Statistics Canada Statistique Canada

Methodology Branch

Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes-ménages

Canada

WORKING PAPER
METHODOLOGY BRANCH

Surveys of Skewed Populations: Optimal Sample Redesign Under the Generalized Regression Estimator with Application to the Local Government Finance Survey

HSMD - 95 - 006E

Gurupdesh S. Pandher

Survey Analysis and Methods Development Section
Household Survey Methods Division, Statistics Canada

August 1995

The work presented in this paper is the responsibility of the author and does not necessarily represent the views or policies of Statistics Canada

Surveys of Skewed Populations: Optimal Sample Redesign Under the Generalized Regression Estimator with Application to the Local Government Finance Survey

Gurupdesh S. Pandher
Survey Analysis and Methods Development Section
Household Survey Methods Division, Methodology Branch
Statistics Canada, Ottawa, K1A 0T6

Abstract

In a survey re-engineering context, the most efficient sample design and estimation strategy holds the promise of offering the largest reduction in the sample size (and survey costs) for any given level of desired precision. The redesign methodology developed in this paper is directed towards identifying an efficient sample design for surveys of skewed populations (eg. business, agricultural, and institutional populations) under the generalized regression estimator.

A scheme called the "Transfer Algorithm" is proposed to find an optimal partitioning of the skewed population into the take-all and sampled groups. The criterion for constructing these groups is based directly on the design variance of the regression estimator under a flexible range of sample selection designs (eg. SRS, pps, generalized pps). The Transfer Algorithm is then iteratively integrated with a sample size determination step. The combined procedure identifies the globally minimal sample size and population allocation required to meet the precision constraint under the desired sample design.

Desirable mathematical properties of the Transfer Algorithm such as existence and optimality of solution are established. An equivalence result is obtained allowing the solution to be alternatively determined in terms of simple quantities computable directly from the population auxiliary data. The optimality of the combined procedure is also established. The theoretical results are reported in Theorems 1 to 4. The completed methodology is illustrated using Ontario provincial data from the Local Government Finance Survey and a graphical representation of the combined methodology is given. Using total provincial expenditures as the control variable, a 52% reduction in the sample size was achieved at a precision of 2% coefficient of variation.

Méthodologie optimale de remaniement du plan d'échantillonnage des enquêtes à populations asymétriques suivant l'estimateur par régression généralisé, avec application au remaniement de l'Enquête sur les finances des administrations locales

Gurupdesh S. Pandher
Section du développement des méthodes d'enquête et d'analyse
Division des méthodes d'enquêtes sociales
Direction de la méthodologie
Statistique Canada, Ottawa, K1A OT6

Résumé

Lorsqu'il faut remanier une enquête, la meilleure stratégie d'échantillonnage et d'estimation est celle qui permet de réduire au maximum la taille de l'échantillon (et donc les coûts d'enquête) tout en garantissant le degré de précision souhaité. La méthodologie combinée présentée dans le document s'attaque au problème qui se pose à qui veut trouver la taille minimale de l'échantillon. Elle permet de déterminer la taille optimale de la population et la meilleure répartition de l'échantillon entre les unités des groupes «à tirage complet» et «échantillonnés» que l'on retrouve souvent dans les enquêtes à populations asymétriques (p. ex., les populations d'entreprises, d'exploitations agricoles et d'établissements institutionnels).

Un ensemble de règles appelé «algorithme de transfert» est proposé pour un cloisonnement optimal de la population asymétrique entre les groupes à tirage complet et échantillonnés. Le critère utilisé pour construire ces groupes est fondé directement sur la variance de plan de l'estimateur par régression suivant une gamme variable de plans d'échantillonnage (p. ex., ÉAS, échantillonnage PPT, échantillonnage stratifié). L'algorithme de transfert est ensuite intégré itérativement à un pas de détermination de la taille de l'échantillon pour produire une méthodologie combinée qui permet de connaître la taille minimale de l'échantillon global, le cloisonnement optimal de la population et la répartition optimale de l'échantillon compte tenu du degré de précision exigé par le plan d'échantillonnage retenu.

Les propriétés mathématiques souhaitables de l'algorithme de transfert, notamment l'existence et l'optimalité de la solution, sont établies. Un résultat d'équivalence permet aussi d'exprimer la solution en quantités simples calculables directement à partir des données auxiliaires de population. L'optimalité de la procédure combinée de remaniement du plan d'échantillonnage est également établie. Ces résultats théoriques sont démontrés dans les théorèmes 1 à 4. La méthodologie est illustrée par les données provinciales tirées de l'Enquête sur les finances des administrations locales; la méthodologie combinée est représentée graphiquement.

Surveys of Skewed Populations: Optimal Sample Redesign Under the Generalized Regression Estimator with Application to the Local Government Finance Survey

Gurupdesh S. Pandher [*]

1. INTRODUCTION

In many survey situations additional information is available on all population units before the survey is undertaken. This auxiliary information is frequently useful in devising a more efficient sample design and estimation strategy. In a survey redesign context, the most optimal strategy holds the promise of offering the largest reduction in survey costs by requiring the lowest sample size necessary to meet the desired precision constraint on the estimates. In repeat surveys of skewed populations, an efficient sample design and estimation strategy may be realized by exploiting a) the correlation structure between the size-related auxiliary information x (eg. population of municipality, employees in a firm, farm acreage) and the survey variables y (eg. municipality expenditures, value of shipments, farm yield) and b) the variance relationship between the survey variables and the auxiliary size information.

In this paper, a comprehensive sample redesign methodology is developed for skewed populations with the ultimate objective of bringing about maximal reductions in the current sample size while ensuring a desired level of precision for the generalized regression estimator (GREG) of the population total. This work was motivated by the redesign of the Local Government Finance Survey (LGFS) conducted by Statistics Canada's Public Institutions Division. Financial information (eg. revenues, expenditures, debt, etc.) obtained from local government units is used in the estimation and publication of financial statistics on a provincial and national basis. Although the work presented in this paper is motivated by a concrete application, the sample design methodology devised applies generally to all surveys based on skewed populations (eg. agricultural, business, and institutional surveys).

[*] Gurupdesh Pandher is Methodologist with Survey Analysis and Methods Development Section, Household Survey Methods Division, Methodology Branch, Statistics Canada, 16th Floor, Coats Building, Ottawa, Ontario K1A 0T6, Canada. The research reported in the paper was sponsored by Public Institutions Division.

In identifying an efficient new sample design, the overall methodology addresses and integrates the solution to three problems:

1) Creation of the "Take-all" and "Sampled Groups"

Since the variability of the survey response y_k tends to increase with the size of the unit x_k, it is common in skewed populations to sample the largest x-valued units with certainty in order to improve the efficiency of the population estimators. The demarcation of the population into the non-overlapping "take-all" $U_a = \{1, \ldots, N_a\}$ and "sampled" groups $U_b = \{1, \ldots, N_b\}$ is obtained through a new scheme named the "Transfer Algorithm".

2) Choosing an Efficient Sample Selection Scheme

Let $p(s;\lambda) = \big(p_a(s_a), p_b(s_b;\lambda)\big)$ represent the complete sample design for U_a and U_b where the sample design parameter λ determines the type of sample selection implemented in the sampled group U_b. The sample inclusion probabilities due to $p_b(s_b;\lambda)$ may be expressed as $\pi_k(\lambda) = n_b \left(x_k^{\lambda/2} / \sum_{U_b} x_j^{\lambda/2}\right)$, $k \in U_b$. Note that the parameter λ defines a broad class of sample designs with simple random sampling (SRS: $\lambda = 0$) and probability proportional to size sampling (pps: $\lambda = 2$) as particular cases. Design optimality results (Godambe and Joshi, 1965) allow the identification of the most optimal value for the sample design parameter λ.

3) Minimal Sample Size Determination

The third component of the overall methodology is aimed at finding the minimal sample size required to meet the imposed precision constraint for the estimator.

The combined procedure developed in the paper integrates these components to enable a new globally minimal sample size and optimal population partitioning to be determined under a flexible range of sample selection strategies (eg. SRS, pps, generalized pps).

First, an iterative scheme - called the "Transfer Algorithm" - is developed which finds an optimal allocation of population units between the take-all and sampled population groups in the sense of minimizing the variance of the generalized regression estimator. Desirable mathematical properties of this algorithm such as existence and optimality of solution are established. An equivalence result is also obtained allowing the solution to be determined in terms of simple quantities computable directly from the population auxiliary data.

The Transfer Algorithm is then synthesized iteratively with the sample size determination step to find the minimal sample size needed to satisfy the imposed precision constraints. It is shown that the combined methodology produces a sequence of sample sizes and population partitionings which converges to a globally optimal solution where further reduction in the sample size is not possible given the imposed precision constraint. The procedure is illustrated using provincial data from the Local Government Finance Survey in Ontario and a graphical representation of the combined methodology is given.

Lavallee and Hidiroglou (1988), Hidiroglou and Srinath (1993) (subsequently denoted as L&H and H&S, respectively), and Glasser (1962) have proposed alternative methodologies for constructing the take-all and sampled groups within the context of stratified SRS design. The proposed approach differs from other methods in three respects. Firstly, the population demarcation is obtained under a flexible range of sample selection strategies (eg. SRS, pps, generalized pps). Secondly, the criterion for constructing the population demarcation is based on minimizing the variance of the GREG estimator of the total under the desired sample selection strategy (Glasser and L&H base their allocation on minimizing the within-stratum sum-of-squares x; H&S use the total regression sum-of-squares under a regression model with a compulsory intercept assuming SRS). Thirdly, the proposed methodology explicitly captures the size-induced heteroscedasticity present in skewed survey populations which has been ignored in other frameworks.

2. SURVEY FRAMEWORK

The model assisted survey framework is adopted for the skewed population whose auxiliary and survey characteristics are denoted by $C_U = \{(x_1, y_1), \ldots, (x_N, y_N)\}$. In this framework, underlying the class of generalized regression estimators for the population total (GREG) are regression models (Sarndal et. al. 1992, p.255) which enable the estimators to exploit the correlation between the survey variables y and the auxiliary covariates x. Different model assumptions on the deterministic and stochastic components of the underlying model lead to different regression estimators for the population total. For example, a ratio-form heteroscedastic model

$$y_k = \beta\, x_k + \epsilon_k, \qquad (2.1)$$

with the error $\epsilon_k \sim (0, \sigma_k^2)$ and the variance structure given by $\sigma_k^2 = c\, x_k^\gamma$ (2.2) leads to the following GREG estimator for $t_b = \sum_{k \epsilon U_b} y_k$:

$$\hat{t}_{Rb} = \sum_{U_b} x_k \hat{B} + \sum_{s_b} \frac{(y_k - x_k \hat{B})}{\pi_k} \quad (2.3)$$

where $\hat{B} = \left(\sum_{s_b} y_k / \pi_k\right) / \left(\sum_{s_b} x_k / \pi_k\right)$ is the sample-based probability weighted estimate of the population regression parameter $B = \left(\sum_{U_b} y_k\right) / \left(\sum_{U_b} x_k\right)$.

It is helpful to visualize the population data given by $C_U = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ as a scatterplot of N points (see Figures 1 and 2). In skewed survey populations the variability in the survey response y tends to increase with the size (x) of the population unit. In the population scatter, this behaviour shows up as a "fanning out" pattern of points (x_k, y_k) along the population line Bx as the value of x increases. Hence, the error variances σ_k^2 are an increasing function of x_k. This relationship may be specified as $\sigma_k^2 \propto h(x_k)$. A parameterized form of $h(x_k)$ proposed by Sarndal et. al. (1992, p.462), which is general, yet simple enough to be practically useful is $h(x_k) = c\, x_k^{\gamma}$, where $\gamma \geq 0$ is the heteroscedasticity parameter.

Fig.1 Example of Population Scatterplot (X, Y)

X = Population Size, Y = Revenues (Ontario)

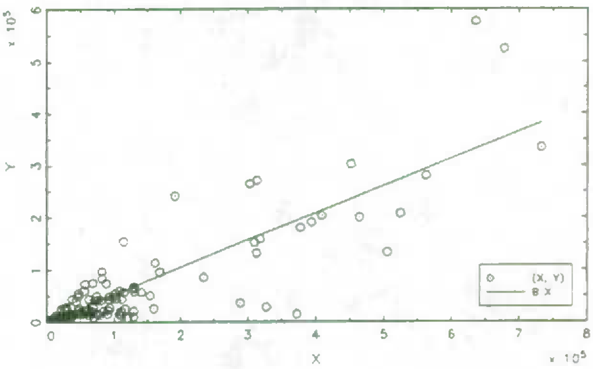
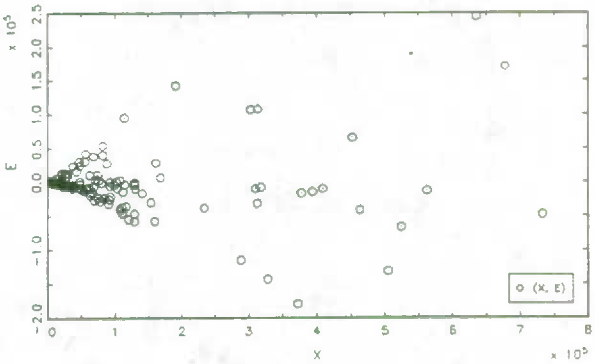


Fig.2 Example of Population Residual Scatterplot (X, E)

X = Population Size, Y = Revenues, E = Y − B*X (Ontario)



Once the sample has been selected, the ultimate purpose is to estimate the population total $t = t_a + t_b$ for the survey variable y where t_a and t_b represent the sub-totals for the take-all and sampled groups, respectively. The total across both groups $t = t_a + t_b$ is estimated by $\hat{t} = t_a + \hat{t}_{Rb}$ where $\hat{t}_a = t_a = \sum_{U_a} y_k$ since all units are sampled in the take-all group and $\hat{t}_{Rb}$ is the GREG estimator under the relevant model. The anticipated variance of $\hat{t}_{Rb}$ (defined as the variance with respect to both the design and the model, denoted p and ξ, respectively) is expressible as

$$V(\hat{t}_{Rb}) \equiv \varepsilon_{\xi} V_p(\hat{t}_{Rb}) \doteq \sum_{k \epsilon U_b} (\frac{1}{\pi_k} - 1)\, \sigma_k^2 \qquad (2.3)$$

Furthermore, if σ_k^2 is given by $\sigma_k^2 = c\, x_k^{\gamma}$ (2.2), then design optimality (Godambe and Joshi, 1965) implies that the optimal sample inclusion probabilities are $\pi_k^*(\gamma) \propto x_k^{\gamma/2}$, $k \epsilon U_b$. Therefore, the sample design $p_b^*(s_b; \lambda{=}\gamma)$ in the sampled sub-population, defining the first order inclusion probabilities $\pi_k^*(\gamma) = n_b \left(x_k^{\gamma/2} / \sum_{U_b} x_j^{\gamma/2}\right)$, $k \epsilon U_b$, minimizes the anticipated variance $V(\hat{t}_{Rb})$.

Three methods for estimating the heteroscedasticity parameter γ from past survey data called the "Least Squares Method", the "Maximum Likelihood Method", and the "Graphical Method" are described in Appendix A. Results from applying these methods to Local Government Finance Survey data are also reported.

3. OPTIMAL DETERMINATION OF TAKE-ALL AND SAMPLED SUB-POPULATIONS

In this section, an iterative scheme named the "Transfer Algorithm" is proposed to determine the optimal demarcation between the take-all and sampled sub-populations under the sample design $p(s; \lambda)$. The criterion for this construction is based on finding a population partitioning minimizing the estimated anticipated variance of $\hat{t}_{Rb}$. An equivalence result is obtained, providing an alternative and simpler method of solution based entirely on auxiliary population data. Desirable mathematical properties such as existence and optimality of solution are also established.

3.1 The Transfer Algorithm

The proposed scheme for constructing the take-all and sampled sub-populations, U_a and U_b, respectively, is based on the following idea. Initially, place all population units in the sampled group, labelling it $U_b^{(0)}$ (the superscript l represents the iteration cycle). Hence, the take-all group is an empty set $U_a^{(0)} = \varnothing$. The resulting population and sample size allocation at $l=0$ is given by $N_a^{(0)} = 0$, $n_a^{(0)} = 0$, $N_b^{(0)} = N$, and $n_b^{(0)} = n_0$ where n_0 is the current sample size.

In a repeat survey setting, the variances σ_k^2 in (2.3) can be empirically modelled using the relation $\sigma_k^2 = c\,x_k^{\gamma}$ (2.2) where γ and c are estimated from past sample data using the methods of Appendix A. Using the estimated version of (2.2) in (2.3) yields the following estimator for $V^{(l)}(\hat{t}_{Rb};\cdot)$:

$$\hat{V}^{(l)}(\hat{t}_R;\lambda, N_b^{(l)}, n_b^{(l)}) = \sum_{k\in U_b^{(l)}} \left[\frac{1}{\pi_k(\lambda)} - 1\right] \hat{c}\,x_k^{\hat{\gamma}} \tag{3.1}$$

where the largest l x-valued units have been removed from $U_b^{(0)}$. Note that λ is used here to parameterize the sample design to allow greater generality when $\lambda \neq \gamma$. This distinction is important because for any given auxiliary-size variable x, different γ values will hold for different target survey variables.

In the iterative algorithm, we start initially with all population units placed in $U_b^{(0)}$. Then at each iteration $l, 0 \le l < n$, the largest l+1 x-valued unit $x_{(N-l-1)}$ is transferred from $U_b^{(l)}$ to $U_a^{(l)}$ and the difference

$$\Delta(l) = \hat{V}^{(l+1)}(\hat{t}_{Rb};\lambda, N-l-1, n-l-1) - \hat{V}^{(l)}(\hat{t}_{Rb};\lambda, N-l, n-l) \tag{3.2}$$

is computed. Negative values of $\Delta(l)$ mean that the transfer of the unit corresponding to the ordered value $x_{(N-l-1)}$ lead to a decrease in the variance. Moreover, such transfers continue to result in a reduction in the variance of $\hat{t}_{Rb}$ as long as $\Delta(l) < 0$. In general, for any iteration l, the relationship between the population and sample size allocations is described by the following relations: $N_b^{(l)} = N - l$, $n_b^{(l)} = n - l$, and $N_a^{(l)} = n_a^{(l)} = l$. These relations hold because the overall population and sample sizes must remain constant ($N = N_a^{(l)} + N_b^{(l)}$ and $n = n_a^{(l)} + n_b^{(l)}$) for all iterations.

The solution is also constrained by the condition $\pi_k(\lambda) < 1, k\in U_b^*(l^*)$. Note that if $\pi_{(N-l^*)} < 1$, then $\pi_{(N-k)} < 1, l^* \le k \le n$, since $x_{(N-k)}^{\lambda/2} \le x_{(N-l^*)}^{\lambda/2}, l^* \le k \le n$. Let $l^*(\lambda), 0 \le l^* < n$, represent the solution to the Transfer Algorithm. Given the discussion above, the solution to the Transfer Algorithm under the sample design $p(s;\lambda)$ may be formulated as

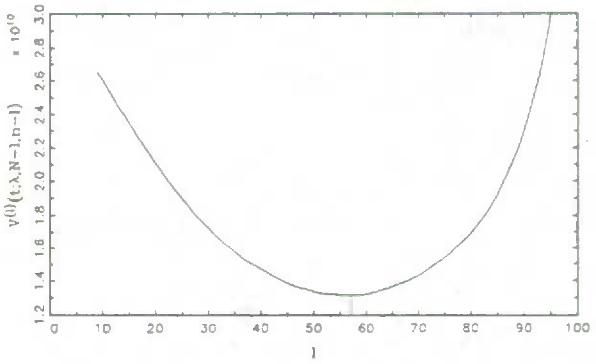
$$l^*(\lambda) = \min_{l}\left\{l:\ \left[\pi_{(N-l)}(\lambda) < 1\right] \mathit{and}\ \Delta(l) = \left[\hat{V}^{(l+1)}(\hat{t}_{Rb};\lambda) - \hat{V}^{(l)}(\hat{t}_{Rb};\lambda)\right] \ge 0,\ 0 \le l < n\right\} \tag{3.3}$$

The optimal population allocation to the take-all group $U_a^*(l^*)$ is then given by the population units coinciding with the l^* ordered units transferred to the take-all auxiliary vector $\mathrm{X}_a^* = \left(x_{(N-l^*)}, x_{(N-l^*+1)}, \ldots, x_{(N)}\right)$; correspondingly the sampled group $U_b^*(l^*)$ consists of the units corresponding to $\mathrm{X}_b^* = \left(x_{(1)}, x_{(2)}, \ldots, x_{(N-l^*-1)}\right)$. It is clear from (3.3) that the solution to the Transfer Algorithm $l^*(\lambda)$ also depends on the sample design $p(s;\lambda)$ - indexed by λ - in effect.

Mathematical properties of the solution (3.3) of the Transfer algorithm such as existence and optimality are studied in detail in Section 4.2, however, the following general observations can be made here. Transferring a unit from $U_b^{(l)}$ to $U_a^{(l)}$ causes two opposite effects on the variance $V^{(l)}(\hat{t}_{Rb};\cdot)$. The reduction in the population size $(N_b^{(l+1)} = N_b^{(l)} - 1)$ has the impact of decreasing the variance, while the equivalent reduction in the sample size $(n_b^{(l+1)} = n_b^{(l)} - 1)$ has the reverse effect of increasing $V^{(l)}(\hat{t}_{Rb};\cdot)$. Somewhere in this process, a critical value $l^*, 0 \leq l^* < n$, exists which gives the optimal breakdown $\left\{U_a^*(l^*), U_b^*(l^*)\right\}$.

Fig.3 Variance of Regression Estimator at l (λ=1)

N=Population Size, n=Current Sample Size, l=Units Transferred



The behaviour of this system is also affected by the initial sample size $n^{(0)} = n$ and the distribution of the values $x_k^{\gamma/2}, k \in U$, in the population. It is also possible that for certain configurations, $\Delta(l) > 0$ holds for all $0 \leq l < n$. This means that no efficiency gains can be realized from transferring units as described in the proposed methodology; the optimal construction of take-all and sampled groups is then given by $U_a^* = \varnothing$ and $U_b^* = U$, with $l^* = 0$.

An example of the application of the Transfer Algorithm to the LGF survey population of local municipalities in Ontario (with N=793, n=108, γ=2, and λ=1) is given in Figures 3 and 4. The curves are plotted for $l > 8$ because in the interval $0 < l \leq 8$, the first condition of (3.3), namely $\left[\pi_{(N-l)}(\lambda) < 1\right]$, is not satisfied. Note that in Figure 3, the minimum value of $\hat{V}^{(l)}(\hat{t}_{Rb})$ is achieved at $l^* = 57$ and in Figure 4 this point coincides with $\Delta(l^*) = \hat{V}^{(l^*+1)} - \hat{V}^{(l^*)} \geq 0$. In Table 3.1, the solution $l^*(\lambda)$ to the Transfer Algorithm as defined in (3.3) and (3.18) are reported for $0 \leq \lambda \leq 2\gamma$ (γ=2).

Fig.4 Changes in Variance of Regression Estimator (λ=1) :

$\Delta(l) = V^{(l+1)}(t;1,N-l-1,n-l-1) - V^{(l)}(t;1,N-l,n-l)$

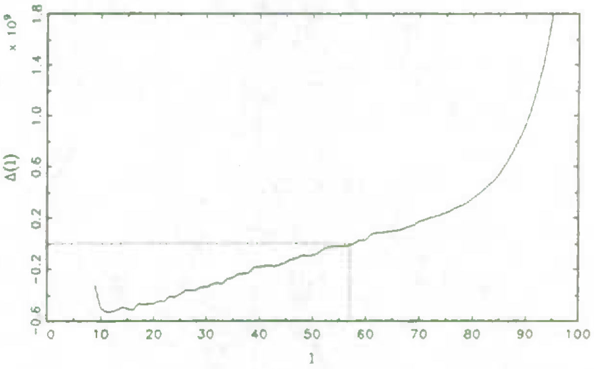


Table 3.1 Solution to Transfer Algorithm $l^*(\lambda)$ for $0 \leq \lambda \leq 2\gamma$ (γ=2)

λ	$l^*(\lambda)$ **Definition (3.3)**	$l^*(\lambda)$ **Definition (3.18)**
0	64	64
.5	60	60
1.0	57	57
1.5	50	50
2.0	39	39
2.5	50	50
3.0	57	57
3.5	60	60
4.0	64	64

3.2 Analysis of the Transfer Algorithm

In this Section, the Transfer Algorithm described in the previous section is analyzed. This is done with two questions in mind: i) does the algorithm converge to a solution? and ii) is the solution optimal? Furthermore, the analysis below reveals that the solution defined by (3.5) can be expressed equivalently in terms of quantities which are much simpler to compute. This equivalence result is established first before investigating the properties of the solution.

3.2.1 Equivalence Result

From the expression for the variance of $V^{(l)}(\hat{t}_{Rb};\cdot)$ given in (2.3), we have after substituting for $\pi_k(\lambda)$

$$V^{(l)}(\hat{t}_{Rb};\lambda,N^{(l)},n^{(l)})=\sum_{k=1}^{N^{(l)}}\left[\frac{\sum_{j=1}^{N^{(l)}} x_{(j)}^{\lambda/2}}{n^{(l)}\,x_{(k)}^{\lambda/2}}-1\right]\sigma_{(k)}^2 \qquad (3.4)$$

or, equivalently,

$$V^{(l)}(\hat{t}_{Rb};\lambda,N-l,n-l)=\sum_{k=1}^{N-l-1}\left[\frac{\sum_{j=1}^{N-l-1} x_{(j)}^{\lambda/2}+x_{(N-l)}^{\lambda/2}}{(n-l)\,x_{(k)}^{\lambda/2}}-1\right]\sigma_{(k)}^2+\left[\frac{\sum_{j=1}^{N-l} x_{(j)}^{\lambda/2}}{(n-l)\,x_{(N-l)}^{\lambda/2}}-1\right]\sigma_{(N-l)}^2 . \qquad (3.5)$$

The subscript b in $N_b^{(l)}$ and $n_b^{(l)}$ may be dropped since by definition $N^{(l)}=N_b^{(l)}=N-l$ is the population size of $U_b^{(l)}$ and $n^{(l)}=n_b^{(l)}=n-l$ is the resulting sample size. Moreover, at iteration $l+1$, the variance expression $V^{(l-1)}(\hat{t}_{Rb};\cdot)$ may be written as

$$V^{(l+1)}(\hat{t}_{Rb};\lambda,N-l-1,n-l-1)=\sum_{k=1}^{N-l-1}\left[\frac{\sum_{j=1}^{N-l-1} x_{(j)}^{\lambda/2}-(n-l-1)\,x_{(k)}^{\lambda/2}}{n-l-1}\right]\frac{\sigma_{(k)}^2}{x_{(k)}^{\lambda/2}} \qquad (3.6)$$

After matching common terms and some further reductions, the difference of the variances $\Delta(l)=V^{(l+1)}-V^{(l)}$ may be written as

$$V^{(l+1)} - V^{(l)} = \sum_{k=1}^{N-l-1} \left[\frac{\sum_{j=1}^{N-l-1} x_{(j)}^{\lambda/2} - (n-l-1)\, x_{(N-l)}^{\lambda/2}}{(n-l)\,(n-l-1)} \right] \frac{\sigma_{(k)}^{2}}{x_{(k)}^{\lambda/2}} - \left[\frac{\sum_{j=1}^{N-l} x_{(j)}^{\lambda/2} - (n-l)\, x_{(N-l)}^{\lambda/2}}{(n-l)} \right] \frac{\sigma_{(N-l)}^{2}}{x_{(N-l)}^{\lambda/2}} . \tag{3.7}$$

An estimator of the above expression based on (3.1) is given by

$$\hat{V}^{(l+1)} - \hat{V}^{(l)} = c \sum_{k=1}^{N-l-1} \left[\frac{\sum_{j=1}^{N-l-1} x_{(j)}^{\lambda/2} - (n-l-1)\, x_{(N-l)}^{\lambda/2}}{(n-l)\,(n-l-1)} \right] x_{(k)}^{\gamma-\lambda/2} - c \left[\frac{\sum_{j=1}^{N-l} x_{(j)}^{\lambda/2} - (n-l)\, x_{(N-l)}^{\lambda/2}}{(n-l)} \right] x_{(N-l)}^{\gamma-\lambda/2} . \tag{3.8}$$

where $c > 0$ and $\gamma \geq 0$ are estimated from modelling the relation $\sigma_k^2 = c\, x_k^\gamma$ (2.2) using the methods of Appendix A.

Expression (3.8) further reduces to

$$\hat{V}^{(l+1)} - \hat{V}^{(l)} = c\, \frac{A(l)\; B(l)}{(n-l)\;(n-l-1)} . \tag{3.9}$$

where $A(l) = \sum_{j=1}^{N-l} x_{(j)}^{\lambda/2} - (n-l)\, x_{(N-l)}^{\lambda/2}$ and $B(l) = \sum_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2} - (n-l)\; x_{(N-l)}^{\gamma-\lambda/2}$. For notational convenience, in the remainder of the paper $V^{(l)}$ will be re-defined to represent the estimator of the anticipated variance defined in (3.1).

Next, note that $V^{(l+1)} - V^{(l)} < 0$ in the cases i) $[A(l) > 0 \text{ and } B(l) < 0]$ and ii) $[A(l) < 0 \text{ and } B(l) > 0]$ and $V^{(l+1)} - V^{(l)} \geq 0$ when iii) $[A(l) \geq 0 \text{ and } B(l) \geq 0]$ and iv) $[A(l) \leq 0 \text{ and } B(l) \leq 0]$. In case i), the condition on $n^{(l)} = n - l$ under which $B(l) < 0$ is determined to be

$$n^{(l)} > \left[\frac{\sum_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2}}{x_{(N-l)}^{\gamma-\lambda/2}} \right] . \tag{3.10}$$

Moreover, defining

$$R(l; \gamma-\lambda/2) = \left[\frac{\sum_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2}}{x_{(N-l)}^{\gamma-\lambda/2}} \right] \tag{3.11}$$

allows (3.10) to be written compactly as

$$n^{(l)} > R(l;\gamma-\lambda/2)\,. \tag{3.12}$$

Similarly, the condition on $n^{(l)}=n-l$ required for $A(l)>0$ is obtained by solving $A(l)=\sum_{j=1}^{N-l} x_{(j)}^{\lambda/2} - (n-l)\,x_{(N-l)}^{\lambda/2} > 0$. This yields

$$n^{(l)} < \left[\frac{\sum_{k=1}^{N-l} x_{(k)}^{\lambda/2}}{x_{(N-l)}^{\lambda/2}}\right] \tag{3.13}$$

which using an analogous definition to (3.12) may be re-expressed as

$$n^{(l)} < R(l;\,\lambda/2)\,. \tag{3.14}$$

Similar conditions for $n^{(l)}=n-l$ can be derived for the remaining cases ii), iii), and iv) mentioned above. The results are summarized in Table 3.2.

Table 3.2 Outcomes for $V^{(l+1)}-V^{(l)}<0$ and $V^{(l+1)}-V^{(l)}\geq 0$ in Terms of $n^{(l)}=n-l$.

	$V^{(l+1)}-V^{(l)}<0$		$V^{(l+1)}-V^{(l)}\geq 0$
Behaviour of A and B.	Condition on $n^{(l)}=n-l$	Behaviour of A and B.	Condition on $n^{(l)}=n-l$
$A(l)>0$ $B(l)<0$	$R(l;\gamma-\lambda/2)<n-l<R(l;\lambda/2)$ (T.1)	$A(l)>0$ $B(l)\geq 0$	$n-l \leq \min\{R(l;\lambda/2),\ R(l;\gamma-\lambda/2)\}$ (T.2)
$A(l)<0$ $B(l)>0$	$R(l;\lambda/2)<n-l<R(l;\gamma-\lambda/2)$ (T.3)	$A(l)\leq 0$ $B(l)\leq 0$	$n-l \geq \max\{R(l;\lambda/2),\ R(l;\gamma-\lambda/2)\}$ (T.4)

The first and second columns of Table 3.2 describes the behaviour of $A(l)$ and $B(l)$ leading to the outcomes $V^{(l+1)}-V^{(l)}<0$ and $V^{(l+1)}-V^{(l)}\geq 0$. The second and fourth columns describe the equivalent condition in terms of $n^{(l)}=n-l$ which yield $V^{(l)}-V^{(l-1)}<0$ and $V^{(l)}-V^{(l-1)}\geq 0$. Since all ranges for $n^{(l)}=n-l$ depend on $R(l;\lambda/2)$ and $R(l;\gamma-\lambda/2)$, the solution to the Transfer Algorithm $l^{*}(\lambda)$ given in (3.3), along with the mathematical properties of the solution, will depend on the distribution of the size measure x and the values of γ and λ.

The behaviour of the system described in Table 3.2 also depends on the sample design $p(s;\lambda)$ employed. Three cases are distinguished and discussed below: a) $0 \leq \lambda < \gamma \Rightarrow \left[R(l;\gamma-\lambda/2) < R(l;\lambda/2)\right]$, b) $\lambda = \gamma \Rightarrow \left[R(l;\gamma-\lambda/2) = R(l;\lambda/2)\right]$, and c) $\gamma < \lambda \leq 2\gamma \Rightarrow \left[R(l;\gamma-\lambda/2) > R(l;\lambda/2)\right]$. Although $\lambda > 2\gamma$ is also possible and mathematically defined, this situation is arbitrarily ruled out because it leads to $\gamma-\lambda/2 < 0$ (this term appears as the exponent in $R(l;\gamma-\lambda/2)$).

An important condition required for the solution to the Transfer Algorithm $l^*(\lambda)$ is that $\pi_{(N-l)}(\lambda) < 1$. It is easy to verify that $\pi_{(N-l)}(\lambda) < 1 \Leftrightarrow A(l) > 0$. In terms of the description for the Transfer Algorithm given in Table 3.2, this condition means that the solution can occur only when both $A(l) > 0$ and $B(l) \geq 0$ or, equivalently, when $n-l$ satisfies condition (T.2). Using the implications of each case for λ on the relative ordering of $R(l;\gamma-\lambda/2)$ and $R(l;\lambda/2)$ in Table 3.2, the following equivalence theorem can be readily constructed from the intermediate tables.

Theorem 1. Equivalence Theorem

Let $p(s;\lambda)$ represent the sample design in effect, defining the inclusion probabilities $\pi_k^{(l)}(\lambda) = (n-l)\, x_{(k)}^{\lambda/2} \;/\; \sum_{k=1}^{N-l} x_{(k)}^{\lambda/2}$, $k \in U_b$. Further, with $R(l;\lambda/2) = \sum_{k=1}^{N-l} x_{(k)}^{\lambda/2} \;/\; x_{(N-l)}^{\lambda/2}$ and $R(l;\gamma-\lambda/2) = \sum_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2} \;/\; x_{(N-l)}^{\gamma-\lambda/2}$ defining the critical values for $n-l$, the following equivalences hold for the Transfer Algorithm.

a) $0 \leq \lambda < \gamma$:

$$\Big[\left[\pi_{(N-l)}(\lambda) < 1\right] \mathit{and} \left[V^{(l+1)}(\hat{t}_{Rb};\lambda, N-l-1, n-l-1) - V^{(l)}(\hat{t}_{Rb};\lambda, N-l, n-l)\right] \geq 0\Big] \tag{3.15a}$$

$$\Leftrightarrow \quad \left[n-l \leq R(l;\gamma-\lambda/2)\right]$$

$$\left[V^{(l+1)}(\hat{t}_{Rb};\lambda, N-l-1, n-l-1) - V^{(l)}(\hat{t}_{Rb};\lambda, N-l, n-l) < 0\right] \tag{3.16a}$$

$$\Leftrightarrow \quad \left[R(l;\gamma-\lambda/2) < n-l < R(l;\lambda/2)\right]$$

$$\Big[\,\big[\pi_{(N-l)}(\lambda)\ge 1\big]\ \textit{and}\ \big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\lambda,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\lambda,N-l,n-l)\big]\ge 0\Big] \qquad (3.17a)$$

$$\Leftrightarrow\quad \big[n-l\ \ge\ R(l;\lambda/2)\big]$$

b) $\lambda=\gamma$:

$$\Big[\,\big[\pi_{(N-l)}(\gamma)<1\big]\ \textit{and}\ \big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\gamma,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\gamma,N-l,n-l)\big]\ge 0\Big] \qquad (3.15b)$$

$$\Leftrightarrow\quad \big[n-l\ \le\ R(l;\gamma/2)\big]$$

$$\Big[\,\big[\pi_{(N-l)}(\gamma)\ge 1\big]\ \textit{and}\ \big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\gamma,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\gamma,N-l,n-l)\big]\ge 0\Big] \qquad (3.16b)$$

$$\Leftrightarrow\quad \big[n-l\ \ge\ R(l;\gamma/2)\big]$$

c) $\gamma<\lambda\le 2\gamma$:

$$\Big[\,\big[\pi_{(N-l)}(\lambda)<1\big]\ \textit{and}\ \big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\lambda,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\lambda,N-l,n-l)\big]\ge 0\Big] \qquad (3.15c)$$

$$\Leftrightarrow\quad \big[n-l\ \ge\ R(l;\lambda/2)\big]$$

$$\big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\lambda,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\lambda,N-l,n-l)\ <0\big] \qquad (3.16c)$$

$$\Leftrightarrow\quad \big[R(l;\lambda/2)\ <\ n-l\ <\ R(l;\gamma-\lambda/2)\big]$$

$$\Big[\,\big[\pi_{(N-l)}(\lambda)\ge 1\big]\ \textit{and}\ \big[V^{(l+1)}(\hat{t}_{Rb}\,;\,\lambda,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\lambda,N-l,n-l)\big]\ge 0\Big] \qquad (3.17c)$$

$$\Leftrightarrow\quad \big[n-l\ \ge\ R(l;\gamma-\lambda/2)\big]$$

3.2.2 Simpler Alternative Method of Solution

The methodology of finding the optimal allocation of the population to the take-all and sampled groups was originally developed in terms of the behaviour of the difference

$$\Delta(l) = V^{(l+1)}(\hat{t}_{Rb}\,;\,\gamma,N-l-1\,,n-l-1)\ -\ V^{(l)}(\hat{t}_{Rb}\,;\,\gamma,N-l,n-l)\,. \qquad (3.4)$$

The equivalences established between the behaviour of $V^{(l+1)}-V^{(l)}$ and $\pi^{(l)}_{(N-l)}(\lambda)$ and the terms $n^{(l)}=n-l$, $R(l;\lambda/2)$, and $R(l;\gamma-\lambda/2)$ in Theorem 1 allow the solution $l^*(\lambda)$ to be stated in a greatly simplified - yet equivalent - form. This result is stated in Theorem 2. It is directly obtained from the three components of Theorem 1 keeping in view that the solution $l^*(\lambda)$ must satisfy the two conditions: i) $V^{(l^*)}-V^{(l^*-1)}<0$ and $V^{(l^*+1)}-V^{(l^*)}\ge 0$ and ii) $\big[\pi^{(l^*)}_{(N-l^*)}<1\big]\Leftrightarrow\big[A(l^*)>0\big]$.

Theorem 2. Equivalent Solution to Transfer Algorithm

The solution $l^*(\lambda)$ to the Transfer Algorithm stated in (3.3) in terms of $V^{(l)} - V^{(l-1)}$ and $\pi^{(l)}_{(N-l)}(\lambda)$ may also be equivalently expressed as

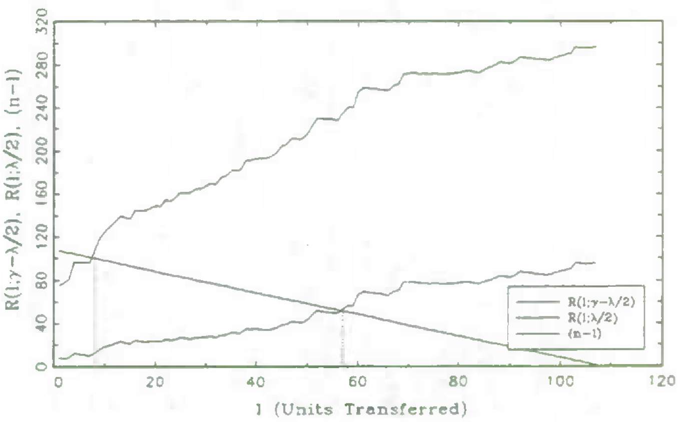
$$l^*(\lambda) = \left\{ \begin{array}{l} \min\limits_{l} \{l:\ n-l \ \le\ R(l;\gamma-\lambda/2)\ ,\ 0 \le l < n\}\ ,\ 0 \le \lambda < \gamma \\ \min\limits_{l}\ \{l:\ n-l \ \le\ R(l;\gamma/2)\ ,\ 0 \le l < n\}\ ,\ \lambda=\gamma \\ \min\limits_{l}\ \{l:\ n-l \ \le\ R(l;\lambda/2)\ ,\ 0 \le l < n\}\ ,\ \gamma < \lambda \le 2\gamma \end{array} \right\} . \tag{3.18}$$

where $R(l;\gamma-\lambda/2) = \sum\limits_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2} \ /\ x_{(N-l)}^{\gamma-\lambda/2}$ and $R(l;\lambda/2) = \sum\limits_{k=1}^{N-l} x_{(k)}^{\lambda/2} \ /\ x_{(N-l)}^{\lambda/2}$.

An example of how (3.18) can be used to find the optimal population allocation is illustrated in Figure 5 (the same Ontario data for the population of local municipalities is used as in Figures 3 and 4) with $\gamma=2$ and $\lambda=1$. In this case $0 \le \lambda < \gamma$, and the solution is determined by the behaviour of functions $R(l;\gamma-\lambda/2)$ and $n-l$ (see Theorem 2). The same solution $l^*=57$ is obtained as before. Moreover, near $l=8$, the functions $R(l;\lambda/2)$ and $n-l$ cross. From Table 3.2 it is clear that $\left[\pi^{(l)}_{(N-l)} \ge 1\right] \Leftrightarrow \left[A(l) \le 0\right]$ for $l \le 8$ and $\left[\pi^{(l)}_{(N-l)} < 1\right] \Leftrightarrow \left[A(l) > 0\right]$ for $l > 8$.

Fig.5 Use of R(l;γ−λ/2), R(l;λ/2), and (n−l)

to Construct Optimal Take-all/Sampled Groups (Ontario)



3.2.3 Existence and Optimality

The issue of existence and optimality is concerned with the question of whether 1) the Transfer Algorithm always converge to a solution and 2) is the solution reached globally optimal ? Note that by construction, the solution $l^*(\lambda)$ guarantees local optimality in the region $[0, l^*(\lambda)]$. The second question is concerned with the conditions required for global optimality. The solution $l^*(\lambda)$ will be optimal if the conditions leading to $l^*(\lambda)$ remain unchanged (stable) in the system defined over the remaining region $(l^*(\lambda), n-1]$. Stability (of conditions) in this region ensures the (global) optimality of $l^*(\lambda)$. In terms of the original formulation, these concepts may be defined as follows:

1) Existence: $\exists\, l^*, 0 \leq l^* < n$, such that $V^{(l^*+1)} - V^{(l^*)} \geq 0$ and $\pi^{(l^*)}_{(N-l^*)} < 1$.

2) Stability: If $V^{(l^*+1)} - V^{(l^*)} \geq 0$, then $V^{(l+1)} - V^{(l)} \geq 0$ and $\pi^{(l)}_{(N-l)} < 1$ for $0 \leq l^* < l < n$.

In Section 3.2.1, $V^{(l+1)} - V^{(l)}$ was expressed as

$$V^{(l+1)} - V^{(l)} = \frac{A(l)\ B(l)}{(n-l)\ (n-l-1)}. \tag{3.9}$$

where $A(l) = \sum_{j=1}^{N-l} x_{(j)}^{\lambda/2} - (n-l)\, x_{(N-l)}^{\lambda/2}$ and $B(l) = \sum_{k=1}^{N-l} x_{(k)}^{\gamma-\lambda/2} - (n-l)\ x_{(N-l)}^{\gamma-\lambda/2}$. Additionally, denote l_A^* to be the smallest value of $0 \leq l < n$ satisfying $A(l) > 0$; similarly, let l_B^* be the smallest value of $0 \leq l < n$ satisfying $B(l) \geq 0$. At the solution $l^*(\lambda)$, the following two conditions are required: i) $V^{(l^*+1)} - V^{(l^*)} \geq 0$ and ii) $\left[\pi^{(l^*)}_{(N-l^*)} < 1\right] \Leftrightarrow \left[A(l^*) > 0\right]$. Keeping (3.9) in view, this implies that at the solution, the condition $B(l^*) \geq 0$ must also hold. Therefore, the solution to the Transfer Algorithm can also be stated as

$$l^* = \max\ \{l_A^*, l_B^*\}\ ,\ 0 \leq l^* < n. \tag{3.19}$$

Further, note that because l_A^* and l_B^* are solutions to two independent systems defined over $0 \leq l < n$, we can re-define existence and stability as follows:

1) Existence: $\exists\, l_A^*, 0 \leq l_A^* < n$, such that $A(l_A^*) > 0$, and (3.20)

$\exists\, l_B^*, 0 \leq l_B^* < n$, such that $B(l_B^*) \geq 0$.

2) Stability: If $A(l_A^*) > 0$, then $A(l) > 0$ for $0 \leq l_A^* < l < n$, and (3.21)

If $B(l_B^*) \geq 0$, then $B(l) \geq 0$ for $0 \leq l_B^* < l < n$.

These properties and the conditions under which they hold are now established for the three cases a) $0 \le \lambda < \gamma \Rightarrow [R(l;\gamma-\lambda/2) < R(l;\lambda/2)]$, b) $\lambda = \gamma \Rightarrow [R(l;\gamma-\lambda/2) = R(l;\lambda/2)]$, and c) $\gamma < \lambda \le 2\gamma \Rightarrow [R(l;\gamma-\lambda/2) > R(l;\lambda/2)]$.

3.2.3.A Existence and Optimality of l^* $(0 \le \lambda < \gamma)$

Existence of l^* $(0 \le \lambda < \gamma)$

It follows from (3.19) above that l^* exists if both l_A^* and l_B^* exist. Recall that $\lambda < \gamma \Rightarrow [R(l;\gamma-\lambda/2) < R(l;\lambda/2)]$. The fact that $n-l$ is a decreasing function over $0 \le l < n$ means that the event $A(l_A^*) > 0 \Leftrightarrow [n-l_A^* < R(l_A^*;\lambda/2)]$ will occur before $B(l_B^*) > 0 \Leftrightarrow [n-l_B^* < R(l_B^*;\gamma-\lambda/2)]$: $l_A^* < l_B^*$.

Existence of l_A^* $(0 \le \lambda < \gamma)$

Initially $(l=0)$, two outcomes are possible: i) either $[A(l) \le 0] \Leftrightarrow [n-l \ge R(l;\lambda/2)]$ or ii) $[A(l) > 0] \Leftrightarrow [n-l < R(l;\lambda/2)]$. The outcome for l^* will depend on which of these cases occurs.

i) $[A(0) \le 0] \Leftrightarrow [n \ge R(0;\lambda/2)]$

Note that $n^{(l)} = n-l$ is a strictly decreasing linear function of l with $n^{(l)}|_{l=n-1} = 1$. On the other hand, $R(l;\lambda/2)|_{l=0} \ge R(l;\lambda/2)|_{l=N-1} = 1$. Therefore, given the fact that $R(l;\lambda/2)|_{l=0} = 1$, there exists a $l_A^*, 0 \le l_A^* < n$, such that $R(l_A^*;\lambda/2) > n-l_A^*$ (with $A(l_A^*) > 0$).

ii) $[A(0) > 0] \Leftrightarrow [n < R(0;\lambda/2)]$

Here, initially the function value $R(0;\lambda/2)$ is above $n-0$ so that $A(0) > 0$. Therefore, existence is satisfied at $l_A^* = 0$.

Existence of l_B^* $(0 \le \lambda < \gamma)$

The proof for the existence of l_B^* is analogous to that for l_A^* with $A(l)$ and $R(l;\lambda/2)$

replaced by $B(l)$ and $R(l;\gamma\text{-}\lambda/2)$, respectively.

Optimality of l^* $(0 \leq \lambda < \gamma)$

For optimality, the conditions which lead to the solution l^* must hold stable in the region $(l^*(\lambda), n-1]$. By $l^* = \max\ \{l_A^*, l_B^*\}$, $0 \leq l^* < n$, stability prevails if the conditions leading to l_A^* and l_B^* in the two independent sub-systems of the Transform Algorithm (defined by $A(l)$ and $B(l)$, respectively) continue to hold in $(l_A^*, n-1]$ and $(l_B^*, n-1]$, respectively.

Stability in $(l_A^*(\lambda), n-1]$ $(0 \leq \lambda < \gamma)$

Again consider the two possible cases initially $(l=0)$ possible.

i) $[A(0) \leq 0] \Leftrightarrow [n \geq R(0;\lambda/2)]$

We are assured of at least one solution by existence, however, the system may be unstable if the function $R(l;\lambda/2)$ crosses $n^{(l)} = n-l$ more than once in the range $[0, n)$.

The behaviour of the function $R(l;\lambda/2)$ depends on the value of λ and the distribution of the auxiliary characteristic x in the population. For example, if $\lambda=0$, then $R(l;\lambda/2) = N-l$ always lies above $n-l$ so that $A(l) > 0, 0 \leq l < n$ with $l_A^* = 0$. Similarly, the distribution of the x-values has an effect on the shape of $R(l;\lambda/2)$. For example, if all $x_{(j)} = c, j \in U$, are constant, then again $R(l;\lambda/2) = N-l$ and $l_A^* = 0$. In these situations, no crossings of $R(l;\lambda/2)$ and $n-l$ are realized over $[0, n)$. The more interesting non-trivial cases occur for $\lambda > 0$ and non-homogeneous values of x.

The step change in the function $R(l;\lambda/2)$ over consecutive values of l roughly parallels the idea of a slope for continuous functions. This jump may be expressed as

$$R(l+1;\lambda/2) - R(l;\lambda/2) = \sum_{k=1}^{N-l-1} x_{(k)}^{\lambda/2} \left[\frac{1}{x_{(N-l-1)}^{\lambda/2}} - \frac{1}{x_{(N-l)}^{\lambda/2}} \right] - 1 . \qquad (3.20)$$

Although the term in ( ) of (3.20) is non-negative since $\left(x_{(N-l)}^{\lambda/2} - x_{(N-l-1)}^{\lambda/2}\right) \geq 0$, the jump can be either positive or negative. However, if all negative jumps of $R(l;\lambda/2)$ are greater than or equal to all jumps of $n^{(l)} = n-l$ given by $n^{(l+1)} - n^{(l)} = -1$, $0 \leq l < n$, then stability is guaranteed: $R(l;\lambda/2)$ crosses $n^{(l)}=n-l$ in only one period $[l_A^*, l_A^*+1]$. Formally, this condition may be expressed as

$$R(l+1;\lambda/2)-R(l;\lambda/2) \geq -1, 0 \leq l < n. \tag{3.21}$$

Solving (3.21) yields

$$\left(x_{(N-l)}^{\lambda/2}-x_{(N-l-1)}^{\lambda/2}\right) \geq 0, 0 \leq l < n, \tag{3.22}$$

Hence, the inequality of (3.22) always holds ensuring unconditional stability for $A(l)$ in $(l_A^*(\lambda), n-1]$. Note that condition (3.22) allows for ties among consecutive auxiliary values.

ii) $[A(0)>0] \Leftrightarrow [n < R(0;\lambda/2)]$

This is the second outcome initially possible when $l=0$. It was shown earlier that in this case $l_A^*=0$. Stability in this case requires that $R(l;\lambda/2)$ remain above $n^{(l)}=n-l$ for all $0 \leq l \leq n$, with the two curves never crossing. Again the stability condition (3.22) assures that this does not happen.

Stability in $(l_B^*(\lambda), n-1]$ $(0 \leq \lambda < \gamma)$

The proof for stability in $(l_B^*(\lambda), n-1]$ is analogous to that for l_B^*, with $A(l)$ and $R(l;\lambda/2)$ replaced by $B(l)$ and $R(l;\gamma-\lambda/2)$, respectively. The condition required for stability turns out to be the same as (3.22).

3.2.3.B Existence and Optimality of l^* $(\lambda=\gamma$ and $\gamma<\lambda\leq 2\gamma)$

The proof for the existence and optimality of $l^*(\lambda)$ in the remaining two cases b) $\lambda=\gamma$ and c) $\gamma<\lambda\leq 2\gamma$ is analogous to that for case a) $0\leq\lambda<\gamma$ given above and lead to the same results. These results follow easily from using the relevant choice of the functions $R(l;\lambda/2)$ and $R(l;\gamma-\lambda/2)$ and are not repeated again in the interest of brevity.

The results regarding the existence and optimality of the solution delivered by the Transfer Algorithm proved above are summarized in the theorem below.

Theorem 3 Existence and Optimality of Solution to Transfer Algorithm

The Transfer Algorithm always converges to a solution $0 \leq l^*(\lambda) < n$ defined in Theorem 2. For $0 \leq \lambda \leq 2\gamma$, $\gamma \geq 0$, the solution $l^*(\lambda)$ reached is optimal under the conditions stated below:

a) $\lambda = 0$ and/or $\gamma = 0$ and/or $x_{(j)} = c, j \in U$, are constant: the solution $l^*(\lambda) = 0$ is optimal.

b) $\gamma > 0$ and $0 < \lambda \leq 2\gamma$: the solution $0 < l^*(\lambda) < n$ is optimal because

$$\left(x_{(N-l)}^{\lambda/2} - x_{(N-l-1)}^{\lambda/2}\right) \geq 0 \text{ and } \left(x_{(N-l)}^{\gamma-\lambda/2} - x_{(N-l-1)}^{\gamma-\lambda/2}\right) \geq 0,\ 0 \leq l < n.$$

Note that due to the ordering imposed on the population auxiliary values $\{x_1, x_2, \ldots, x_N\}$, conditions $\left(x_{(N-l)}^{\lambda/2} - x_{(N-l-1)}^{\lambda/2}\right) \geq 0$ and $\left(x_{(N-l)}^{\gamma-\lambda/2} - x_{(N-l-1)}^{\gamma-\lambda/2}\right) \geq 0$ hold for all $0 \leq l < n$. Graphically, this ensures that i) the $R(l;\lambda/2)$ and $R(l;\gamma-\lambda/2)$ curves do not cross $n^{(l)} = n - l$ from above and ii) the $R(l;\lambda/2)$ and $R(l;\gamma-\lambda/2)$ curves cross $n^{(l)} = n - l$ from below only once.

4. SAMPLE SIZE DETERMINATION & COMBINED ITERATIVE PROCEDURE

Given a sample design $p(s, \lambda)$, $0 \leq \lambda \leq 2\gamma$, with sample size n, the Transfer Algorithm yields an optimal construction of the take-all and sampled sub-populations, $U_a^*(l^*)$ and $U_b^*(l^*)$, respectively. Next, an expression for finding the minimal sample size is obtained which meets the imposed precision constraint - expressed in terms of the coefficient of variation $CV_{\min}$. The sample determination step is then integrated with the Transfer Algorithm to develop a combined procedure which allows the survey designer to find the globally minimal sample size and optimal population partitioning.

4.1 Expression for New Sample Size

Let q represent the iteration cycle for the combined procedure and $n_q^* = n_{aq}^* + n_{bq}^*$ denote the total minimal sample size required to satisfy the precision constraint. Given the sample design $p_q\left(s, \lambda, l_q^*(\lambda, n_q)\right)$, current sample size n_q, and the population partitioning $\left\{U_{aq}^*(l_q^*),\ U_{bq}^*(l_q^*)\right\}$, the precision constraint for $\hat{t}_R = t_a + \hat{t}_{Rb}$ may be stated formally as

$$CV_{\min} \geq \frac{\hat{V}_q^{1/2}(\hat{t}_{Rb}; \lambda, N - l_q^*, n_q - l_q^*)}{\hat{t}_R} \quad (4.1)$$

Solving this inequality for n_{bq}^* gives the following expression for the minimal sample size needed in the sampled group $U_{bq}^*(l_q^*)$ to meet the precision constraint:

$$n_{bq}^* = n_q^* - l_q^*(n_q) = \frac{X(l_q^*, \lambda/2)\ X(l_q^*, \hat{\gamma} - \lambda/2)\ \hat{c}}{\hat{t}_R^{\ 2}\ CV_{\min} + X(l_q^*, \hat{\gamma})\ \hat{c}} \tag{4.2}$$

where $X(l_q^*, \lambda/2) = \sum_{k=1}^{N-l_q^*} x_{(k)}^{\lambda/2}$, $X(l_q^*, \hat{\gamma} - \lambda/2) = \sum_{k=1}^{N-l_q^*} x_{(k)}^{\hat{\gamma}-\lambda/2}$, and $\hat{t}_R$ may be estimated from past survey data corresponding to the period of the auxiliary information. The total new minimal sample size required to meet the precision constraint is then given by

$$n_q^* = n_{aq}^* + n_{bq}^* = l_q^*(n_q) + n_{bq}^* \tag{4.3}$$

4.2 Combined Sample Redesign Methodology

Next, note that the solution to the Transfer Algorithm l^* depends on the current total sample size: $l_q^*(\lambda) \equiv l_q^*(\lambda, n_q)$. Once the new minimal sample size n_q^* is determined, the existing partitioning

$\{U_{aq}^*(l_q^*), U_{bq}^*(l_q^*)\}$ which was optimal at n_q is no longer optimal at the new minimal sample size n_q^* because $l^*(\lambda, n_q^*) \neq l^*(\lambda, n_q)$ if $n_q^* \neq n_q$. Therefore, letting $n_{q+1} = n_q^*$, a new population partitioning from the Transfer Algorithm based on $l_{q+1}^*(\lambda, n_{q+1})$, given by $\{U_{a.q+1}^*(l_{q+1}^*), U_{b.q+1}^*(l_{q+1}^*)\}$, is required to optimize the construction of the take-all and sampled sub-populations. Next, applying (4.2) over $U_{b.q+1}^*(l_{q+1}^*)$ gives a new minimal sample size $n_{q+1}^* = l_{q+1}^*(n_{q+1}) + n_{b.q+1}^*$ required to achieve the desired precision $CV_{\min}$. Proceeding in this fashion, the combined scheme produces a sequence of population partitioning, sample sizes, and sample allocations

$$\left(l^*(\lambda, n_q), \left(n_{aq} = l_q^*, n_{bq} = n_q - l_q^*\right), \left(N_{aq}^* = l_q^*, N_{bq}^* = N - l_q^*\right), \left(n_{aq}^* = l_q^*, n_{bq}^*\right)\right), \quad q = 0, 1, \ldots (4.4)$$

with $n_{q+1} = n_q^* = n_{aq}^* + n_{bq}^*$ and the initial value n_0 (current survey sample size). The combined procedure is repeated until further reductions in the minimal sample size can no longer be achieved. This leads to the stopping rule

$$q^* = \min_q \left\{ q : n_{q+1}^* - n_q^* \geq 0 \right\}. \tag{4.5}$$

4.3 Optimality of Combined Iterative Scheme

The optimality of the combined methodology yielding the sequence (4.4) under the stopping rule (4.5) is analyzed here. First, note that the new total sample sizes $n_q^* = n_{aq}^* + n_{bq}^* = l_q^*(n_q) + n_{bq}^*$, $0 \leq q \leq q^*$ (with $n_q = n_{q-1}^*$ and $n_0 = n$) are found iteratively by i) finding the optimal number $l_q^*(\lambda, n_q^*)$ of the largest $x_{(k)}$-valued units to be transferred (yielding the take-all and sampled population allocation $\left\{ U_{aq}^*(l_q^*), U_{bq}^*(l_q^*) \right\}$ with corresponding sample sizes $n_{aq} = l_q^*(n_q^*)$ and $n_{bq} = n_q - l_q^*(n_q)$) and ii) finding the new minimum sample size n_{bq}^* for $U_{bq}^*(l_q^*)$ required to satisfy the minimum desired coefficient of variation $CV_{\min}$. By construction, all new resulting total sample sizes $n_q^* = l_q^*(n_q) + n_{bq}^*$, $0 \leq q \leq q^*$, meet the pre-specified precision constraint for the estimator $\hat{t}_R = t_a + \hat{t}_{Rb}$.

According to the stopping rule (4.5), future iterations $q > 0$ will be realized only as long as $n_{q+1}^* < n_q^*$ or as long as the minimum precision constraint is met and further sample size reductions are possible. The possibility for this exists because at the new total sample size n_q^*, the old partitioning based on $l_q^*(\lambda, n_q)$ is no longer optimal and a new partitioning $\left\{ U_{a,q+1}^*(l_{q+1}^*), U_{b,q+1}^*(l_{q+1}^*) \right\}$ based on $l_{q+1}^*(n_{q+1}^*)$ (with $n_{q+1} = n_q^*$) may improve the efficiency of the estimator $\hat{t}_R = t_a + \hat{t}_{Rb}$. This assertion may be formalized as follows:

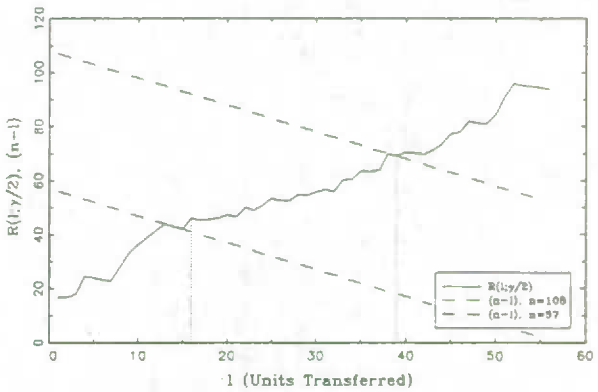
$$V_{q+1}(\hat{t}_{Rb}; \lambda, N - l_{q+1}^*(n_{q+1}), n_{q+1} - l_q^*(n_{q+1})) < V_q(\hat{t}_{Rb}; \lambda, N - l_q^*(n_q), n_{q+1} - l_q^*(n_q)), \quad q = 0, 1, \ldots, q^*, \tag{4.6}$$

with the sample sizes defined by $n_q = n_{q-1}^* = l_{q-1}^*(n_{q-1}) + n_{b,q-1}^*$.

The truth of assertion (4.6) can be established using Theorem 1 (Equivalence) for the sample design $p_q(s, \lambda, l_q^*(\lambda, n_q))$ corresponding to the three cases a) $\lambda < \gamma$, b) $\lambda = \gamma$, and c) $\gamma < \lambda \leq 2\gamma$.

Expression (4.6) is shown to hold for case b) $\lambda = \gamma$ below; extension to the other two cases is straightforward and is omitted to preserve space.

Fig.6 Impact on l(λ,n) of Overall Sample Size (n) Reduction



Now, if $n_{q+1} < n_q$, then $n_{q+1} - l < n_q - l$ while $R(l;\gamma/2)$ remains unchanged. Graphically, the curve $n_q - l$ falls to $n_{q+1} - l$ by a constant amount as shown in Figure 6. At the sample size n_q, the optimal solution given by the Transfer Algorithm is $l_q^*(\gamma, n_q)$ with the two curves $R(l;\gamma/2)$ and $n_q - l$ crossing in the interval $[l_q^*, l_q^* + 1]$. Moreover, it is clear that if $n_{q+1} < n_q$, then $l_{q+1}^*(\gamma, n_{q+1}) < l_q^*(\gamma, n_q)$, and therefore $R(l_q^*;\gamma/2) \geq n_q - l_q^*(\gamma, n_q)$ $> n_{q+1} - l_q^*(\gamma, n_q)$; in fact $R(l;\gamma/2) > n_{q+1} - l$, $l > l_{q+1}^*(n_{q+1})$. By (3.15b), this then leads to the desired result that

$$V_{q+1}(\hat{t}_{Rb}\,;\gamma^*, N - l_{q+1}^*(n_{q+1}), n_{q+1} - l_{q+1}^*(n_{q+1})) < V_q(\hat{t}_{Rb}\,;\gamma^*, N - l, n_{q+1} - l)\ , \quad l > l_{q+1}^*(n_{q+1}) \quad (4.7)$$

Note that for the case $\lambda = \gamma$, the other possibility $\left[l < l_{q+1}^*(n_{q+1})\right] \Leftrightarrow \left[R(l;\gamma/2) < n_{q+1} - l\right]$ is untenable because in this interval $\pi_{N-l}(\gamma) > 1$ so that the variance of $\hat{t}_{Rb}$ is not defined in this region. This establishes the truth of assertion (4.6) under all sample designs $p_q(s, \lambda = \gamma, l_q^*(\gamma, n_q))$, $q = 1, \ldots, q*$. A similar result can also be readily obtained for the other two cases, namely a) $0 \leq \lambda < \gamma$ and c) $\gamma < \lambda \leq 2\gamma$. This leads us to state the following Theorem.

Theorem 4. Optimality of the Combined Sample Redesign Algorithm

Let $p_q(s;\lambda, l_q^*) = \left(p_a(s_a; l_q^*), p_b(s_b;\lambda, l_q^*)\right)$, $q = 0, 1, \ldots$, represent a sequence of sample designs where $l_q^*(\lambda, n_q)$ is the solution to the Transfer Algorithm stated in Theorem 2. Each sample design $p_q(s\,;\lambda\,,l_q^*)$ defines the sample inclusion probabilities $\pi_k = 1\,,\ k \epsilon\ U_{aq}(l_q^*)$ and $\pi_k(\lambda) = (n - l_q^*) \left[x_{(k)}^{\lambda/2} \ / \ \sum_{k=1}^{N-l_q^*} x_{(k)}^{\lambda/2} \right]$, $k \epsilon\ U_{bq}(l_q^*)$. Then, the sequence of population partitioning, sample sizes, and sample allocations

$$\Big(l^*(\lambda, n_q)\,, \big(n_{aq} = l_q^*\,, n_{bq} = n_q - l_q^*\big),\ \big(N_{aq}^* = l_q^*\,, N_{bq}^* = N - l_q^*\big)\,, \big(n_{aq}^* = l_q^*\,, n_{bq}^*\big)\Big)\,,\quad q = 0\,, 1\,, \ldots \text{(4.4)}$$

produced by the combined methodology, where n_{bq}^* is defined by (4.3) and $n_{q+1} = n_q^*$ (with $n_0 = n$), under the stoping rule

$$q^* = \min_q \left\{ q : n_{q+1}^* - n_q^* \geq 0 \right\}. \tag{4.5}$$

is an optimal path in the sense that the variances $V_q(\hat{t}_{Rb}\,;\lambda, N - l_q^*(n_q), n_q - l_q^*(n_q))$, $0 \leq q \leq q^*$, reach their minimum along this path at each prevailing sample size n_q.

Therefore, by construction, the final sample size $n_{q^*}^* = n_{aq^*}^* + n_{bq^*}^* = l_{q^*}^*(n_{q^*}^*) + n_{bq^*}^*$ found by the combined scheme under the stopping rule (4.5) is the globally minimal sample size respecting the pre-set precision constraint (CV_{MIN}) for the estimated total.

The combined procedure leads to an optimal solution along the path defined by (4.4) to a point where further reductions in the sample size are not possible (by reconstructing U_a^* and U_b^*) given the imposed precision constraint.

4.4 Application of Proposed Sample Reduction Methodology

In this section, the combined methodology is applied to data from the Local Government Finance Survey. The survey response y in this application is the actual revenues reported for sampled local government units for Ontario in 1989. The actual estimates are prepared 30 months after the end of the survey year from financial statements submitted by the local government units to the Department of Municipal Affairs (provincial). Population counts for the local government units from the nearest census (1991) are used as the auxiliary variable x. The population of local-level municipalities for Ontario consists of a total of 793 units of which a sample of 108 units is currently taken.

The results of applying the combined methodology to Ontario LGFS data are reported in Table 4.1. The level of desired precision $CV_{\min}$ was set at 2% for the total regression estimator $\hat{t}_R=t_a+\hat{t}_{Rb}$. Using the methods described in Appendix A, the best value for the heteroscedasticity parameter γ was determined to be $\hat{\gamma}=2$ for Ontario; the corresponding proportionality constant was estimated to be $\hat{c}=.0825$. The near optimal sample design defined by $\lambda=\hat{\gamma}$ $(p(s;\hat{\gamma}))$ was used. For each iteration cycle q, the combined methodology (with $n_0=n$) calls for the following steps:

a) Apply the solution to the Transfer Algorithm given in Theorem 2 to find $l_q^*(\lambda,n_q)$. This leads to the population and sample allocation between the take-all and sampled sub-populations given by

$$\left(l^*(\lambda,n_q),\left(n_{aq}=l_q^*\ ,\ n_{bq}=n_q-l_q^*\right),\ \left(N_{aq}^*=l_q^*\ ,\ N_{bq}^*=N-l_q^*\right)\right). \qquad (4.5)$$

b) Use expressions (4.2) and (4.3) to find the minimal sample required to achieve the desired precision $CV_{\min}$ under the population allocation obtained in a). This yields the minimal sample sizes

$$\left(n_{aq}^*=l_q^*(\lambda,n_q),\ n_{bq}^*\right) \qquad (4.9)$$

where $n_q^*=l_q^*+n_{bq}^*$.

c) Set $n_{q+1}=n_q^*$ and repeat steps a) and b) above until the stoping rule (4.5) leads to q^*.

Table 4.1 Application of Combined Methodology to LGF Survey Data (Ontario, 1989)

Iteration (q)	n_q	$l_q(\lambda,n_q)$	n_{aq}^*	n_{bq}^*	n_q^*
0	108	39	39	18	57
1	57	16	16	34	50
2	50	12	12	38	50

For Ontario the combined scheme stopped at iteration $q^* = 1$. The globally optimal population partitioning between the take-all and sampled groups is $N_a^* = 16$ and $N_b^* = 777$. The new minimal total sample size is $n^* = 50$ with allocations $n_a^* = 16$ and $n_b^* = 34$. A total sample size reduction of $n_0 - n_2^* = 108 - 50 = 58$ is achieved at the desired correlation coefficient (CV) of 2% for the regression estimator $\hat{t}_R = t_a + \hat{t}_{Rb}$.

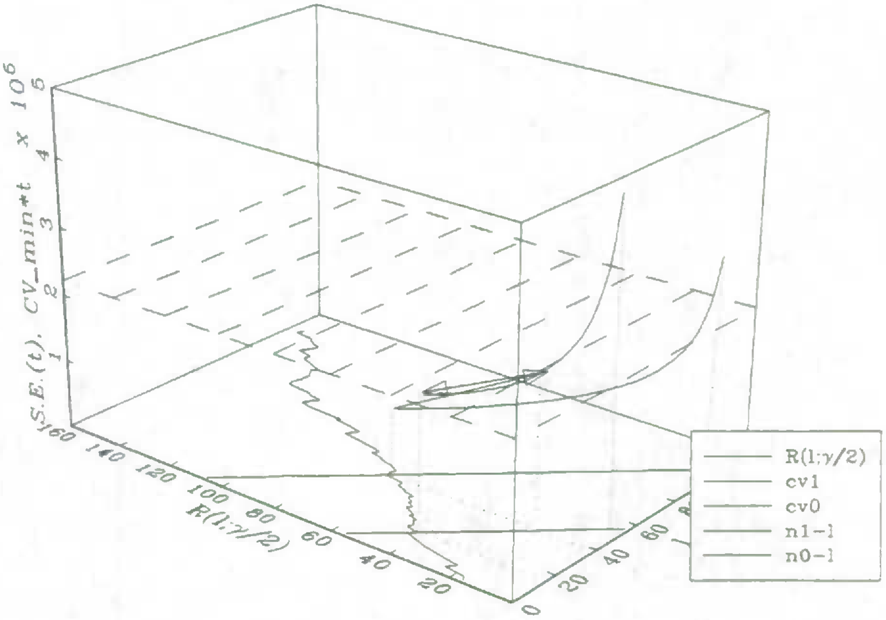
4.5 Graphical Representation of Combined Sample Size Determination Methodology

A graphical representation of the working of the combined procedure using Ontario LGFS data for cycle $q=0$ and $q=1$ is given in Figure 7. The hashed plane represents the precision constraint expressed as $CV_{\min} * \hat{t}_R = 222{,}788$ (with $CV_{\min} = .02$ and $\hat{t}_R = 11{,}139{,}424$). Initially, an application of the Transfer Algorithm gives $l_0^* = 39$. Graphically, this corresponds approximately to the point where the curves $n_0 - l$ ($n_0 = 108$) and $R(l;\gamma/2)$ cross. Above the $n_0 - l$ line, lie the square root of the estimated variances $\hat{V}^{1/2}(\hat{t}_{Rb};\gamma^*, N-l, n_0-l)$ defined over the region $l_{A_0}^* < l < n_0$ (l_A^* is defined in (3.19); in the case $\lambda=\gamma$, $l_A^* = l^*$). It is apparent from the graph that the minimum value of these variances occurs at $l_0^* = 39$ (this was proved in Section 4.3).

After determining l_0^*, a new optimal sample size n_0^* is found using (4.3) which satisfies the precision constraint. In the graph, the system moves as depicted by the first arrow from the point with coordinates $\left(l_0^*, n_0 - l_0^*, \hat{V}^{1/2}(\hat{t}_{Rb};\gamma^*, N - l_0^*, n_0 - l_0^*)\right) = (39, 69, 98507)$ to the point at $\left(l_0^*, n_1 - l_0^*, \hat{V}^{1/2}(\hat{t}_{Rb};\gamma^*, N - l_0^*, n_1 - l_0^*)\right) = (39, 18, 226157)$, yielding a reduction in the sample size of $n_0^* - n_0 = 51$ units. However, at this new point, the variance is not the minimum over the new sample size line defined by $n_1 - l$ ($n_1 = n_0^* = 57$) and, therefore, further reductions in the sample size are still possible. In the next cycle $q=1$, a new population partitioning $l_1^*(n_1)$ optimal at the new sample size is found using the Transfer Algorithm. In Figure 7, this is depicted by the second arrow, moving the system to the point $\left(l_1^*, n_1 - l_1^*, \hat{V}^{1/2}(\hat{t}_{Rb};\gamma^*, N - l_1^*, n_1 - l_1^*)\right) = (16, 41, 197003)$. In this way, the combined procedure continues to move the system along an optimal sequence of decreasing sample sizes and new population and sample partitionings until the precision constraint (given by the plane) allows no further reductions.

Fig.7 Graphical Depiction of Combined Sample

Size Determination Methodology



4. SUMMARY & CONCLUDING REMARKS

The most efficient sample design and estimation strategy holds the promise of offering the largest reduction in the sample size (and hence survey costs) for any desired level of precision in the estimates. This paper provides a comprehensive methodology for identifying and implementing an efficient sample design for recurrent surveys of skewed populations. In the context of using the GREG estimator to estimate the population total, the combined procedure integrates the solution to the following three problems: i) identifying an efficient sample selection scheme, ii) constructing an efficient demarcation between the take-all and sampled population groups at a given sample size, and iii) determining the minimal sample size required to meet the precision constraint(s).

The Transfer Algorithm allows the survey designer to find an optimal allocation of population units between the take-all and sampled population groups in the sense of minimizing the anticipated variance of the regression estimator under the desired sample design. Desirable mathematical properties of this algorithm such as existence and optimality of solution were established and an equivalence result was obtained allowing the solution to be determined in terms of simple quantities computable directly from the population auxiliary data.

The first two components of the overall sample design methodology were then integrated with a sample size determination step through an iterative scheme. This involves a) application of the Transfer Algorithm at the current overall sample size to create the take-all and sampled sub-populations, b) using the precision constraint to find the required sample size in the resulting sampled group, and c) repeating the above two steps as long as further reductions continue to be observed in the overall sample size. Iteration under the stoping rule (4.5) allows convergence to a globally minimal sample size and optimal population partitioning under the imposed precision constraint. The optimality of the combined procedure was also established. The application of the procedure to the Local Government Survey in Ontario resulted in a 52% reduction in the total sample size for the GREG estimator of the total at a minimum coefficient of variation set at 2%.

At the final stage, the desired sample design $p(s;\lambda,l^*(\lambda,n^*))$ (indexed by the design parameter λ) may be implemented in the sampled group $U_b^*(l^*)$. Note that the solution $l^*(\lambda,n)$ given by the Transfer Algorithm applies generally to any sample design defined in the range $0 \leq \lambda \leq 2\gamma$; not merely at the optimal value of the design parameter $\lambda = \gamma$. The presence of λ in the specification of first order inclusion probabilities gives rise to a wide class of generalized pps designs which yield the SRS ($\lambda = 0$) and the standard pps design ($\lambda = 2$) as special cases. A stratified design based on the transformed size-values $x_k^{\gamma/2}, k \in U_b^*$, may be seen as an approximation to the optimal design ($\lambda = \gamma$).

The proposed approach differs from existing methods for constructing the take-all and sampled groups in the literature in three respects. Firstly, an optimal population demarcation can be obtained for a flexible range of sample selection designs (eg. SRS, pps, generalized pps). Secondly, the criterion used to find the optimal population allocation is based directly on minimizing the design-based variance of the regression estimator under the desired sample design. Thirdly, the proposed methodology explicitly captures the size-induced heteroscedasticity evident in skewed survey populations.

Acknowledgements

The author would like to acknowledge the support of Public Institutions Division for sponsoring this work and thank M.P. Singh, M.S. Kovacevic, and H. Lee for their valuable comments on earlier drafts of the paper.

References

Glasser, G.J. (1962). On the Complete Coverage of large Units in a Statistical Study. International Review of the International Statistical Institute, 30, 28-32.

Godambe, V.P., Joshi, V.M. (1965). Admissibility and Bayes Estimation in Sampling Finite Populations. Annals of Mathematical Statistics, 36, 1702-1722.

Hidiroglou, M.A., Srinath, K.P. (1993). Problems Associated With Designing Subannual Business Surveys. Journal of Business and Economic Statistics, Vol. 11, 4, 397-405.

Lavallee, P., Hidiroglou, M.A. (1988). On the Stratification of Skewed Populations. Survey Methodology, Vol. 14.1, 33-43.

Pandher, G.S. (1995). Skewed Survey Populations: Optimal Construction of "Take-all" and "Sampled" Groups with Application to the Local Government Finance Survey Redesign, Methodology Branch Working Paper: SSMD-95-001, p.36. Statistics Canada.

Sarndal, C.E, Swensson, B., Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.

APPENDIX A. ESTIMATION OF FINITE POPULATION HETEROSCEDASTICITY PARAMETER γ

This Appendix is concerned with the estimation of the heteroscedasticity parameter γ in the relation

$$\sigma_k^2 = c\,x_k^\gamma\,\eta_k \qquad \text{(A.1)}$$

where η_k is the multiplicative error explaining the fact that all σ_k^2 do not behave deterministically. (A.1) is a stochastic version of the relation (2.2) used in Section 2.1 to describe the heteroscedasticity of skewed populations. Although γ is not known for the population to be sampled, in repeat surveys like the LGF, data from previous surveys permits estimation of γ.

Three methods are discussed below for estimating γ. No one method was used uniformly to determine the value of γ in each province. Estimates from the different methods were compared. The stability of γ over the observation set was also examined by excluding the largest x-valued observations from the data. This gave a profile of the behaviour of γ over different size ranges. The values of the heteroscedasticity parameter finally chosen in each province also took this analysis into account.

A.1 Least Squares (LS) Approach

This LS approach involves linearizing the relationship between the variance σ_k^2 and x_k given in (A.1) and using the sample estimates of σ_k^2 to then fit the linearized equation. First, the regression

$$y_k = x_k\beta + \epsilon_k \qquad \text{(A.2)}$$

is fitted to obtain the estimated residuals $\hat{\epsilon}_k = y_k - x_k\hat{\beta}$. Empirical investigations into the relationship between the survey variables y based on past sample data reveal that the models (A.2) and (A.1) captures quiet well ($R^2 = .85$) the scatter-plot phenomena - an increasing linear trend and increasing heteroscedasticity with x_k - between revenues (and expenditures) y and the population size x.

Taking the natural logarithm of both sides of (A.1) yields

$$\ln(\sigma_k^2) = \ln c + \gamma \ln x_k + \eta_k^* \qquad \text{(A.3)}$$

where $\eta_k^* = \ln(\eta_k)$ is the additive error component in the linearized form of (A.1). Using $\hat{\epsilon}_k^2$ for $\sigma_k^2, k \epsilon s$, in (A.3) and fitting the model gives the least squares estimate of the heteroscedasticity parameter γ.

A.2 Maximum Likelihood Approach

This method assumes the following normality structure for the errors in model (A.2):

$$\epsilon_k \sim N(0, \sigma^2 x_k^\gamma) \tag{A.4}$$

where $\sigma^2 x_k^\gamma$ is the variance function completely specified upon determining γ. The MLE for γ is given by the solution to

$$g(\gamma, \hat{\beta}_\gamma) = \sum_{k=1}^{n} \frac{(y_k - x_k \hat{\beta}_\gamma)}{x_k^\gamma} \left[\ln x_k - \sum_{h=1}^{n} \frac{\ln x_k}{n} \right] = 0 \tag{A.5}$$

where

$$\hat{\beta}_\gamma = \frac{\sum_{i=1}^{n} y_i x_i^{(1-\gamma)}}{\sum_{i=1}^{n} x_i^{(2-\gamma)}}. \tag{A.6}$$

Expression (A.5) is obtained by first solving the score functions for σ^2 and substituting it into the log-likelihood. The equation for $g(\gamma, \beta_\gamma)$ above is then obtained by differentiating the resulting concentrated-log likelihood function, with respect to γ and setting it to zero. The estimator for $\hat{\beta}_\gamma$ given in (A.6) follows from solving the score function for β and using this expression in conjunction with the score $g(\gamma, \beta_\gamma)$ enables both parameters γ and β to be solved iteratively. A Newton-Raphson algorithm was programmed in GAUSS to obtain estimates for γ and β_γ.

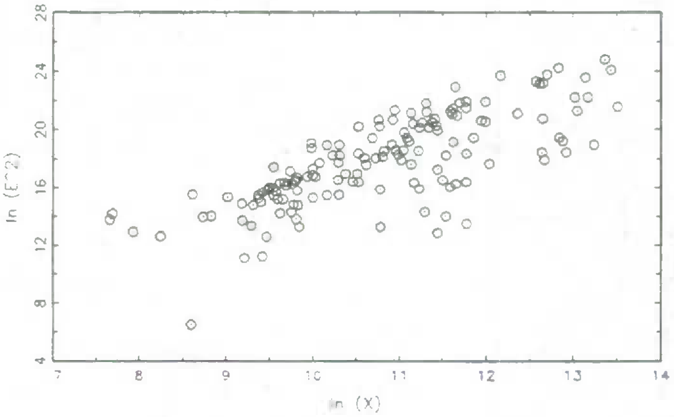
The assumption of normality implicit in the maximum likelihood approach is a drawback since the distribution of local government financial information (e.g., revenues and expenditures) strongly departs from normality. However, the method does yield an alternative estimation methodology which may be used to check and compare the results obtained under other approaches.

A.3 Graphical Approach

In some provinces, due to small sample sizes, the estimation methods discussed above yielded suspicious and unstable estimates for γ when the larger observations are sequentially dropped. This problem was addressed by obtaining some graphical insights into the value for γ. Plots of $\ln(\hat{E}_k^2)$ and $\ln x_k$ (see Figure 3) were examined to ascertain visually the slope of a line through the sample cluster. This rough estimate of the slope should be close to the least squares estimate of γ if a sufficiently large number of points had been available.

Fig.3 Graphical Method to Find Gamma

Plot of ln(E^2) vs. ln(X) (Ontario)



Information about plausible values of γ using this approach was used in addition to the numerical methods discussed above in provinces with small sample sizes and in cases where estimates of γ show instability over reduced observations.

A.4 Application to Local Government Finance Survey Data

The estimates of the heteroscedasticity parameter γ under the least squares (LS) and MLE methods (denoted $\hat{\gamma}_{LS}$ and $\hat{\gamma}_{MLE}$, respectively), after excluding the m largest x-valued observations (effective sample size $n-m$), are reported in Table 2.1. The dependent (survey) variable y was defined as the revenues reported by local government units in the 1989 actual estimates; the independent variable x is the 1991 census count for the municipality.

Table A.1 Least Squares and Maximum Likelihood Estimates of γ and Estimates of Proportionality Constant c

Largest Units Removed (m)	Effective Sample Size (n-m)	$\hat{\gamma}_{LS}$	$\hat{\gamma}_{MLE}$	$\hat{c}$
0	108	1.97	2.05	.0825
1	106	1.72	2.07	.0803
8	100	1.90	2.14	.0853
18	90	1.94	2.10	.0817
28	80	2.15	2.14	.0857
38	70	2.18	2.07	.0737

The graph of $\ln(\hat{E}_k^2)$ vs. $\ln(x_k)$ is exhibited in Figure 3. The slope of this cluster of points is a rough estimate of the value of γ. Based on the insights given by the three methods for possible estimates of γ, the value of the heteroscedasticity parameter was set to $\hat{\gamma} = 2$.

A.5 Estimation of Proportionality Constant *c*

For the purpose of estimating the design variance of the regression estimator given in (3.3), modelling of the error variances σ_k^2, $k \in U$, is required. If the relationship between σ_k^2 and the size value x_k defined by (A.1) holds well in the population, then the disturbance η_k will have a small influence on x_k^γ and the modelled form of (A.1) given by

$$\hat{\sigma}_k^2 = \hat{c}\ x_k^{\hat{\gamma}} \tag{A.7}$$

will give a good empirical approximation for the error variances σ_k^2.

After the best estimate for γ has been identified, the proportionality constant c appearing in (A.1) is estimated. This value is needed to facilitate the estimation of the variance $V(\hat{t}_R)$ given in (3.3). Equation (A.7) suggests the following estimator for c:

$$\hat{c} = \frac{1}{n} \sum_{k \epsilon s} \hat{\epsilon}_k^2 / x_k^{\hat{\gamma}} \tag{A.8}$$

where $\hat{\epsilon}_k = y_k - \hat{\beta} x_k$ serves as the estimate for σ_k^2.

Estimates of c using the estimator (A.8) over different subsets of the Ontario data (y = revenues for 1989, x = 1991 census population counts) excluding the m largest x-valued observations are given in Table 2.1. These estimates over the reduced datasets give some indication as to the sensitivity and stability of the estimation procedure and the behaviour of the data. The estimates $\hat{c}$ for $m = 1, \ldots, 38$ are relatively stable. The value of $\hat{c} = .0825$ at $m = 0$ was chosen for later work.