MIXED DESIGN BASED AND MODEL BASED STRATEGIES IN SURVEY SAMPLING

Stephen M. Woodruff, Bureau of Labor Statistics

I. INTRODUCTION

This paper examines combinations of model based and design based strategies for estimating population totals where a model holds for an unknown subset of the population and the model properties for the remainder of the population are unknown to the statistician. The estimator considered is a composite of the best linear unbiased estimator and the expansion estimator. The sample design is a mixture of a purposive sampling distribution (optimal if the model holds for the entire population) and the simple random sampling distribution. The results of this paper suggest strategies which may be useful in certain cases of model failure and provide information on robustness of certain model based strategies. This work compares model based strategies with design based strategies within a more general framework which contains both sets of strategies as special cases. The purpose of this generalization or mixing of strategies is to see when a mixture of both may be a feasible alternative to either alone.

The motivation for this paper is the desire to formulate a sampling problem and mixed strategy which may help answer some of the current questions (Sarndal[1], Lindley[2], Hansen, Madow[4] and Tepping[3 4]) about model based versus design based procedures. These questions mainly concern the robustness of model based procedures. Although empirical evidence suggested that the terminology "Robust Procedures" is an appropriate synonym for model based procedures, logically rigorous demonstrations of this are meager.

A strategy is a pair consisting of a sampling scheme and an estimator (or predictor). Sarndal[1] mentions a homogeneous model where the best strategy is the sample mean from any sample. The method of sample selection is not important. Thus, for this case, it won't hurt to use simple random sampling and the Horvitz-Thompson estimator, a pure design based strategy. The results presented here allow one to make the stronger statement that given the right circumstances, then this design based strategy is best. Not only will it not hurt to use simple random sampling, but simple random sampling minimizes expected mean square error; this is squared deviation averaged over both the model and sampling scheme. The terms model based and design based follow the usage of Cassel, Sarndal and Wretman in their book and papers. The term "design based" refers to that set of sampling and estimation procedures presented in "Sampling Theory," by Des Raj[6]; "Sampling Techniques," by William Cochran[5]; and "Sample Survey Methods and Theory," by Hansen, Hurwitz and Madow[7], for example. The term "model based" refers to the superpopulation procedures advocated by R. Royall, K. Brewer, Scott, Ho[8 9 10], etcetera.

The problem is to estimate a population total from a sample. Let $\vec{y} = (y_1, y_2, \ldots y_N)$ denote the vector of population values who's sum is to be estimated. The model based approach to sampling and estimation assumes this vector is a realization of a vector valued random variable $\vec{Y} = (Y_1, Y_2, \ldots Y_N)$. The distribution of Y is denoted as ζ. A model is a set of conditions on ζ. A given model often implies a single best sample and best estimator (or predictor) among a particular class of estimators.

In design based procedures, ζ, specifies a single point in R^N with probability one. The stochastic nature of the problem is derived solely through a sampling distribution, P, which is placed on the power set of $\{y_1, y_2, \ldots y_N\}$. The term "design" refers to the characteristics of P.

The criterion by which strategies are judged is expected mean square error denoted $\varepsilon E(T(S) - Y)^2$.

ε denotes expectation with respect to ζ, E denotes expectation with respect to P.

T(S) is the real valued predictor of Y which is distributed according to both ζ and P.

s denotes the sample outcome of P and $Y = \sum_{i=1}^{N} Y_i$

When no superpopulation is assumed (ζ is a single point in R^N with probability one) or no model for ζ can be assumed, then for any sampling design, P, such that $\pi_i > 0$ for all i = 1, 2, ... N (π_i = probability of inclusion in the sample for the i^{TH} universe member) the Horvitz-Thompson estimator is admissible in the set of all P-unbiased estimators for $y = \sum_{i=1}^{N} y_i$ and it is the unique hyperadmissible estimator in this class (Hanurav[11]). When no information is available about the universe, then simple random sampling seems justifiable. Indeed, if nothing is known about ξ, then inferences about the universe, based on a sample, can only be legitimately made via the sampling distribution.

Although a model will often imply a specific estimator and sampling scheme, what kind of a strategy is best when the model holds for only some unknown subset containing a proportion, ϱ, of the universe and for the rest of the universe, no information is available? This paper applies mixed strategies to this problem. The estimators that are considered are of the form.

$$T = (1-\alpha)H + \alpha M \qquad (1.1)$$

where: H denotes the Horvitz-Thompson estimator
M denotes the model based estimator and
α is a real number such the $0 \le \alpha \le 1$

The sampling scheme approximates a mixture of sampling distributions. It is a function of a real valued parameter λ which varies between zero and one. When λ=0, the optimal model based sampling scheme is used and when λ=1 simple random sampling is used. For λ strictly between zero and one a mixed sampling scheme is used. The sampling scheme is roughly a continuous function of λ. For example, this means that as $\lambda \to 1$, the sampling distribution approaches the simple random sampling distribution.

The set of strategies considered can be characterized by the set of ordered pairs (α,λ) in the unit square. For example, if $(\alpha,\lambda) = (0,1)$ then a pure design based strategy is implied. The problem is to find the ordered pairs (α,λ) that minimizes expected mean square error in a given situation.

The term "chaos" will be used to denote the unknown subset of the universe for which no information is available. When this chaotic portion of the universe is either small or well behaved, then pure model based procedures (MBP's) do quite well. It is shown that, for the type of model failure which chaos inflicts upon this estimation problem, model based procedures can be very robust.

Another interesting conclusion is that in many cases these strategies should not be mixed. That is, the optimum strategy is either pure MBP or pure design based procedures (DBP) depending on Q. Thus, the optimum strategy may be nearly a step function of Q.

Analytically, this means that there exists some number a, $0 < \underline{a} < \underline{1}$ such that for $Q > a$, $(\alpha_0 \lambda_0) \doteq (1,0)$ and for $Q \leq a, (\alpha_0 \lambda_0) \doteq (0,1)$ where (α_0, λ_0) denotes the pair (α, λ) which minimize expected mean square error of T.

This implies that balanced sampling (which approximates simple random sampling) and the BLUE (best linear ζ-unbiased estimator), may often not be a good strategy.

II. DESCRIPTION OF THE MODEL AND THE MIXED STRATEGIES

The set $\{Y_1, Y_2 \dots Y_N\}$ are assumed to be uncorrelated random variables such that for each i:

$$P(Y_i = \mu_i) = 1 - Q \text{ and}$$
$$P(Y_i \sim D(\beta x_i, \sigma x_i)) = Q$$

where $\{\mu_1, \mu_2, \dots \mu_N\}$ and $\{x_1, x_2, \dots x_N\}$ are constants. The set of x's are known to the statistician and the set of μ's are unknown. $D(\beta x_i, \sigma x_i)$ denotes the distribution of a random variable with mean, βx_i, and standard error, σx_i where β and σ are unknown constants.

Thus, either $y_i = \mu_i$ or $y_i = \beta x_i + \varepsilon_i$ where ε_i is the outcome of a random variable with mean zero and variance $\sigma^2 x_i^2$. This implies that for each i:

$$\varepsilon(Y_i) = (1-Q)\mu_i + Q \beta x_i$$
$$\gamma(Y_i) = Q\sigma^2 x_i^2 + Q(1-Q)(\mu_i - \beta x_i)^2$$
$$\varepsilon(Y_i^2) = (1-Q)\mu_i^2 + Q(\sigma^2 x_i^2 + \beta^2 x_i^2)$$

where ε and γ denote expectation and variance respectively with respect to ξ, the distribution of the vector $(Y_1 \dots Y_N)$ described above.

The problem is to estimate $y = \sum_{i=1}^{N} y_i$, given $\{x_1, x_2 \dots x_N\}$ and a sample of size n from $\{y_1, y_2, \dots, y_N\}$. The form of the estimator and the sampling scheme are to be choosen from the set of mixed strategies so that expected mean square error is minimized.

If $Q=1$ the linear model holds and the best linear unbiased estimator (BLUE) is:

$$\hat{\beta}\Big(\sum_{i\varepsilon S^c} x_i\Big) + \Big(\sum_{i\varepsilon S} y_i\Big) \qquad (2.1)$$

where $\hat{\beta} = (1/n) \cdot \sum_{i\varepsilon S} y_i/x_i$

where S denotes the sample of size n and S^c is its compliment in $\{y_1, y_2 \dots y_N\}$. This is BLUE independent of how the sample is selected but the sample of size n, which corresponds to the n largest x's, minimizes expected mean square error (Brewer, Royall). Therefore, this model implies both a unique best linear estimator and unique best sample.

If $Q=0$ then the auxilary variables $\{X_1, X_2 \dots X_N\}$ provide no information about the $\{Y_1, Y_2 \dots Y_N\}$. This situation is referred to as complete chaos. $y_i = \mu_i$ for all i and nothing is known about the set $\{\mu_1, \mu_2, \dots, \mu_N\}$. It was stated in the introduction that in this case there is strong motivation for using the Horvitz-Thompson estimator and simple random sampling.

For simple random sampling, the Horvitz-Thompson estimator is $N\bar{y}_n$ where $\bar{y}_n = \frac{1}{n} \sum_{i\varepsilon S} y_i$.

When Q is strictly between zero and one then an estimator of the form (1.1) will be used. The sampling scheme will be a compromise between a simple random sample of size n and that sample which corresponds to the n largest x-values.

This sampling scheme, which is a function of a parameter λ such that $0 \leq \lambda \leq 1$ is achieved by using a stratified sampling scheme where the sample allocation and strata boundaries are functions of λ. Without loss of generality, it is assumed that the x's are increasing functions of their subscripts.

$$x_1 \leq x_2 \leq x_3 \leq \dots \leq x_N.$$

The sampling scheme will consist of 2 strata. $[\lambda n]$ will be sampled using simple random sampling without replacement from the units with the smallest $N-(n-[\lambda n])$ x-values. In the strata consisting of the rest of the units everything is sampled. The function $[\cdot]$ is defined as:

$[a]$ = largest integer less than or equal to a.

Within each of the two strata, the sample is a simple random sample of the allocated size. When $\lambda < 1/n$, then the sample consists of the n units with the largest X values, $\{x_{N-n+1}, x_{N-n+2} \dots x_N\}$. This is the optimum sampling strategy when the linear model holds for the entire set, $\{Y_1, Y_2 \dots Y_N\}$. When $\lambda=1$, then the sampling scheme reduces to a simple random sample of size n from $\{y_1 \dots Y_N\}$.

The estimator (predictor) (1.1) is

$$T(S) = \alpha M + (1-\alpha)H$$

where $H = (N-(n-n(s)))(1/n(s))Y_{\mathcal{S}} + Y_\ell$

and $M = \hat{\beta}\Big(\sum_{i\varepsilon S^c} x_i\Big) + \Big(\sum_{i\varepsilon S} y_i\Big)$

where $n(\mathcal{S})$ = sample size in the strata of units with small x-values.

$Y_{\mathcal{S}} = \sum_{i\varepsilon \mathcal{S}} Y_i$, $\mathcal{S}$ = sample of units from the stratum with small x-values

$Y_\ell = \sum_{i\varepsilon \ell} Y_i$, ℓ = set of $n-n(\mathcal{S})$ units with largest x-values

Now recall that it is desired to choose (α,λ) such that $\varepsilon E(T(S)-Y)^2$, the expected mean square error is minimized. The solution ,

(α_0, λ_0), is a function of the μ's, the x's, β, σ^2 and Q. The rest of this paper is devoted to the properties of this function.

The particular linear model, $D(\beta x_i, \sigma x_i)$, was chosen so that the double expectation could be evaluated without the need of linearizing approximations. For example, if $V(Y_i) \sim x_i$ instead of $V(Y_i) \sim x_i^2$ then a Taylor series type approximation would be necessary in order to evalute the expectation $\varepsilon E(T(S)-Y)^2$, with respect to P, the sampling distribution. Therefore, except for rounding error, the calculations presented here are exact.

An exact algebraic solution expressing (α_0, λ_0) as a function of $\{\mu_1, \mu_2 \dots \mu_N, x_1, x_2 \dots x_N, \beta, \sigma^2, Q\}$ is extremely tedious to write out and probably quite uninformative. This obstacle was surmounted via computer by evaluating $\varepsilon E(T(S)-Y)^2$ at a sufficiently dense set of points containing the unit square $\{(\alpha,\lambda): \alpha = -.2 + (.05)i, \lambda = k/8$ for $i=0, 1, 2, \dots, 28, k=1, 2, \dots 8\}$. This solution allows one not only to find $(\alpha o, \lambda o)$ but it also shows the behavior of $\varepsilon E(T(S)-Y)^2$ near $(\alpha o, \lambda o)$.

It is clear at this point that the motivation for solving this problem may be purely theoretical since $(\alpha o, \lambda o)$ will seldom be available to the survey sampler. It may still be of practical interest to know how MBP's compare to DBP's when they are considered in this fashion. The characteristics of the surface generated by $\varepsilon E(T(S)-Y)^2$ as a function of (α,λ) may also prove enlightening to practitioners.

The concept of model failure as formulated in this paper was designed as a means of constructing a more general set of strategies which contains both MBP's and DBP's as special cases. Perhaps there is some middle ground between the two extremes which employes the best features of both sets of procedures.

The problem of robustness in case of model failure is addressed by Royall and Herson. They introduced the concept of balanced sampling as a means to correct for model misspecification. Balancing reduces the bias portion of expected mean square error when the true underlying superpopulation follows a polynomial model the degree of which is greater than the model on which the BLUE is based. This deals with an all or nothing situation in the universe to be sampled. That is, what happens if an alternate model holds for all members of the population.

This paper explores a more generalized form of model failure. This degree of generality makes analysis difficult. The explicit expressions for the expected mean square errors are to long to be enlightening, and they have been omitted. Instead, the algebraic formula are left in the computer and only the actual population parameters and expected mean square errors are tabulated and graphed under a variety of conditions. These explicit expressions for mean square error are available from the author.

III. SOME TABULATION OF EXPECTED MEAN SQUARE ERROR

The tables in this section show the expected mean square errors for three estimators: the BLUE, the Horvitz-Thompson and the best composite of these two. These mean square errors are rounded to the nearest hundred and tabulated in hundreds.

The universe size is 25 and the sample size is 8. This sample of 8 units is allocated among 2 strata. The large stratum (corresponding to large x-valued units) is the certainty stratum. The sample size in this stratum varies from zero to seven as λ varies from 1 to 1/8. When $\lambda=1$ (sample size in large stratum is zero) then a simple random sample of size 8 is chosen from among the 25 units. When $\lambda=3/8$, then the large stratum consists of the 5 largest units, which are sampled with certainty, and the small stratum consists of the 20 smallest units from which a simple random sample of size 3 is chosen.

Note that the case $\lambda=0$ is not considered. This is because the Horvitz-Thompson estimator is not defined in this case. Thus, the best sample for the BLUE when the model obtains is not considered. Nevertheless, when $\lambda=1/8$, then the sample is very nearly optimal for the BLUE given that the model holds for all 25 units (Q=1).

The values of the model parameters are as follows:

$\beta = 1 \qquad \sigma^2 = .1$

$x_i = (1/625) \cdot i^3 + .4999$ for $i = 1,2,\dots,25$

Tables 1 through 8 differ only in the way their μ-vectors were generated (chaos). The column labeled "optimal alpha" shows the best value for alpha for the given sampling scheme and Q. The next three numbers in the row are expected mean square errors in hundreds. "Min MSE" is the expected mean square error of the composite estimator $T = (1-\alpha)H + \alpha M$ where α is the "Optimal Alpha" given in that row. "MSE H" and "MSE M" are the expected mean square errors of the Horvitz-Thompson and Model based estimators respectively.

Let the set of random variables $\{Z_{ij}$: i = 1,2 and j=1,2,3 ... 25$\}$ be i.i.d. uniform $[0,1]$. In table 1, each μ_i, i=1,2, ... 25 was generated as follows:

$$\mu_i = I_1(Z_{1i}) \cdot 30 \cdot Z_{2i} + I_2(Z_{1i}) \cdot 14 + I_3(Z_{1i}) \cdot (5x_i + \sqrt{x_i}(Z_{2i} - .5))$$

where I_k is the indicator function on the interval $[(k-1)/3, k/3)$. The expected mean square error of T, H and M and the optimal alpha are conditional on the set of $\{\mu_i\}$ so generated.

Table 1 shows the optimal strategy (α_0, λ_0) for a given Q is as follows:

Q=	(α_0, λ_0) =
.67	(.95, 1/8)
.33	(.02, 7/8)
.0	(.01, 7/8)

When Q = .67, a fair degree of model failure, the MBP is robust (i.e., MIN MSE $\doteq$ MSE M). When $Q \leq .33$, then DBP's should be used. Note that

TABLE 1

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.95	33	167	33
		0.33	0.79	70	229	80
		0.00	0.61	91	217	143
6	2	0.67	0.81	37	89	40
		0.33	0.69	56	117	68
		0.00	0.56	51	110	86
5	3	0.67	0.50	44	62	63
		0.33	0.42	57	79	100
		0.00	0.38	47	73	115
4	4	0.67	0.22	44	49	111
		0.33	0.15	55	60	213
		0.00	0.14	48	55	308
3	5	0.67	0.11	41	43	192
		0.33	0.06	48	49	430
		0.00	0.05	42	44	717
2	6	0.67	0.07	41	42	307
		0.33	0.04	46	47	747
		0.00	0.04	40	42	1323
1	7	0.67	0.04	40	41	501
		0.33	0.02	42	43	1340
		0.00	0.01	38	39	2520
0	8	0.67	0.15	128	143	616
		0.33	0.11	214	236	1535
		0.00	0.10	284	317	2759

(α_0, λ_0) is either close to (1,0) or (0,1). That is, the best strategy is nearly H and simple random sampling or M and a sample of the largest x-valued units.

For table 2 through 7, the set of $\{\mu_i\}$ are given as follows:

Table	μ_i (i = 1,2, ... 25)
2	$30 \cdot Z_{1i}$
3	$5x_i + x_i(Z_{1i} - .5)$
4	$x_i + x_i(Z_{1i} - .5) + 4$
5	$x_i^{3/2} + x_i(Z_{1i} - .5)$
6	$x_i^{1.7} + x_i(z_{1i} - .5)$
7	$5 + Z_{1i}$

TABLE 2

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.97	30	133	30
		0.33	0.78	69	182	78
		0.00	0.57	94	180	144
6	2	0.67	0.84	35	72	36
		0.33	0.66	56	95	67
		0.00	0.50	60	92	93
5	3	0.67	0.47	40	52	55
		0.33	0.35	55	65	90
		0.00	0.31	51	62	106
4	4	0.67	0.18	39	42	94
		0.33	0.09	48	50	177
		0.00	0.07	45	46	250
3	5	0.67	0.08	36	37	161
		0.33	0.02	41	42	348
		0.00	0.02	37	37	562
2	6	0.67	0.06	37	38	256
		0.33	0.02	41	41	600
		0.00	0.02	37	37	1036
1	7	0.67	0.04	37	37	419
		0.33	0.01	38	38	1086
		0.00	0.00	34	34	2005
0	8	0.67	0.03	39	39	622
		0.33	0.01	38	38	1687
		0.00	0.01	33	33	3199

TABLE 3

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	1.00	23	330	23
		0.33	1.00	23	533	23
		0.00	1.00	0	639	0
6	2	0.67	0.99	31	241	31
		0.33	1.00	31	391	31
		0.00	1.00	0	476	0
5	3	0.67	0.99	40	226	41
		0.33	1.00	40	368	40
		0.00	1.00	0	451	0
4	4	0.67	0.99	52	230	52
		0.33	0.99	52	376	52
		0.00	1.00	1	461	1
3	5	0.67	0.99	67	254	67
		0.33	1.00	67	417	67
		0.00	1.01	1	514	1
2	6	0.67	1.01	87	300	87
		0.33	1.01	87	497	87
		0.00	1.02	1	619	2
1	7	0.67	0.99	108	325	108
		0.33	1.00	108	537	108
		0.00	1.01	2	666	2
0	8	0.67	1.01	138	390	138
		0.33	1.02	137	648	138
		0.00	1.03	3	811	3

For table 8:

$$\mu_i = I_1(Z_{1i})\cdot(3x_i + \sqrt{x}_i(Z_{2i} - .5)) + I_2(Z_{1i})\cdot(x_i + 4 + \sqrt{x}_i(Z_{2i} - .5)) + I_3(Z_{1i})\cdot(15Z_{2i} + 17)$$

Table 2 shows how T, with the optimal alpha, stabilizes expected mean square error (MIN MSE) in spite of the degree of model failure. Note also how MSE H decreases as MSE M increases (i.e., sampling scheme approaches simple random sampling). When Q = 0, a simple random sample would almost certainly lead one to reject the model and use H. If, as in the other extreme, the 7 largest units are sampled with certainty,

TABLE 4

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.93	4	39	4
		0.33	0.80	6	37	8
		0.00	0.61	5	24	13
6	2	0.67	0.82	6	27	7
		0.33	0.70	6	26	10
		0.00	0.60	2	20	10
5	3	0.67	0.67	8	24	12
		0.33	0.54	8	23	19
		0.00	0.47	3	18	23
4	4	0.67	0.52	11	24	22
		0.33	0.37	11	22	43
		0.00	0.30	6	17	65
3	5	0.67	0.41	14	26	37
		0.33	0.27	14	24	83
		0.00	0.22	10	20	140
2	6	0.67	0.37	18	32	57
		0.33	0.25	21	33	135
		0.00	0.20	18	32	235
1	7	0.67	0.28	22	33	94
		0.33	0.17	23	31	244
		0.00	0.12	20	28	454
0	8	0.67	0.26	28	40	134
		0.33	0.16	31	42	358
		0.00	0.12	30	41	676

TABLE 5

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.88	17	110	18
		0.33	0.82	37	180	44
		0.00	0.76	60	241	77
6	2	0.67	0.89	18	94	19
		0.33	0.85	33	157	37
		0.00	0.82	47	212	55
5	3	0.67	0.93	20	101	21
		0.33	0.93	31	170	31
		0.00	0.95	32	230	32
4	4	0.67	0.97	24	115	24
		0.33	1.03	29	196	29
		0.00	1.10	13	265	15
3	5	0.67	1.04	31	146	31
		0.33	1.12	32	250	35
		0.00	1.23	1	338	13
2	6	0.67	1.10	46	198	47
		0.33	1.14	64	345	68
		0.00	1.17	57	469	66
1	7	0.67	1.11	66	236	68
		0.33	1.06	122	410	123
		0.00	1.00	167	555	167
0	8	0.67	1.08	115	321	116
		0.33	0.88	263	564	268
		0.00	0.73	409	765	459

it may not be so clear that the model fails but in this case MSE M is not vastly greater than MIN MSE. Thus, a little post sampling data juggling would lead one to an estimator similar to the optimal T.

Table 3 shows that a combined strategy in the case of model failure of this type is of little help. Regardless of the degree of model failure, a pure model based strategy is best. A less dramatic, but similar result, is seen in table 4 (model failure in the form of a small y-intercept term).

Tables 5 and 6 show what happens in the case when model failure is an upward opening curve. In both cases, balancing improves the model

TABLE 6

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.78	63	249	77
		0.33	0.70	142	438	198
		0.00	0.63	229	600	362
6	2	0.67	0.78	68	225	81
		0.33	0.72	135	400	174
		0.00	0.67	197	549	281
5	3	0.67	0.83	79	257	86
		0.33	0.83	131	459	146
		0.00	0.82	159	629	181
4	4	0.67	0.91	94	316	97
		0.33	0.97	124	567	125
		0.00	1.04	85	775	86
3	5	0.67	1.01	120	421	120
		0.33	1.11	128	757	134
		0.00	1.25	2	1033	44
2	6	0.67	1.09	173	589	176
		0.33	1.15	227	1064	241
		0.00	1.23	156	1455	200
1	7	0.67	1.10	266	761	270
		0.33	1.03	486	1373	487
		0.00	0.97	654	1867	655
0	8	0.67	1.05	468	1067	469
		0.33	0.81	1045	1930	1091
		0.00	0.66	1589	2626	1871

TABLE 7

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	0.84	6	25	7
		0.33	0.46	9	14	16
		0.00	0.02	0	0	29
6	2	0.67	0.66	7	16	10
		0.33	0.34	6	9	15
		0.00	0.03	0	0	19
5	3	0.67	0.46	9	15	17
		0.33	0.19	6	8	30
		0.00	0.02	0	0	39
4	4	0.67	0.32	11	15	32
		0.33	0.10	7	8	66
		0.00	0.01	0	0	104
3	5	0.67	0.23	13	16	56
		0.33	0.06	8	8	133
		0.00	0.00	0	0	233
2	6	0.67	0.17	15	19	93
		0.33	0.05	9	9	241
		0.00	0.00	0	0	448
1	7	0.67	0.14	18	21	147
		0.33	0.04	10	11	407
		0.00	0.00	0	0	783
0	8	0.67	0.12	21	25	222
		0.33	0.03	12	13	642
		0.00	0.00	0	0	1265

based estimator and the optimal alpha stays generally close to 1.0. The combined strategy provides a dramatic improvement only in the case $Q = 0$ and $\alpha = 1.25$ (negative weight on H).

In table 7 model failure takes the form of a constant term plus small shock. Recall that these numbers are expected mean square errors in 100's rounded to the nearest whole number, thus, the zeros represent expected mean square errors less than 50. When $Q = .67$, the model based strategy gives good results, otherwise, the Horvitz-Thompson estimator is better.

Table 8 is similar to table 1 in that for $Q = .67$ the model based strategy is best and for $Q = .33$ or 0, the design based strategy is best.

TABLE 8

STRATUM SIZE LARGE	SMALL	Q	OPTIMAL ALPHA	MIN MSE	MSE H	MSE M
7	1	0.67	1.04	31	138	31
		0.33	0.95	64	202	64
		0.00	0.82	93	225	99
6	2	0.67	0.68	51	83	59
		0.33	0.55	84	118	107
		0.00	0.45	98	130	146
5	3	0.67	0.29	56	64	101
		0.33	0.20	81	88	184
		0.00	0.17	88	95	250
4	4	0.67	0.11	52	54	168
		0.33	0.06	70	71	322
		0.00	0.05	74	75	463
3	5	0.67	0.05	48	48	267
		0.33	0.02	60	61	546
		0.00	0.02	62	62	838
2	6	0.67	0.04	48	48	404
		0.33	0.02	59	59	865
		0.00	0.02	60	60	1384
1	7	0.67	0.02	46	46	606
		0.33	0.01	53	53	1373
		0.00	0.01	52	52	2306
0	8	0.67	0.05	77	79	814
		0.33	0.04	108	111	1818
		0.00	0.04	127	132	3016

IV. CONCLUSIONS

Judging by the results of this study, model based procedures are quite efficient and robust even in many cases of severe model failure. Thus, for the cases considered here, the inferences based on the working model are still good when the data are generated via a very different process. The Horvitz-Thompson estimator, as it is used in conjunction with the BLUE in this paper, does add both robustness and stability to the inference. It also provides a yardstick by which to judge the BLUE and its purposive sampling scheme (largest x-valued units). For reasons that are both political and operational, a fully purposive sampling plan should rarely be used. Some degree of randomization is necessary to avoid designer bias, and robustness is often improved by randomization.

Robustness is the statistical analog to the mathematical concept of continuity . The structure needed to make this statement precise is contained in "Robust Statistics" by Peter Huber. If the working model, which is used to design and analyze survey data, is "close" to the actual process by which the data was generated, then the inference based on the working model must be nearly as good as the inference based on the actual process that generated the data. If this is not the case for a given set ofstrategies, then they should be abandoned in favor of strategies which satisfy this condition.

If $\vec{Y}$ is distributed according to ζ and the statistician hypothesizes ζ_c as the distribution of $\vec{Y}$, then let T and T_c denote the BLUE's under ζ and ζ_c respectively. Let P and P_c denote the optimal sampling distributions under ζ and ζ_c respectively, then a desireable property of the strategy (T_c, P_c) is that $\varepsilon E_c(T_c - Y)^2 \to \varepsilon E(T - Y)^2$ when $\zeta_c \to \zeta$ (in distribution) where E_c denotes expectation with respect to P_c. Using this as the definition of robustness, it can be shown using some convergence results from measure theory that model based procedures are robust with respect to much larger varieties of model failure than the approach suggested by, Huber or Hampel, for example, which employs Frechet or Gâteau derivatives.

A large amount of empirical evidence accumulated over recent years by the advocates of model based procedures (prediction theory, robust estimation), strongly supports the conjecture that these procedures are indeed quite robust. This paper adds to this increasing body of evidence. It also suggests an alternative set of strategies to further enhance both model based and design based procedures by combining them. These combined strategies have several advantages over either pure model based procedures or pure design based procedures.

In most cases, this combined strategy stabilizes mean square error as the sampling scheme varies between simple random sampling and purposive sampling. This kind of sample insensitivity is certainly desireable. In some cases, neither Horvitz-Thompson nor the BLUE give very satisfactory results when compared to the best composite of the two. Often the combined strategy highlights the strengths or weaknesses of MBP's versus DBP's. This is the case when the optimal α is either zero or one; that is, when the best combination of the Horvitz-Thompson and the BLUE is either the Horvitz-Thompson estimator or the BLUE.

In conclusion, this study has increased my confidence in model based procedures. If the scatter plot of the sample data pairs (x_i, μ_i), appear, even vaguely, to lie around a straight line through the origin, then a BLUE should be considered. If enough auxiliary information is at hand, then a combined strategy of the type studied in this document will further enhance the BLUE both with respect to increased robustness and decreased mean square error.

REFERENCES

[1]Sarndal, Carl-Erik, (1978) "Design-based and Model-based Inference in Survey Sampling," Scandanavian Journal of Statistics, 5:27-52, 1978.

[2]Lindley, DV, (1978) "Discussion of Design-based and Model-based Inference in Survey Sampling," Scandanavian Journal of Statistics 5:27-52, 1978.

[3]Hansen, M.H., Madow, W.G., (1978) "Estimation and Inferences from Sample Surveys-Some Comments on Recent Developments," Section on Survey Research Methods, Proceedings of the American Statistical Society.

[4]Hansen, M.H., Madow, W.G., Tepping, J.T., (1983) "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys," JASA Vol 78, Number 384, December 1983.

[5]Cochran, W.G., (1963) "Sampling Techniques," 2nd Edition, Wiley.

[6]Des Raj., (1968) "Sampling Theory," McGraw-Hill, Inc.

[7]Hansen, M.H., Horvitz, W.H., Madow, W.G., (1953) "Sample Survey Methods and Theory," Wiley.

[8]Royall, R.M., (1970) "On Finite Population Sampling under Certain Linear Regression Models," Biometrika, 57, 377-387.

[9]Royall, R.M., (1971) "Linear Regression Models in Finite Population Sampling Theory," in Foundations of Statistical Inference (V.P. Godambe and D.A. Sprott, eds) Holt, Rinehart and Winston, Toronto.

[10]Scott, A.J., Brewer, K.R.W., Ho, E.W.H., (1978) "Finite Population Sampling Robust Estimation," JASA 1978, Vol 73, Number 362.

[11]Hanurav, T.V., (1968) "Hyperadmissibility and Optimum Estimators for Sampling Finite Populations," Annals of Mathematical Statistics, 39,