A GENERAL COMPOSITE ESTIMATOR

Stephen M. Woodruff, Bureau of Labor Statistics

1. Introduction

Estimates of some population parameter can usually be improved when supplementary information is used. For example, if we are estimating average hourly earnings at a given time we can usually improve the sample survey measurement at that time by averaging it with an updated estimate from the preceeding time (or times). This technique is used by the Bureau of the Census in their retail trade survey.(1) This type of improved estimator usually takes the form $aY + bX$ where X and Y are 2 estimators of the same thing and a and b are fixed real numbers.

This type of composite estimator often results from empirical Bayes estimation, for example, the James-Stein estimator.(11) Another example is the empirical Bayes estimate of variance that was derived in order to improve sample allocation.(9)

This paper examines the least restricted case of composite estimation. Let's start with two quite arbitrary real valued random variables, X and Y, defined on a probability space (X, B, P).(8) The only restriction placed on X and Y is that their second moments exist.

Then let w be any real number. How should we choose real numbers a and b so that $E(w-aY-bX)^2$ is minimized? When does a solution exist? What is the $\inf_{(a,b)} E(w-aY-bX)^2$ and how does this minimum compare to the minimum when a and b are restricted as: $a=1-b$, $0 < a < 1$? What happens if $EX = w$ or $EY = w$ or both? How robust is $Z(a,b) = aY + bX$ when a and b deviate from their optimal values? This paper examines these questions using geometric and tabular presentations.

2. Mathematical Structure

Let X be the composite estimator. That is $Z = aY + bX$ and let the covariance matrix of the vector $D = (Y, X, Z)$ be:

$$E(D'D) - E'(D)E(D) = \begin{bmatrix} V(Y) & C_{XY} & C_{YZ} \\ C_{XY} & V(X) & C_{XZ} \\ C_{YZ} & C_{XZ} & V(Z) \end{bmatrix}$$

This matrix can be written as a function of the variances and covariance of X and Y as:

$$\begin{bmatrix} V(Y) & C_{XY} & aV(Y) + bC_{XY} \\ C_{XY} & V(X) & aC_{XY} + bV(X) \\ aV(Y)+bC_{XY} & aC_{XY}+bV(X) & a^2V(Y)+b^2V(X)+2abC_{XY} \end{bmatrix}$$

The mean square error of Z is denoted as $M(Z) = E(Z-w)^2 = V(Z)+B^2(Z)$ where $B(Z) = E(Z)-w$ is the bias of Z in estimating w.

Thus from the covariance matrix of D the mean square error of Z is:

$$M(Z) = a^2V(Y)+b^2V(X)+2abC_{XY}+(aEY+bEX-w)^2$$

where EY denotes the expected value of Y. Given the covariance structure of X and Y, M(Z) is a function of a and b. It can be written as

$$M(Z) = a^2EY^2 + b^2EX^2 + 2abEXY$$
$$-2awEY - 2bwEX + w^2 .$$

Thus M(Z) is a continuous differentiable function of a and b which gets large as a or b or both get large in absolute value. It is also clear that it has an unique minimum which can be found by solving the system:

$$\frac{\partial M(Z)}{\partial a} = 0 \qquad \frac{\partial M(Z)}{\partial b} = 0 \tag{2.1}$$

The solution of this system of linear equations is the point (a^*, b^*) in the (a,b) plane given by:

$$a^* = w(EYEX^2 - EXEXY)/F \qquad b^* = w(EXEY^2 - EYEXY)/F \tag{2.2}$$

where $F = EX^2EY^2 - (EXY)^2 \geqslant 0$ by the Cauchy-Schwartz inequality.

The one restriction is $F \neq 0$ which is equivalent to the linear independence of X and Y.

Let $Z^* = a^*Y + b^*X$. We studied the properties of Z^* via computer simulation. These computer studies gave rise to the following conjecture.

(Note that $-w\,B(Z) = w\,(w-aEY-bEX)$ is a planar surface in the (a,b,u) coordinate system.) The point $(a^*, b^*, M(Z^*))$ is contained in the intersection of the surface M(Z) and $-wB(Z)$. The proof of the conjecture is a routine computation. Thus we get theorem 1.

THEOREM 1: $M(Z^*) = -wB(Z^*)$ (2.3)
where $Z^* = a^*Y + b^*X$.

The consequences of this theorem are.

COROLLARY 1: $|EZ^*| \leqslant |w|$ (2.4)

COROLLARY 2: $\sqrt{V(Z^*)} \leqslant |w/2|$ (2.5)

PROOF: Corollary one is trivial. To see corollary 2 write $M(Z^*)=V(Z^*)+B^2(Z^*)$ and substitute $-wB(Z^*)$ for $M(Z^*)$. This quadratic in $B(Z^*)$ then yields

$B(Z^*) = (-w \pm \sqrt{w^2 - 4v(z^*)})/2$ this implies
$w^2 - 4V(Z^*) \geq 0$ and this implies
$\sqrt{V(Z^*)} \leq |w/2|$.

Corollary 1 is another example of the well known fact that by shrinking an unbiased estimator toward the origin its mean square error can be reduced. corollary 2 gives an unconditional upper bound on the Relative Error of 1/2.

These results on the composite of two estimators can be generalized to any finite number of estimators. If we are given random variables, $X_1, X_2, ----, X_n$ and wish to choose real numbers, $a_1, a_2, ----a_n$ such that $E(w-\sum a_i X_i)^2$ is minimized we get:

$\hat{A} = w M^{-1} EX$ where

$\hat{A}' = (a_1^*, a_2^*, \ldots\ldots a_n^*)$

$M = E(X X')$

& $X' = (X_1, X_2 \ldots\ldots, X_n)$

The composite estimator with optimal weights is then, $Z^* = \hat{A}'X$ and its expected value is $w \cdot EX'M^{-1}EX$. Theorem 1 and its two corollaries obtain in this case as well as in the case of two random variables. A short computation gives: $M(Z^*) = w^2 (1 - EX'M^{-1}EX) = -wB(Z^*)$.

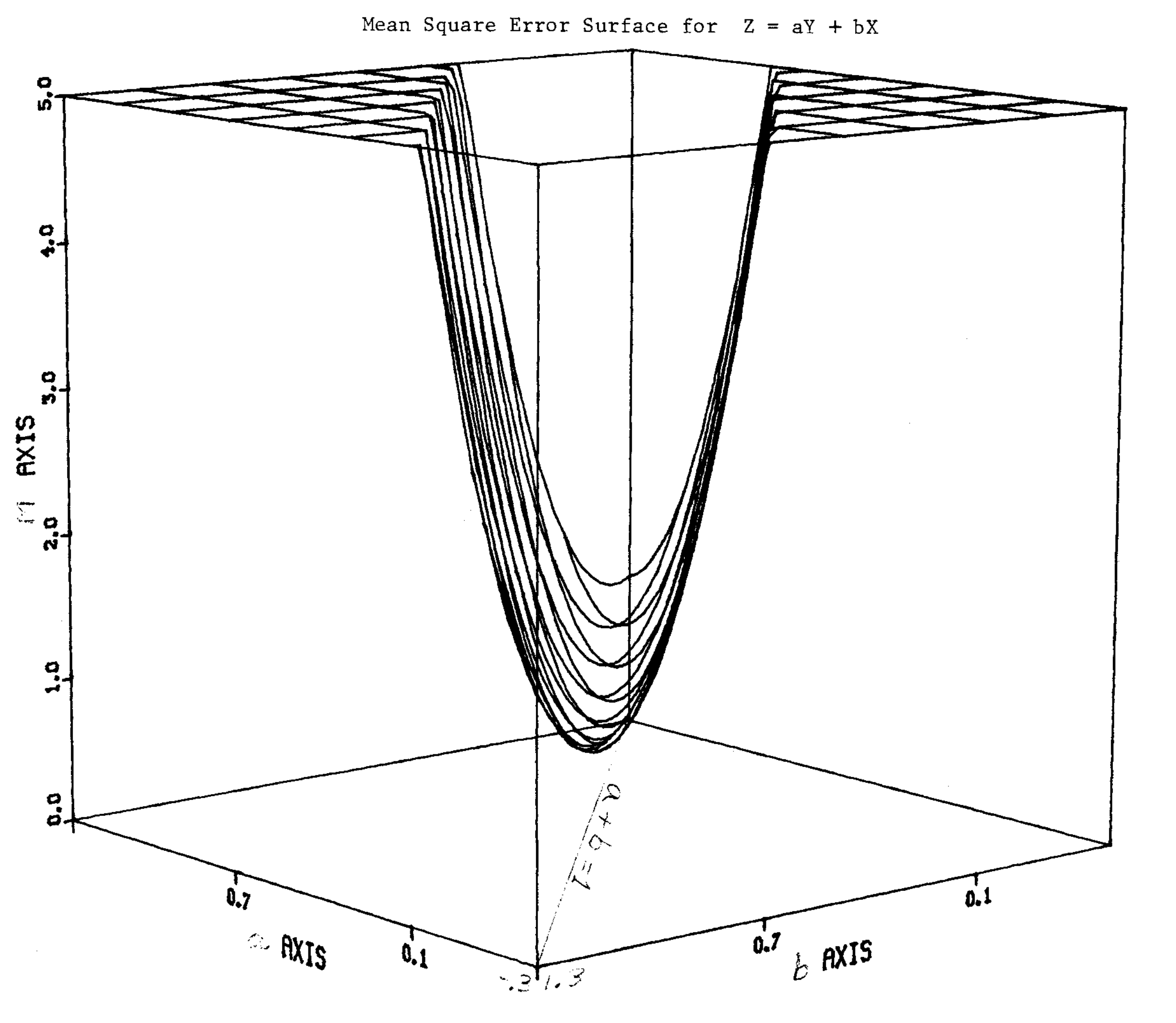
3. An Empirical Investigation

This general composite estimator was tested empirically on a sampling problem. Samples of various sizes were drawn from Normal populations according to the following model:

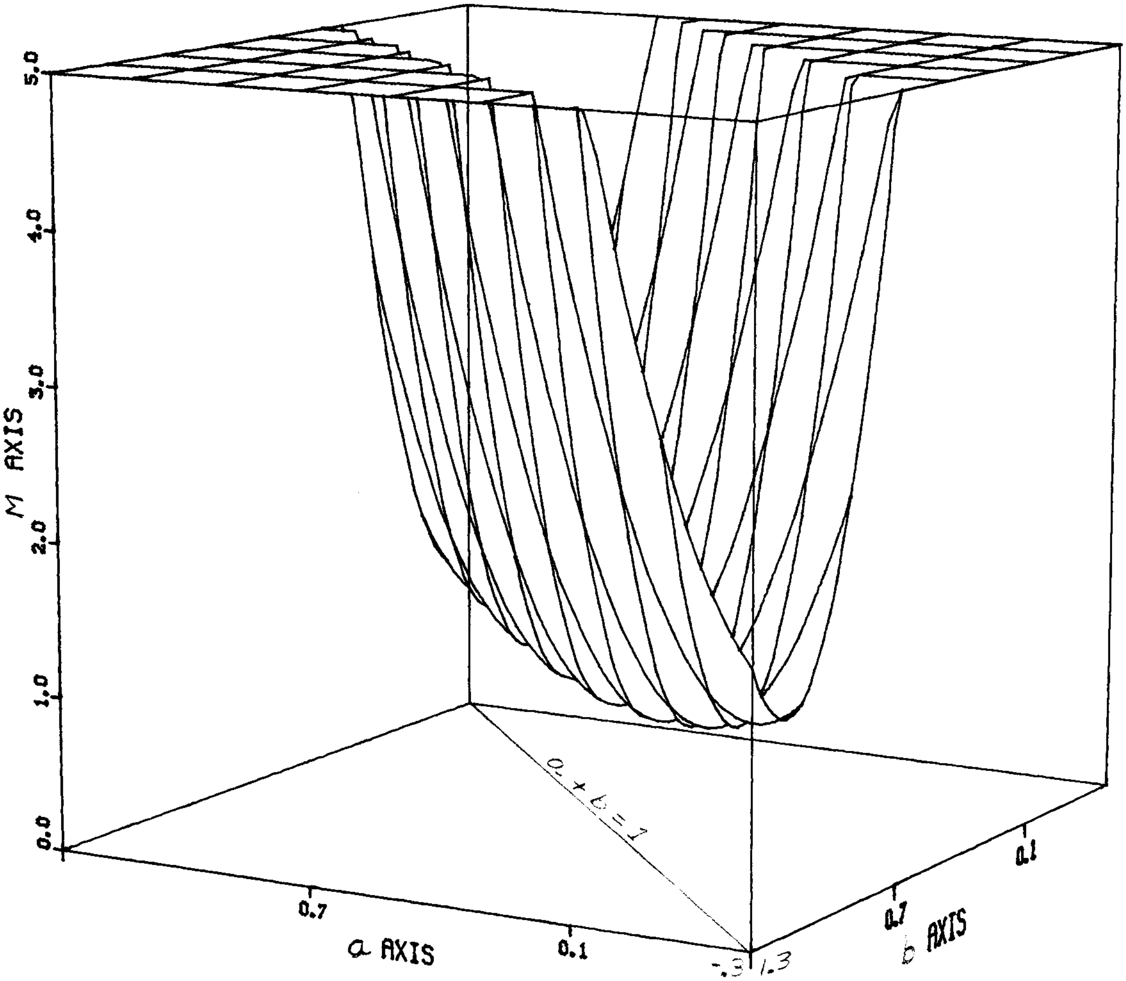
$$x_{ij} = f + gj + t_i \quad \text{for} \quad \begin{cases} j = 1,2,---,5 \\ i = 1,2,---,N \end{cases}$$

$$\& \quad x_{ij} = f + gj + t_{ij} \quad \text{for} \begin{cases} j = 1,2,---,5 \\ i = N+1, N+2, ----, N+M \end{cases}$$

where the t_i and t_{ij} are independent N(0,1) variates. Thus exactly N independent normal variates are generated for the first set of x's and 5M are generated for the second set of x's. This model can be thought of



Mean Square Error Surface for $Z = aY + bX$



as two independent samples of x's for each of five time periods. The first set of x's can be thought of as the overlap portion of the sample. That is, the units which remain in the sample for all five time periods. Note that the time period to time period correlation within this group is 1 since the same normal variate is used for all five time periods for any given i, $1 \leq i \leq N$. Saying this another way we have $\rho(x_{ik}, x_{il}) = 1$.

The second set of x's can be thought of as the M sample members which are drawn anew at each time period. All the variates in this group are independent for all $i > N$ and j, $1 \leq j \leq 5$. Thus the sample mean for time period j can be written as:

$$\bar{x}_j = (1/(N+M))(N\,\bar{x}_{Nj} + M\,\bar{x}_{Mj})$$

where $\bar{x}_{Nj}$ is the sample mean at time period j of the N units in the first group of x's. $\bar{x}_{Mj}$ is sample mean of the second group at time period j. Note that, $\bar{x}_{N1}$ + g = $\bar{x}_{N2}$, $\bar{x}_{N2}$ + g = $\bar{x}_{N3}$, etc. We compared estimators, $\sum_{k \leq j} a_k \bar{x}_k$, with $\bar{x}_j$ for j = 2, 3, 4, and 5. The measure of comparison is an estimate of average mean square error for 20 replications of the experiment.

These simulations were done using a variety of sample sizes for the 2 portions of the sample at each time period. N, the size of the overlap portion, ran from 25 to 100 and M, the size of the independent portion ran from N/2 to 2.5 N, for each different N that was used.

When the covariance matrix of the $\bar{x}_j$ is known then the simulations showed that the composite estimator provides substantial gaines. When this matrix must be estimated from the sample data the general composite estimator did poorly . This lack of robustness with respect to estimating the covariance matrix is a serious drawback. it would certainly present a problem in sample survey applications where the covariance matrix must be estimated.

The three dimensional plots show very clearly why this estimator can be very sensitive to minor deviations from the optimum weights. These are graphs of the truncated mean square error surface as a function of the weights in the two component case of the general composite estimator. The two plots shown here are views of the same surface from different angles. They show the mean square error surface in the case when V(X) = 1.2, V(Y) = .8, CXY = .6, EX = 4.2 and EY = 3.8 and $w = 4$.

This visual aid shows very clearly that restricting the weights to the line, $a + b = 1$, can have a very positive effect on robustness. If the equation of the straight line lying directly under the "gutter" can be estimated with confidence then one should restrict (a,b) to lie on it. In any case whenever the bias of the components is not large then $a + b = 1$ as a good rule of thumb.

More simulations were done while using the restriction that the weights lie on a hyperplane. The equation of this hyperplane was estimated from the generated data. These restrictions provided some improvement in the performance of the general composite estimator but more needs to be done to improve their robustness.

4. Conclusions

In the cases where the general composite estimator is robust it's use should certainly be considered because of it's simplicity compared to the more usual methods of adjusting each component of a composite estimator for bias (1-8). These bias adjustments in the general composite estimator are unnecessary as they are done automatically by the weights.

One measure of robustness (in the n - component case) is the second derivative of M(Z) with respect to a_i evaluated at a_i^* for $i=1,2,\ldots,n$. These values indicate how flat the mean square error surface is at the point $(a^*_1, a^*_2, \ldots, a^*_n)$. A short computation shows that these values are $E(X_i^2)$ for $i=1,2,\ldots,n$. Thus when $E(X_i^2)$ are small for $i=1,2,\ldots,n$ then the general composite estimator should be considered.

I would like to express my thanks to Alfreda Reeves who typed the final paper.

BIBLIOGRAPHY

1. Wolter K.M., "Composite Estimation in Finite Populations", JASA, Sept. 1979.
2. Woodruff, R.S., "The Use of Rotating Samples in the Census Bureau's Monthly Surveys", JASA, June 1963.
3. Eckler, A.R., "Rotation Sampling", JRSS, 1955.
4. Patterson, H.D., "Sampling on Successive Occasions with Partial Replacement of Units", JRSS, 1950.
5. Cochran, G.W., "Sampling Techniques" (second edition), Wiley.
6. Raj, D., "Sampling Theory", McGraw-Hill.
7. Hansen, M.H., Hurwitz, W.N., Madow, W.J., (1953)"Sample Survey Methods and Theory." John Wiley & Sons, Inc. New York.
8. Loeve, Michel, (1963), "Probability Theory", Van Nostrand Reinhold Company.
9. Marks, Harry M.; Woodruff, Stephen M., (1979), "Allocation Using Empirical Bayes Estimators", American Statistical Association 1979 Proceedings of Section on Survey Research Methods.
10. Woodruff, Stephen M., (1980), "A Class of Ratio Composite Estimators", American Statistical Association 1980 Proceedings of Section on Survey Research Methods.
11. Efron, Bradley; Morris, Carl (1975), "Data Analysis Using Stein's Estimator and its Generalizations", JASA, June 1975.