*}$.

A couple, $(\hat{A}=\sum_S a_i q_i, D)$ defines a <u>linear estimation strategy</u>. We say this strategy is <u>unbiased</u> if $E[\sum_S a_i q_i - A|S]=0$ for all $S\in\mathcal{S}_D$. An unbiased estimation strategy, (A, D), is <u>better</u> than an alternative unbiased strategy, $(\hat{A}', D')$, when the <u>conditional variance</u>, $E[(\hat{A}-A)^2|S]$, for <u>all</u> $S\in\mathcal{S}_D$ is no greater than $E[(\hat{A}'-A)^2|S]$ for <u>any</u> $S\in\mathcal{S}_{D'}$. A strategy is <u>optimal</u> among a class of strategies if it is better than all other strategies in the class. Note that "unbiased" and "variance" are defined with respect to the random q_i values and <u>not</u> the sampling probabilities.

In this paper, we speak of the optimal estimation strategy unbiased under a particular model rather than the best linear unbiased (BLU) estimator. The former is a much stronger concept; an optimal strategy is better than all other couples of an estimator and a sample design, while a BLU estimator need only be "optimal" given a particular sample.

2.2 Royall's Strategy

Royall (1970) shows that if the q_i are independent and identically distributed (iid), then an optimal, linear unbiased strategy, $(\hat{A}^*, D^*)$, obtains by relabeling the units so that $w_1\geq w_2\geq\ldots\geq w_N$, letting $\mathcal{S}_{D^*}$ contain the single sample, $S^*=\{1,2,\ldots,n\}$, and

$$\hat{A}^*=\sum_S w_i q_i + \sum_{\bar{S}} w_j \sum_S q_i/n .$$

Notice that the BLU estimator given <u>any</u> sample takes on the same form as above. In an optimal strategy, however, $\mathcal{S}_{D^*}$ may contain only samples in which $\min_S\{w_i\}\geq\max_{\bar{S}}\{w_i\}$.

The problem with Royall's strategy is that it is not very robust. If the q_i were not identically distributed, for example if each q_i were correlated with its respective w_i, $\hat{A}^*$ would not be unbiased given the sample in $\mathcal{S}_{D^*}$.

2.3 A More Robust Model

Let us expand Royall's specification of the q_i slightly:

$$q_i = b_0 + b_1 w_i + \varepsilon_i , \qquad (2.1)$$

where the ε_i are independent and identically distributed random variables. Note that nothing is lost by assuming $E(\varepsilon_i)=0$.

The expansion of the basic model to (2.1) is more intuitive than the usual expansion in the standard notation. When the initial model is expressed as $y_i=bx_i+\eta_i$, the obvious mathematical extension is to add a constant term, a: $y_i=a+bx_i+\eta_i$. Unfortunately, there is no convenient story attached to a non-zero intercept in many cases. For example, suppose y_i is the revenue collected by retailer i from sales of a designated commodity, while x_i is the quantity of

the commodity sold by i. A non-zero intercept literally means that on average either a retailer with no quantity will collect a positive revenue or a threshold quantity is needed before a retailer can collect a single dollar from sales.

Using the same example, $b_1<(>)0$ in (2.1) simply states that as the relative quantity sold by a retailer, w_i, increases, her price, on average, decreases (increases). In equation (2.1), the original model of a constant price among retailers (give or take a random error), has been extended to allow for economies or diseconomies of scale.

2.4 The Mean-of-Ratios Strategy

For a linear estimation strategy to be bias-robust under the extended model (2.1) - that is, unbiased for every set of b_k values - these two conditions must be satisfied for all $S\in\mathcal{S}_D$:

A1. $\sum_S a_i=1$;

A2. $\sum_S a_i w_i=\sum^N w_i^2$.

Optimality is attained among the class of strategies obeying both A1 and A2 when

$$\begin{aligned}\mathrm{Var}(\hat{A}\mid S\in\mathcal{S}_D)&=E\left[(\textstyle\sum_S a_i q_i-\sum^N w_i q_i\mid\cdot)^2\right]\\&=E\left[(\textstyle\sum_S a_i\varepsilon_i-\sum^N w_i q_i\mid\cdot)^2\right]\\&=\sigma_\varepsilon^2(\textstyle\sum_S a_i^2-\sum^N w_i^2)\end{aligned}\qquad(2.2)$$

is minimized (the last step depends on A2). Let us call the expression in the last line of (2.2) "V".

Consider the strategy $(\hat{A}_M=\sum_S q_i/n, D)$ where for all $S\in\mathcal{S}_D$, $\sum_S w_i/n=\sum^N w_i^2$. The strategy satisfies both constraints and results in a V value of $\sigma_\varepsilon^2(1/n-\sum w_i^2)$. This value is equal to the minimum value V attains under the single constraint A1; therefore, it also must be the minimum under the dual constraints A1 and A2.

Let us explore why the optimal, linear, robust strategy described above is restricted to samples satisfying $\sum_S w_i/n=\sum^N w_i^2$. The minimization of the last line of (2.2) constrained by A1 alone results in the estimator $\hat{A}_M=\sum_S q_i/n$ coupled with any sample design. The additional constraint A2 does not affect the estimator only the allowable samples. Recall that we are estimating a weighted average, $A=\sum^N w_i q_i$, yet the estimator, $\hat{A}_M$, is a straight arithmetic average. When each q_i is correlated with the auxiliary variable, w_i, it is evident that the average of the sampled w_i should equal the weighted average of the population w_i. In other words, $\sum_S w_i/n=\sum^N w_i^2$. This insight does not come easily using the standard notation. For example, Cumberland and Royall impose the following odd looking restriction on the sample: $\bar{x}\sum_s x_i/n-N^{-1}\sum^N x_i^2=0$ (p. 357).

We call any strategy containing the estimator $\hat{A}_M=\sum_S q_i/n$ a mean-of-ratios strategy. The name derives from the fact that in the standard notation, $q_i=y_i/x_i$. One familiar mor strategy is the Horvitz-Thompson (1952), in which sampled units are chosen randomly, without replacement, with probabilities proportionate to the w_i; i.e., $\mathrm{pr}(i\in S)=\pi_i=nw_i$. Note that πps samples will on average satisfy $\sum_S w_i/n=\sum^N w_i^2$ (the selection probablity weighted average of $\sum_S(w_i/n)$ over all the possible samples, $\sum^N(w_i/n)\mathrm{pr}(i\in S)$, equals $\sum^N w_i^2$). That is why samples satisfying this condition are called π-balanced on the w_i.

3. EXTENSIONS

It is a simple matter to extend the analysis of the previous section to allow for additional explanatory variables. To simplify matters, let us say that z_i is a lone additional auxiliary; i.e.,

$$q_i=b_0+b_1w_i+b_2z_i+\varepsilon_i,\qquad(3.1)$$

where $E(\varepsilon_i)=0$, and the ε_i are iid. Then $(\hat{A}_M, D)$ is optimal among linear robust estimation strategies when $\mathcal{S}_D$ is non-empty and $S\in\mathcal{S}_D$ only if

$$\sum_S w_i/n=\sum^N w_i^2$$

and

$$\sum_S z_i/n=\sum^N w_i z_i.\qquad(3.2)$$

As before the minimum variance a linear, robust strategy can attain is $\sigma_\varepsilon^2(1/n-\sum^N w_i^2)$. Clearly $(\hat{A}_M, D)$ is robust and has a variance equal to the minimum possible variance. A sample in $\mathcal{S}_D$ is said to be π-balanced with respect to the w_i and the z_i. It should now be obvious that no matter how many auxiliaries are added to the w_i, there is an optimal, linear, robust strategy when a sample π-balanced with respect to every auxiliary exists. It is $(\hat{A}_M, D)$, where contains only samples π-balanced with respect to all the auxiliaries (including the w_i).

Note that πps sampling with the w_i as the measure of size will on average (the selection probability weighted average) produce a π-balanced sample with respect to every possible auxiliary variable. As a result, the Horvitz-Thompson strategy is on average the optimal, linear robust estimation strategy given any specification of the non-random part of the q_i providing that the random parts are identically distributed. Compare this with Godambe (1955) where it is shown that the Horvitz-Thompson strategy has the least variance on average among linear, design-unbiased estimators (estimators that on average equal the value they estimate) when the q_i are iid.

Finally, observe that if the $z_i=1/w_i$ in (3.1), the model can be expressed in the standard notation as $y_i=a+bx_i+cx_i^2+\eta_i$,

where $E(\eta_i)=0$, $E(\eta_i^2)=\sigma_\varepsilon^2x_i^2$. This is the form analyzed in Cumberland and Royall. They required samples be π-balanced on the x_i (w_i) and the $1/x_i$ ($1/w_i$).

4. A MORE GENERAL ERROR STRUCTURE

In this paper the standard problem of estimating a population (P) total, $t=\sum^N y_i$, based on a sample (S) of units, where auxiliary quantities, x_i, are known for all units has been transmuted into estimating a weighted average, $A=\sum^N w_i q_i$, where $w_i=x_i/\sum^N x_j$, and $q_i=y_i/x_i$. The results can be easily translated into standard form as follows. Suppose the y_i are specified by

$$y_i=a+bx_i+cx_i^2+\eta_i,\qquad(4.1)$$

where the η_i are random variables with mean zero. The mean-of-ratios (mor) estimator, $\hat{t}_M=\sum^N x_j\sum_S(y_i/x_i)/n$, is an unbiased estimator of t when

$$\sum^N x_k\textstyle\sum_S x_i^{j-1}/n=\sum^N x_i^j;\ j=0,2.\qquad(4.2)$$

If the η_i are independent, and the variance of each error term satifies $\mathrm{Var}(\eta_i)=kx_i^2$, then the mor estimator coupled with a nonrandom sample obeying (4.2) forms a strategy optimal among all linear estimation strategies that are unbiased no matter what the parameter values. Moreover, if the are indendent, and

$$\mathrm{Var}(\eta_i)=k_1x_i+k_2x_i^2;\ k_1,\ k_2\geq0\qquad(4.3)$$

then this mor strategy is better than the expansion estimator coupled with a balanced sample of any postive order. (a sample, S, is balanced of order J if $\sum_S x_i^j/n=\sum^N x_i^j/N$ for $j=1,\ldots,J$).

The error structure in (4.3) serves a broad range of applications. It has a general explanation, which we will demonstrate using the hospital example in Royall and Herson.

The number of patient-days, y_i, in hospital i is the sum of the number of patient-days in each bed in the hospital, y_{im}, where $\sum_{m=1}^{x_i}y_{im}=y_i$. Each of the y_{im} is itself the sum of a determined term and two independent random error terms with means of zero. The first error term, ε_{im}^1, varies independently over the beds in hospital i and has variance k_1. The second, ε_{im}^2, is the same for every bed in i and has variance k_2. The variance in the number of patient days in hospital i is then

$$\mathrm{Var}(y_i)=\mathrm{Var}(\textstyle\sum^{x_i}y_{im})+\mathrm{Var}(\sum^{x_i}\varepsilon_{im}^1+x_i\varepsilon_{im}^2)=k_1x_i+k_2x_i^2.$$

This is the variance of η_i expressed in (4.3).

This paper offers no empirical verification of its assertion that robust mor strategies are better than Royall-Herson strategies for a broad range of applications. For that, the reader is directed to two articles by Royall and Cumberland (R and C, 1981, and C and R, 1981). The results there, when analyzed, strongly suggest that robust mor is the superior strategy for estimating totals in the six populations investigated. (Royall and Cumberland compute ratio estimators based on unrestricted simple random samples and nearly balanced samples for the six populations, while Cumberland and Royall compute mean-of-ratio estimators based on unrestricted PPS and nearly π-balanced samples).

5. SYSTEMATIC πPS SAMPLING

One convenient way to draw a sample without replacement but with probabilities proportionate to the w_i is by systematic ps sampling (Madow, 1949). In this section, we will confine our attention to systematic ps samples drawn from a population ordered by ascending x values. Consider the possible $D=\sum_s x_i/n$ values for the sample design under consideration. Relabel the units so that $x_1 \le \ldots \le x_N$, and let x_L^k and x_U^k be respectively the x values of the kth systematic draws of the samples with minimum and maximum D values (D_L and D_U). Observe that

B1. $x_L^1 = x_1$,
B2. $x_U^n = x_N$, and
B3. $x_U^k \le x_L^{k+1}$ for k=1,...,n-1.

Therefore

$$D_U - D_L \le (x_N - x_1)/n . \qquad (5.1)$$

As the sample size becomes arbitrarily large, the maximum difference between possible D values tends towards zero.

In order to analyze the asymptotic properties of our sampling design, it is necessary to let both the sample (n) and the population (N) become arbitrarily large. Since the w depend on the sample size, we must re-express (2.1) as

$$q_i = b_0 + b^* x_i + \varepsilon_i , \qquad (5.2)$$

where the ε_i are iid (for simplicity's sake only).

Under reasonable bounding conditions ($|x_i| < B_1$; $n\sum^N w_i^2 < B_2 < 1$), the bias of $\hat{A}_M$ is of order n^{-1} ($|E(\hat{A}_M - A)| \le |b^*(x_N - x_1)/n|$), while its standard deviation is of order $n^{-1/2}$ ($\sigma_\varepsilon \sqrt{(1 - n\sum^N w_i^2)/n}$). As n becomes arbitrarily large, therefore, the contribution of the bias of $\hat{A}_M$ to its mean squared error tends toward zero. Accordingly, we say that a sample from this design is asymptotically π-balanced.

It is possible to extend the model in (5.2) to allow more general expressions for the error structure. That is left for another time. Instead let us replace $b^* x_i$ by a finite linear combination of bounded monotonic transformations of the x_i; i. e.,

$$q_i = b_0 + \sum_{j=1}^{J} b_j h_{ji} + \varepsilon_i , \qquad (5.3)$$

where $h_{ji'} \ge h_{ji}$ for $i' > i$ and $|h_{iN}| < B_3$. As n becomes arbitrarily large, the contribution of the bias of $\hat{A}_M$ to its mean squared error again tends toward zero ($|E(\hat{A}_M - A)| \le \sum^J |b_j (h_{jN} - h_{j1})/n|$.) Thus a systematic πps sample from a x-ordered list is asymptotically π-balanced on all linear combinations of bounded monotonic transformations of the x_i.

The model-based asymptotic properties of the mor estimator coupled with a systematic πps sample from a size ordered list do not depend on the sample having a random start point. Any start point would do. By choosing a start point randomly, this estimation strategy becomes design unbiased (bias-robust on average given any additional auxiliary variable). Unfortunately, the strategy is not asymptotically design consistent (adc) in the Isaki-Fuller (1982) sense. (The estimation strategy is adc, trivially, in the sense of Brewer, 1979, because it is design unbiased.)

For an estimation strategy based an a random sampling design to be adc, its mean squared error must tend toward zero (in design probability) as n grows arbitrarily large no matter how the q_i are specified. To see why the mor strategy under consideration is not adc, consider the following specification of the q_i:

$$q_i = b_0 + b^* x_i + c(-1)^i + \varepsilon_i , \qquad (5.4)$$

where the ε_i are iid, $|c| > 0$, and $x_i = 1 - 2^{-i}$. Let n grow arbitrarily large holding N/n at an even integer value. $A = \sum^N x_i q_i / \sum^N x_i$ converges to $b_0 + b^*$. $\hat{A}_M - A$ converges either to c or -c depending on the randomly chosen start point. Consequently the mean squared error of $\hat{A}_M$ corverges to c rather than to zero.

REFERENCES

BREWER, K. W. (1979), "A Class of Robust Sampling Designs for Large-Scale Surveys," Journal of the American Statistical Association, 74, 911-915.

CASSEL, C. M., SARNDAL C. E., and WRETMAN, J. H. (1977), Foundations of Inference in Survey Sampling, New York: John Wiley and Sons.

CUMBERLAND, W. G. and ROYALL, R. M. (1981), "Prediction Models and Unequal Probability Sampling," Journal of the Royal Statistical Society B, 43, 353-367.

GODAMBE, V. P. (1955), "A Unified Theory of Sampling from Finite Populations," Journal of the Royal Statistical Society B, 17, 269-278.

HORVITZ, D. G. and THOMPSON, D. J. (1952), "A Generalization of Sampling Without Replacement From a Finite Universe," Journal of the American Statistical Association, 47, 663-685.

ISAKI, C. T. and FULLER, W. A. (1982), "Survey Design Under the Regression Superpopulation Model," Journal of the American Statistical Association, 77, 89-96.

MADOW, W. G. (1949), "On the Theory of Systematic Sampling, II," Annals of Mathematical Statistics, 20, 333-354.

ROYALL, R. M. (1970), "On Finite Population Sampling Theory Under Certain Linear Regression Models," Biometrika, 57, 2, 377-87.

___ and CUMBERLAND, W. G. (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance (with discussion)," Journal of the American Statistical Association, 76, 66-88.

___ and HERSON, J. (1973), "Robust Estimation in Finite Populations I," Journal of the American Statistical Association, 68, 880-889.

SCOTT, A. J., BREWER, K. R., and HO, E. W. (1978), "Finite Population Sampling and Robust Estimation," Journal of the American Statistical Association, 78, 359-361.