EVALUATION OF SMALL AREA ESTIMATION TECHNIQUES FOR THE CANADIAN LABOUR FORCE SURVEY

J.D. Drew, M.P. Singh and G.H. Choudhry
Statistics Canada

Estimates from sample surveys are sometimes required for domains whose boundaries do not coincide with those of design strata. Taking the Canadian Labour Force Survey as an example of a survey utilizing a clustered sample design, some alternative small area estimation techniques available in the literature are evaluated empirically including synthetic, post-stratified domain and composite estimators. A sample dependent estimator is also proposed and evaluated.

1. Introduction

With increasing emphasis on planning and administering social and fiscal programs at local levels, there has been demand for more and good quality data at these levels on a wide range of subject matters.

A comprehensive review of existing small area (domain) estimation techniques along with their limitations is given by Purcell and Kish (1979). From the research done to date it is clear that there is not a unique best solution to the small area estimation problem. The choice of a particular method for small area estimation will depend on the data needs and on the richness and availability of data sources, which differ from country to country, and within countries from one subject matter to another. We shall adopt the following classification of domains suggested by Purcell and Kish (1979), which we consider important since the form of the estimator as well as its efficiency would depend on the particular type of application:

(a) Planned domains - for which separate samples have been planned, designed, and selected; e.g. in the Canadian context, provinces or economic planning regions within provinces.

(b) Cross Classes - which cut across the sample design and the sample units (may also be referred to as characteristic domains); e.g., age/sex, occupation, industry.

(c) Unplanned domains - that have not been distinguished at the time of sample design and thus may cut across the design strata; e.g., in the Canadian context Federal Electoral Districts, and Census Divisions or subdivisions.

It should be noted that both types (a) and (c) refer to areal domains.

As pointed out by Purcell and Kish most of the developments in small area estimation techniques in the United States and elsewhere have concentrated on the domains of types (a) and (b). In Canada however, type (a) and (b) domains are not so problematic due to the type of design and the sizes of the national surveys, and the main emphasis has been on the data for the domains of type (c).

The estimators we consider in this article are geared to the Canadian LFS where the domains are unplanned domains (type c) and are of a size such that, had they been planned domains (type a), the reliability of regular unbiased survey estimates would be satisfactory without having to resort to small area estimation techniques.

The nature of the design in relation to the domains of interest also has an important role to play in the choice of an estimator. For instance, in the LFS, primary sampling units are small relative to the sizes of the domains of interest, whereas in the United States the sizes of primary sampling units for most of the large scale surveys are larger or comparable in size to the small areas for which the estimates are desired.

We evaluate estimators in the context of producing Census Division level estimates from the Labour Force Survey, using data from population census in an auxiliary fashion. In addition to synthetic estimators, we evaluate post-stratified domain estimators and composite estimators which are linear combinations of the synthetic and the post-stratified domain estimators. Brief historical background on each of these estimators is provided. Also we propose and evaluate a new estimator which we call a sample dependent estimator, which is of the same form as the composite estimator, except the weight given to the synthetic component is a decreasing function of the amount of sample falling into the domain. Efficiencies of the small area estimators relative to the direct (or simple domain) estimator for the characteristics employed and unemployed are obtained in an empirical (Monte Carlo) study in which the LFS design is simulated using census data.

2. Description of Estimation Procedures

Consider a finite population consisting of N units, (e.g. households or dwellings in household surveys), divided into L (areal) design strata labelled 1, 2, ..., h, ...L. If we denote by 'a' the set of units belonging to an unplanned areal domain (type c) of interest, then the parameter to be estimated is the total of an x-variable in the domain 'a', which we denote by ${}_aX$. Let a_h be the set of those units belonging to the domain which are in stratum h, then

$$a = \bigcup_{h=1}^{L} a_h. \tag{2.1}$$

In practice the domain 'a' will have a non-null intersection with a certain number of design strata and if we denote by $\tilde{h}$ the set of such strata, then we have

$$a = \bigcup_{h \varepsilon \tilde{h}} a_h. \tag{2.2}$$

The particular design under consideration follows a multi-stage clustered sample design which is self-weighting within each stratum with weight W_h for stratum h.

For a particular given sample we can obtain the quantities:

t_h = sample total of x-variate in stratum h, and ${}_at_h$ = sample total of x-variate in a_h for h=1, 2, ..L. Then the direct (or also referred to as design based or simple domain) estimator for the total of x-variate for those units in 'a' say ${}_a\hat{X}$, is given by:

$$ {}_a\hat{X} = \sum_{h \in \underset{\sim}{h}} W_h \cdot {}_at_h. \tag{2.3} $$

It should be noted that the direct estimator (2.3) does not utilize any auxiliary information-all it requires is the identification of those sampled units which belong to the domain. Due to the clustered nature of the design, the sample falling in the domain may on occasion be very small or non-existent, generally resulting in high variance for this estimator.

The other estimators in this section rely in different fashions on auxiliary information for a variable y, which is often taken as the count of persons by population sub-groups (defined on the basis of age/sex etc.) from a recent census. These estimators are:

1) Post-stratified domain
2) Synthetic
3) Composite
4) Sample dependent

Additionally estimators (2) - (4) rely to differing degrees on sample external to the domain.

For each of the above estimators, the adjustments based on the auxiliary information can be made either separately for each stratum intersecting the domain, or by applying an overall adjustment for all strata intersecting the domain. Thus the estimators will be further classified as <u>separate</u> or <u>combined</u> depending on the level at which the adjustment is made. These estimators are denoted by ${}_a\hat{X}_{uv}$, where u is the level of adjustment with values: u=s: separate

=c: combined

and v is the type of estimator taking the following values: v = p: post-stratified domain

= s: synthetic

= c: composite

= d: sample dependent

For example, ${}_a\hat{X}_{cs}$ denotes the <u>combined synthetic</u> estimator, etc.

2.1 Post-Stratified Domain Estimator

Define

Y_{hg} = total of the auxiliary y-variable for population sub-group g in stratum h, and

${}_aY_{hg}$ = total of the auxiliary y-variable for population sub-group g in $a_{\underset{\sim}{h}}$.

Further let ${}_a\tilde{Y}_{hg}$ be an unbiased estimate of ${}_aY_{hg}$ based on the auxiliary y-data for the sampled units only.

Then the separate post-stratified domain estimator is

$$ {}_a\hat{X}_{sp} = \sum_g \sum_{h \in \underset{\sim}{h}} (W_h \cdot {}_at_{hg}) \frac{{}_aY_{hg}}{{}_a\tilde{Y}_{hg}} \tag{2.4} $$

where ${}_at_{hg}$ is the sample total of the x-variate for population sub-group g in the intersection of domain 'a' and stratum h.

Similarly the combined post-stratified domain estimator is:

$$ {}_a\hat{X}_{cp} = \sum_g \sum_{h \in \underset{\sim}{h}} (W_h \cdot {}_at_{hg}) \frac{\sum_{h \in \underset{\sim}{h}} {}_aY_{hg}}{\sum_{h \in \underset{\sim}{h}} {}_a\tilde{Y}_{hg}} \tag{2.5} $$

The post-stratified domain estimator is unbiased except for the effect of ratio estimation bias.

Estimators of the above type have been considered earlier by Singh and Tessier (1976) with a different choice of post-strata.

2.2 Synthetic Estimators

We consider separate and combined synthetic estimators defined respectively as follows:

$$ {}_a\hat{X}_{ss} = \sum_g \sum_{h \in \underset{\sim}{h}} (W_h \cdot t_{hg}) \frac{{}_aY_{hg}}{Y_{hg}} \tag{2.6} $$

$$ {}_a\hat{X}_{cs} = \sum_g \sum_{h \in \underset{\sim}{h}} (W_h \cdot t_{hg}) \frac{\sum_{h \in \underset{\sim}{h}} {}_aY_{hg}}{\sum_{h \in \underset{\sim}{h}} Y_{hg}}, \tag{2.7} $$

where t_{hg} is the sample total for the x-variable for population sub-group g in stratum h.

The above synthetic estimator has been considered by Purcell and Linacre (1976) and by Ghangurde and Singh (1976, 1977, 1978). A different form of synthetic estimator was proposed earlier by the National Centre for Health Statistics (1968) and investigated by Gonzalez (1973, 1975) Gonzalez and Waksberg (1973), and Gonzalez and Hoza (1978).

The synthetic estimator will suffer from bias depending on the degree of departure from the assumption of homogeneity for the x-variate between the domain and the larger area, namely $\underset{\sim}{h}$, within sub-groups of the y-variable. We restricted the larger area to those strata which form part of the domain in the belief that such a choice would lead to less bias. In general however, $\underset{\sim}{h}$ need not be so restricted but may include other areas which are believed to satisfy the homogeneity assumption. Bias and mean square error of such estimators have been reported by some of the earlier referenced authors.

2.3 Composite Estimators

A composite estimator using the direct estimator and the synthetic estimator as the two components was suggested by Royall (1973) and others, and has been studied by Schaible (1978). Such an estimator minimizes the chances of extreme situations (both in terms of bias and mean square error) and therefore may be preferred over either of its components. Empirical evidence of the relative performances of synthetic and direct estimators are available from Gonzalez and Waksberg (1975), Schaible, Brock and Schnack (1977), and Ghangurde and Singh (1977). The composite estimator we consider is obtained by replacing the direct estimator (2.3) by the post-stratified domain estimator which may be slightly biased but is generally more efficient than the direct estimator.

The two types of composite estimators: namely, separate and combined are formed as linear combinations of corresponding post-stratified domain and synthetic estimators; viz,

$$ {}_a\hat{X}_{sc} = \alpha_1\, {}_a\hat{X}_{sp} + (1-\alpha_1)\, {}_a\hat{X}_{ss} \quad (2.8) $$

and

$$ {}_a\hat{X}_{cc} = \alpha_2\, {}_a\hat{X}_{cp} + (1-\alpha_2)\, {}_a\hat{X}_{cs} \quad (2.9) $$

The optimum values for α_1 and α_2 for minimum mse's are given by

$$ \alpha_1^* = \frac{\text{mse}[{}_a\hat{X}_{ss}] - E[{}_a\hat{X}_{ss} - {}_aX][{}_a\hat{X}_{sp} - {}_aX]}{\text{mse}[{}_a\hat{X}_{ss}] + \text{mse}[{}_a\hat{X}_{sp}] - 2E[{}_a\hat{X}_{ss} - {}_aX][{}_a\hat{X}_{sp} - {}_aX]} \quad (2.10) $$

and a similar expression for α_2^*.

2.4 Sample Dependent Estimators:

The sample dependent estimator due to Drew and Choudhry (1979) is a particular case of a composite estimator which depends on the outcome of the given sample, and is quite simple to compute. It is constructed using the result that the performance of the post-stratified domain estimator depends upon the proportion of the sample falling in the domain. If that proportion is 'reasonably large' then the sample dependent estimator is the same as the post-stratified domain estimator, otherwise it becomes a composite estimator with gradual increasing reliance (in the sense of increasing weight) on the synthetic estimator as the size of the sample in the domain decreases. Thus the separate sample dependant estimator is given by

$$ {}_a\hat{X}_{sd} = \sum_g \sum_{h\varepsilon \underset{\sim}{h}} [\delta_{hg} W_h \cdot {}_at_{hg} \frac{{}_aY_{hg}}{{}_a\tilde{Y}_{hg}} + (1-\delta_{hg}) W_h \cdot t_{hg} \frac{{}_aY_{hg}}{Y_{hg}}] \quad (2.11) $$

where

$$ \delta_{hg} = 1, \text{ if } {}_a\tilde{Y}_{hg}/{}_aY_{hg} \geq K_o, \quad = \frac{1}{K_o} \frac{{}_a\tilde{Y}_{hg}}{{}_aY_{hg}}, \text{ otherwise.} $$

Similarly the combined sample dependent estimator is given by

$$ {}_a\hat{X}_{cd} = \sum_g [\delta_g (\sum_{h\varepsilon \underset{\sim}{h}} W_h \cdot {}_at_{hg}) \frac{\sum_{h\varepsilon \underset{\sim}{h}} {}_aY_{hg}}{\sum_{h\varepsilon \underset{\sim}{h}} {}_a\tilde{Y}_{hg}} + (1-\delta_g)(\sum_{h\varepsilon \underset{\sim}{h}} W_h \cdot t_{hg}) \frac{\sum_{h\varepsilon \underset{\sim}{h}} {}_aY_{hg}}{\sum_{h\varepsilon \underset{\sim}{h}} Y_{hg}}] \quad (2.12) $$

where

$$ \delta_g = 1, \text{ if } \sum_{h\varepsilon \underset{\sim}{h}} {}_a\tilde{Y}_{hg} / \sum_{h\varepsilon \underset{\sim}{h}} {}_aY_{hg} \geq K_o $$

$$ = \frac{1}{K_o} \sum_{h\varepsilon \underset{\sim}{h}} {}_a\tilde{Y}_{hg} / \sum_{h\varepsilon \underset{\sim}{h}} {}_aY_{hg}, \text{ otherwise.} $$

The ratios ${}_a\tilde{Y}_{hg}/{}_aY_{hg}$ and $\sum_{h\varepsilon \underset{\sim}{h}} {}_a\tilde{Y}_{hg} / \sum_{h\varepsilon \underset{\sim}{h}} {}_aY_{hg}$

indicate the over- or under-representation of the population sub-groups at the individual stratum or domain level with respect to auxiliary information for the y-variable, conditional upon the selected sample.

The value of K_o may be appropriately chosen.

In this study the efficiency of sample dependent estimator has been investigated for two specific values of K_o namely 1.0 and 0.5.

It should be noted that Holt, Smith and Tomberlin (1979) and Sarndal (1981) have also proposed different estimators relying on synthetic and direct estimates where the weight attached to the direct component is dependent on the sample falling in the domain.

3. Description of the Empirical Study

3.1 Simulation of the LFS Design

The LFS follows a multi-stage area sampling design (see Platek and Singh, (1976). In simulating the LFS design two cases were examined: (I) the case where both the sample design and the auxiliary information are up-to-date, and (II) the case where both are out-of-date.

For (I), the sample design, the auxiliary information, and the study variables were all based on 1971 census data. For each replication in the Monte Carlo study, independent samples of primaries and secondaries were selected based on census population or dwelling counts, while the final stages of sampling were simulated by random samples of persons.

For (II), the primaries and secondaries were selected on the basis of 1971 census counts, however the sample of persons within secondaries was based on 1976 census results. The auxiliary information was based on the 1971 census data.

3.2 Choice of Population Sub-Groups

The estimators defined in section 2 utilize auxiliary information for population sub-groups. For the variables marital status, age and sex, the Automatic Interaction Detection (AID) procedure, due to Sonquist and Morgan (1964) was used on a sample of census data from across Canada to derive optimal population sub-groups, separately for each Labour Force characteristic. Results of the AID analysis showed that for unemployed, no population sub-groups accounted for more than 2% of the variation, while for the characteristics employed and not in Labour Force the following sub-groups accounted for approximately 25% of the variation: (i) age 15-16 and 65+; (ii) age 17-64, sex female; (iii) age 17-64, sex male. Further splitting of these sub-groups did not result in significant additional gains.

In addition to estimators based on the above population sub-groups, estimators based on total population 15+, and on dwelling counts were also considered.

3.3 Evaluation of Efficiency of Small Area Estimators

In the Monte Carlo study, we have considered 16 Census Divisions (CD's) in the province of Nova Scotia. Due to the multi-stage nature of the design and large number of domains in the study, the computational costs involved were high and it was decided to use only 100 replications.

If we let ${}_a\hat{X}_{m(r)}$ be the estimate of the total

$_aX$ (i.e. the total for the x-variable for the domain 'a'), for the r'th replicate, for small area estimation method m, then the average mean square error for the method m over the 16 domains in the study was calculated as:

$$\text{avg mse } (m) = \frac{1}{16} \sum_a \sum_{r=1}^{100} ({}_a\hat{X}_{m(r)} - {}_aX)^2/100. \quad (3.1)$$

The efficiency of the small area estimator (m) relative to the direct estimator, say method m_o was obtained as:

$$\text{Eff } (m \text{ vs } m_o) = \frac{\text{Avg mse } (m_o)}{\text{Avg mse } (m)}. \quad (3.2)$$

4. Analysis of Results

4.1 Efficiency considerations: Auxiliary Information up-to-date (i.e., case I)

Efficiencies of the four small area estimators are presented relative to the direct estimator in Table 1, for separate and combined levels of construction, and for each of the following auxiliary variables - dwellings, total population (15+), and population by age/sex groups. The following observations can be made:

(i) Separate vs Combined Estimator: The level of construction of estimators does not have much impact on the efficiencies of synthetic estimators for both the characteristics employed and unemployed. For the post-stratified domain estimator for employed, however, the combined form is approximately twice as efficient as the separate, due to the effect of the clustering in the sample design being more accentuated with the separate estimator.

Since the post-stratified domain estimator was less efficient in its separate form, a similar result was anticipated for the composite estimator and hence, only the combined composite estimator was considered. On the other hand, the separate form of the sample dependent estimator was found to rely slightly more on the synthetic component, leaving the efficiencies unaffected by the level of construction.

(ii) Effect of Auxiliary Information: The performance of population by age/sex as an auxiliary variable is uniformly superior, although only marginally so, to the total (15+) population for all four estimators using auxiliary information. Further, both these variables out-perform the dwelling count as an auxiliary variable.

(iii) Comparison among the estimators: For unemployed, performance of the composite estimator with optimum α_2* chosen for the characteristic unemployed is marginally superior to the other estimators irrespective of the level of construction and the choice of auxiliary variable. For employed, the sample dependent estimator shows marginal improvement over other estimators.

4.2 Efficiency Considerations: Auxiliary information out-of-date (i.e., case II)

As can be seen from Table (2), although for unemployed the use of small area estimation techniques showed larger gains relative to the direct estimator (than in the up-to-date case), considerably smaller gains were observed for employed. The later is likely due to the reduced correlation between the study variable and the auxiliary information as both the design and auxiliary information become out-of-date. Also in this case, the efficiency of the synthetic estimator is higher for both of the characteristics measured.

4.3 Consideration of Bias:

Given that the post-stratified domain estimator will generally have negligible bias, the bias of both the composite and sample dependent estimator would generally be smaller than that of the synthetic estimator, i.e. stemming from the degree of reliance on the synthetic component. Since the composition of the LFS frame and the Federal Electoral Districts were known for all of Canada in terms of both 1971 and 1976 census units, it was possible to compute exact biases of the synthetic estimators based on census data. The bias was computed when the auxiliary information (population 15+) was up-to-date and also when the auxiliary information was out-of-date. It was observed that the bias exceeded 10% in 13 and 19 (out of 279) FED's when the auxiliary information was up-to-date and out-of-date respectively. Further, in about half the instances for which the bias exceeded 10% for the up-to-date case, the bias also exceeded 10% for the later time period when the auxiliary information was out-of-date. This suggests that for domains with a known high bias at the time to which the auxiliary information refers, less use should be made of the synthetic estimator. However there is still the danger of bias in the synthetic estimator from those cases where the bias is low the auxiliary information is up-to-date but is large when the auxiliary information becomes out-of-date.

4.4 Efficiency vs Bias in Overall Choice of Estimator

The efficiencies of the synthetic, composite and sample dependent estimates were found to be comparable to each other. It was also observed that the synthetic estimator was generally highly biased. Since the sample dependent estimator makes use of the synthetic estimator whenever there is not 'sufficient' sample in the domain, its bias would depend upon the weight attached to the synthetic estimator component and this can be controlled by a proper choice of K_o.

For instance with K_o=1, the average reliance of the sample dependent estimator on the sample dependent component ranged from .01 for domains comprised primarily of complete strata, to .28 for a domain comprised entirely of partial strata.

Also as expected, the average reliance on the synthetic component for the smaller value of K_o (=0.5) was lower (e.g., .01 and .18 respectively for the above extremes). However as illustrated in Table 1, a trade-off between bias and efficiency is involved since lower choices of K_o also result in reduced efficiency. The above values of K_o provided a reasonable degree of confidence for the type of domains discussed here.

In general, however, other values of K_o may be chosen depending upon e.g. the size of the domain, sample size, strata sizes and their geographical configurations with respect to the domain.

4.5 Concluding Remarks

1) The use of population by age/sex fares uniformly better than the other auxiliary variables, although gains over total population (15+) are marginal.
2) The post-stratified domain estimator although more efficient as compared to the simple domain estimator, performs poorly as compared to the other three small area estimators investigated.
3) The combined composite estimator constructed as a linear combination of post-stratified and synthetic estimators is more efficient than either of its component estimators although only marginally so, as compared to the synthetic component, for optimum value of α. Its bias would depend upon the weight attached to the synthetic component since the bias of the post-stratified estimator would generally be negligible. Further, as the computation of the optimum α is quite involved, in practice only an estimated value of α may be used, resulting in a decrease in efficiency of this estimator.
4) Since the bias of the separate synthetic estimator is smaller than that of the combined synthetic estimator, the separate sample dependent estimator would result in smaller relative bias as compared to the combined sample dependent estimator. The bias of the separate post-stratified domain component can be controlled by collapsing those strata for which the intersection with the domain is very small. Thus considering all the three aspects, bias, mean square error and the computational complexities, the sample dependent estimator constructed at the stratum level using population by age and sex would seem to be a better choice over the other estimators examined. The choice of the value for K_o would depend upon several factors. In this case the value of K_o=1 proved to be efficient while affording protection against the bias of the synthetic estimator.

5. Future Direction of Investigation:

The study reported in this paper has focused on evaluation of certain small area estimation methods using only census and survey data, for unplanned domains (type c) for the LFS. Below we point to directions which future investigations in Canada might take.

Since the small area estimation methods for the unplanned domains have out-performed the unbiased design based estimates for comparable planned domains, it would be desirable to extend this investigation to certain small planned domains (type a) as well. Further work on development of methods of variance estimation to be used in practice for these estimators is also needed. Also the Structure Preserving Estimators (SPREE) suggested by Purcell and Kish (1980) should be evaluated.

More generally, for characteristics for which large scale surveys (such as the Labour Force Survey) are undertaken regularly, it would seem the short term demand for data for domains of the size of FED's or Census Divisions may be met through the use of refined estimation techniques (and pooling of estimates over a period of time) utilizing census and survey data alone. However, for meeting demands in the longer term and for other types of data based on smaller surveys and other types and sizes of domains, all three sources of data namely census, surveys and administrative files need to be fully exploited. Multi-variate linear regression estimators of the type considered by Ericksen (1974) and Gonzalez and Hoza (1978) using data from all three sources should be studied in detail for their bias, mean square error and the computational complexities.

Discussions with Mr. R. Platek have been beneficial in the finalization of this paper.

REFERENCES

Drew, J.D. and Choudhry, G.H. (1979), "Small Area Estimation," Technical Report, Census and Household Surveys Methods Division, Statistics Canada.

Ghangurde, P.D. and Singh, M.P. (1976), "Synthetic estimation in the LFS, "Technical Report, Household Surveys Development Division, Statistics Canada.

Ghangurde, P.D. and Singh, M.P. (1977), "Synthetic Estimates in periodic household surveys," Journal of Survey Methodology, Statistics Canada 3, 152-181.

Ghangurde, P.D. and Singh, M.P. (1978), "Evaluation of Efficiency of Synthetic Estimates," Proceedings of the American Statistical Association, Social Statistics Section, 53-61.

Gonzalez, M.E. (1973), "Use and Evaluation of Synthetic Estimates," Proceedings of the American Statistical Association, Social Statistics Section, 33-36.

Gonzalez, M.E. and Waksberg, J. (1973), "Estimation of the Error of Synthetic Estimates," paper presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

Gonzalez, M.E. (1975), "Small Area Estimation of Unemployment," Proceedings of the American Statistical Association, Social Statistics Section, 437-460.

Gonzalez, M.E. and Hoza, C. (1978), "Small Area Estimation with application to Unemployment and Housing Estimates, "Journal of the American Statistical Association 73, 7-15.

Holt, T., Smith, T.M.F. and Tomberlin, T.J. (1979), "A Model Based Approach to Estimation for Small Sub-groups of a Population," Journal of the American Statistical Association, 74, 405-410.

National Center for Health Statistics (1968), "Synthetic State Estimates of Disability," P.H.S. Publication No. 1759, U.S. Government Printing Office, Washington, D.C.

Platek, R. and Singh, M.P. (1976), "Methodology of the Canadian Labour Force Survey," Catalogue No. 71-526, Statistics Canada.

Purcell, N.J. and Linacre, S. (1976), "Techniques for the Estimation of Small Area Characteristics," paper presented at the 3rd Australian Statistical Conference, Melbourne, Australia.

Royall, R.M. (1973), "Discussion of two papers on Recent Developments in Estimation of Local Areas," Proceedings of the American Statistical Association, Social Statistics Section, 43-44.

Royall, R.M. (1978), "Prediction models in Small Area Estimation," NIDA Workshop on Synthetic Estimates, Princeton, N.J.

Sarndal, C.E. (1981), "When Robust Estimation is not an obvious answer: The case of the Synthetic Estimator versus Alternatives for Small Areas," Proceedings of the American Statistical Association, Survey Research Section.

Schaible, W.L., Brock, D.B. and Schnack, G.A. (1977), "An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics," Proceedings of the American Statistical Association, Social Statistics Section, 1017-1021.

Schaible, W.L. (1978), "Choosing Weights for Composite Estimators for Small Area Statistics," Proceedings of the American Statistical Association, Survey Research Section, 741-746.

Schaible, W.L. (1979), "A Composite Estimator for Small Area Statistics," In Synthetic Estimates for Small Areas (J. Steinberg, Ed.) National Institute on Drug Abuse Research Monograph No. 24, U.S. Government Printing Office, Washington, D.C., 36-53.

Singh, M.P. and Tessier, R. (1975), "Some Estimators for Domain Totals", Journal of The American Statistical Association 71, 322-325.

Sonquist, J.N. and Morgan, J.A. (1964), "The Detection of Interaction Effects," Monograph No. 35, Survey Research Centre, Institute for Social Research, University of Michigan.

Table (1): Efficiencies of Small Area Estimators Relative to Direct Estimator - Nova Scotia Census Divisions (Auxiliary data up-to-date).

Characteristic	Auxiliary Variable	Level of Construction	ESTIMATOR Post-Stratified Domain	Synthetic	Composite (α^*_2=0.223)	Sample Dependent K_o=0.5	Sample Dependent K_o=1.0
Employed	Dwelling	combined	4.58	10.17	10.92	9.17	10.42
	Population	combined	4.92	10.75	10.58	10.50	11.67
	Population by age/sex	combined	5.08	10.83	11.25	11.17	12.25
	Dwelling	separate	2.75	10.50	-	9.58	10.50
	Population	separate	2.83	10.92	-	10.58	11.42
	Population by age/sex	separate	2.83	11.00	-	11.00	11.75
Unemployed	Dwelling	combined	1.33	1.70	1.75	1.40	1.55
	Population	combined	1.36	1.70	1.75	1.43	1.58
	Population by age/sex	combined	1.36	1.70	1.75	1.43	1.58
	Dwelling	separate	1.30	1.69	-	1.48	1.58
	Population	separate	1.33	1.69	-	1.51	1.61
	Population by age/sex	separate	1.33	1.69	-	1.51	1.61

Table (2): Efficiencies of Small Area Estimators Relative to Direct Estimator - Nova Scotia Census Divisions (Auxiliary data out-of-date)

Characteristic	Auxiliary Variable	Level of Construction	Post-Stratified Domain	Synthetic	Sample Dependent (K_o=1.0)
Employed	population	combined	3.47	4.73	4.07
	population by age/sex	combined	3.60	4.73	4.23
Unemployed	population	combined	1.44	2.19	1.68
	population by age/sex	combined	1.46	2.21	1.69