SSC Annual Meeting, July 1995
Proceedings of the Survey Methods Section

SAMPLING AND ESTIMATION FOR STAGE ONE OF THE CANADIAN SURVEY OF EMPLOYMENT, PAYROLLS AND HOURS SURVEY REDESIGN

M.A. Hidiroglou[1]

ABSTRACT

The survey of Employment, Payrolls and Hours (SEPH) was designed in the early eighties as a monthly rotating panel survey. Its primary objectives were to provide monthly estimates of the total number of paid employees, average weekly earnings, average weekly hours, and other related variables, at the industry by province level. SEPH covered all industries, except agriculture, fishing and trapping, private household services, religious organizations, and military services. Recently, SEPH has been undergoing a redesign because of the availability (monthly) of two of its key variables (total payrolls and total employment) from an administrative data source, that is based on payroll deduction accounts. Sampling fractions range from 10% to 100%, depending on the region. Starting in January 1995, these administrative variables have been available for the larger businesses as well. A sub-sample of the administrative sample is selected and all variables of interest are collected for this sub-sample through a survey. This sub-sample is used to model the collected variables on the administrative variables. A two-phase regression estimator is then used to construct estimates for domains of interest. The paper will focus on the sampling and estimation procedures of the first stage of this redesign.

RÉSUMÉ

L'enquête sur l'emploi, la rénumération et les heures travaillées (EERH) a été conçue au début des années quatre-vingts comme une enquête avec panel à rotation mensuelle. Ses objectifs premiers étaient de fournir des estimés mensuels du nombre d'employés rénumérés, des revenus hebdomadaires moyens, des heures hebdomadaires moyennes, et d'autres variables reliées, au niveau provincial et par industrie. L'EERH couvrait toutes les industries, à l'exception de l'agriculture, des châsses et pêches, des services domestiques, des organisations religieuses, et des services militaires. Récemment, l'EERH a été remaniée au profit de la disponibilité (sur une base mensuelle) de deux de ses variables-clé (salaires totaux et emplois totaux) à partir d'une source de données administratives, basée sur les montants de déductions salariales. Les fractions de sondage varient entre 10% et 100% selon les régions. Depuis janvier 1995 ces variables administratives ont également été disponibles pour les plus grandes entreprises. Un sous-échantillon de l'échantillon administratif est sélectionné et toutes les variables d'intérêt sont observées pour ce sous-échantillon à travers une enquête. Ce sous-échantillon est utilisé pour relier les variables observées aux variables administratives. Un estimateur de régression à deux phases est alors utilisé pour construire des estimés pour les domaines d'intérêt. Le compte-rendu se concentrera sur les procédures d'échantillonnage et d'estimation associées au premier degré du remaniement.

1. INTRODUCTION

The **Survey of Employment, Payrolls and Hours (SEPH)** is a monthly survey carried out by Statistics Canada. It collects data on: i) employment, ii) payrolls, iii) working hours, iv) overtime pay and hours, v) summarized earnings and vi) categories of employment. The primary objectives of the survey are to provide: i) monthly estimates of the total number of paid employees, ii) payrolls average weekly earnings, iii) average weekly hours and other related variables at the three digit Standard Industrial Classification (SIC) level for Canada, and the provinces. SEPH was designed in the early 1980's as a stratified sample of establishments (Schiopu-Kratina and Srinath, 1991). The survey covered all industries except agriculture, fishing and trapping, private household services, religious organizations and military services. The frame of establishments was stratified in three dimensions. Stratification contained an industrial component, a geographical component and a size (based on number of employees) component. As in many business surveys, the highly skewed population made it necessary to stratify by size. This resulted in strata of large self representing establishments (take-all strata) and strata of smaller establishments. The smaller sized strata were subject to sampling (take-some strata). The sampling design was designed to achieve a coefficient of variation (cv) of 3% at the industry division by province level. The estimation procedure was based on survey weights.

[1] M.A. Hidiroglou, Section Chief, Business Surveys Methods Division, Statistics Canada, 11th floor, Section "J", R.H. Coats Building, Ottawa, Ontario, Canada, K1A 0T6.

This sampling design was recently changed because of the availability of administrative data from Revenue Canada. All businesses in Canada submit payroll deductions for income tax, unemployment insurance, and Canada Pension to Revenue Canada Taxation for their employees. These are submitted via payroll deduction (PD) accounts on at least a monthly basis. There are two types of PD accounts: PD7A and PD7TM. PD7A accounts represent the once monthly remitters whose average monthly remittances in the previous year were less than $15,000. PD7TM remitters are accelerated remitters whose average monthly remittances in the previous year were $15,000 or greater. PD7TM remits one to four times monthly depending on the type of payroll employed (i.e., monthly payrolls, weekly, biweekly, etc.) on an accelerated schedule. This availability of administrative data can be used to advantage. In the past, they were used as a source for updating the frame (Dolson, 1994). Since 1993, two additional pieces of data have become available on the PD forms (PD7A) for the smaller enterprises. These two variables are total number of employees and gross monthly payrolls. The availability of this new source of data provided the stimulus for SEPH's redesign. As for sampling impact, this implied that the number of directly surveyed establishments could be reduced. Also, problems such as data processing, frame processing, time lag, rotation, identified with the previous survey were addressed.

These administrative data will soon become available for all enterprises. Future stages of the SEPH redesign are to use these data. Better and more timely estimates, reduction of the cost of the survey, and reduction of response burden will be possible. The progressive provision of administrative data has led to a corresponding redesign of SEPH in three stages. Stage one has been operational since reference month March 1994. This first stage uses the new administrative variables for the smaller enterprises. Payments for these enterprises are remitted to Revenue Canada once a month using PD7A forms. Stage two will expand the use of the administrative variables: medium sized enterprises and some large enterprises with simple structures will be included. In this stage, most of the enterprises with 300 or more employees will be surveyed with certainty. Finally, in stage three, procedures will be developed to maximize the use of the administrative data for the enterprises. The largest enterprises will be surveyed directly. Procedures associated with these last two stages of the redesign can be found in Hidiroglou et al. (1995). In this paper we focus only on the sampling and estimation aspects of stage one of this redesign.

2. SAMPLING DESIGN

The frame for sampling the establishments (Cuthill, 1989) is known as the **Business Register (BR)**. Enterprises on the BR are linked to payroll deduction (PD) accounts, and to establishments. PD accounts are the main source for updating the BR with births and deaths. The portion of the BR linked to PD7A accounts, with the additional two variables, will be called the **ADMINISTRATIVE** portion. The portion without these variables will be called the **ESTABLISHMENT** portion. The ESTABLISHMENT portion is made up of the larger enterprises. These enterprises are surveyed monthly using a design similar to the old survey (Schiopu-Kratina and Srinath, 1991).

The monthly sample based on the ADMINISTRATIVE portion for SEPH is drawn in two phases. The first phase is a sample of payroll PD7A accounts drawn from the ADMINISTRATIVE frame from Revenue Canada. Once this first phase sample of payroll accounts has been drawn, it is matched to enterprises on the Business Register at Statistics Canada. From this match, each enterprise will be linked either to one sampled payroll deduction account or to more than one sampled payroll deduction account. Only one payroll deduction account per linked enterprise is kept in the first phase sample. The second phase sample is then drawn from this list of uniquely linked payroll deduction accounts.

2.1 The First Phase Sample

The first phase sample is designed as a panel. Although deaths are deleted and births added, no sample rotation takes place. The sampling of the PD accounts (currently 10% of the frame in the largest provinces, 100% in the two territories, 20% elsewhere) is simple systematic. This sample consists of all PD account numbers having a prespecified check digit. For each reference month m, it is first selected before data collection for formation of the frame for the selection of the subsample. The sampling of the ADMINISTRATIVE portion frame will be called first-phase sampling. The sampled units are manually selected from the PD7 forms received from employers concerning month m. Data are captured for total employment and total payrolls. However, because the full range of SEPH variables is not available from this source, a subsample (the second-phase sample) is selected from those accounts on the frame. This selection occurs before the reference month that is potentially alive and classified for both industry and province. All SEPH variables are collected with this sample, including total employees and total payrolls for the month.

2.2 The Second Phase Sample

The first phase sample does not contain the full range of SEPH variables. A subsample (currently 2,500 PD accounts) is contacted each month to collect data for: the estimation of total hours and the allocation of hours, earnings and employment to categories of employees

(paid by the hour, salaried, and other). This subsample is selected from those accounts present in the first phase sample before the reference month. The sampling for this subsample is as follows. A subsample (currently 7,500) of PD accounts is selected from the first-phase sample using a set of constant sampling fractions. Note that the first phase sample is stratified by using province and industry to define 125 strata. There is no stratification by employment size group. Rotation and updates for births and deaths take place every month with each sampled unit being kept in the sample for at least four occasions. The subsample (currently 7,500 units) is split into three portions of equal size (currently 2,500 units), each representative of industry and province. One portion is surveyed each month and each portion is resurveyed quarterly.

For estimation purposes, sets of strata are regrouped. These groupings are labelled model groups in which the data are adequately homogeneous with respect to the parameters in the model being used in estimation. Given the resulting sample allocation, expected coefficients of variation of estimated total hours were computed at the model group level. N-Proportional allocation was used to allocate the sample for each model group to the strata.

The formal description of the sampling process is as follows. The population of units in the ADMINISTRATIVE universe will be denoted as U. This universe is then divided into R regions, denoted as U_r, $r = 1, ..., R$. Let the number of units in the rth region be N_r. A first-phase sample of n_{1r} units is then selected from the N_r population units using Bernoulli sampling. The first-phase sample, resulting in each region r is denoted as s_{1r}. The union of the first-phase sample s_{1r} over all regions will be denoted as s_1 where

$$s_1 = \bigcup_{r=1}^{R} s_{1r} .$$

Each first-phase sample s_{1r} is split into H_{1r} strata, s_{1rh}, and

$$s_{1r} = \bigcup_{h=1}^{H_{1r}} s_{1rh} .$$

The resulting sample sizes in each of these first-phase strata (regions) s_{1rh} is n_{1rh} (note that n_{1rh} is a random variable);

$$n_{1r} = \sum_{h=1}^{H_{1r}} n_{1rh} .$$

A second-phase sample s_{2rh} of sizes n_{2rh}, is then drawn from each of the stratified first-phase samples s_{1rh}, $h = 1, ..., H_{1r}$.

3. ESTIMATION

The estimation process uses auxiliary data for each of the two phases. For the first-phase sample known counts, available for partitions of the population, are used to adjust the first-phase weights. This results in poststratified estimation. For the second-phase sample, a model assisted approach is used for the estimation of totals for variables collected. Using total hours as an example, a model involving linear regression is estimated across groups of strata (model groups) using the subsample data. Total employees and total payrolls for the month are the independent variables, while total hours and summarized earnings are the dependent variables. Ratios of total hours by category of employee to total hours are also estimated. Using total employees and total payrolls reported on the administrative source as the auxiliary variables, parameter values from the regression are used to predict total hours for each unit in the model group in the first-phase sample. Finally, the ratios are used to prorate the total into categories of employees. More notation is required to describe the estimation process.

The estimation process can be viewed as a two-phase regression procedure with stratification being applied at the two levels of sampling: namely, the universe U is stratified into regions U_r and within each region r, the first-phase sample s_{1r} is further stratified into strata s_{1rh}. The notation will be an extension of the one presented in Chapter 9 of Särndal, Swensson, and Wretman (1992).

3.1 Phase One Estimation

Denote the vector of auxiliary information for units in the universe U as x_{1k}. These auxiliary data are available for each sampled unit in phase 1. The first phase sample s_1 is broken out into P_1 size groups s_{1p_1}, $p_1 = 1, ..., P_1$. The corresponding population U is broken out into the same size groups (poststratification groups) U_{1p_1}, $p_1 = 1, ..., P_1$. Each poststratification group is of known size N_{p_1}. Let the portion of the first-phase sample s_1 where

$$s_1 = \bigcup_{r=1}^{R} s_{1r}$$

that belongs to U_{p_1} be s_{1p_1}. The realized sample size will be denoted as n_{1p_1}, where

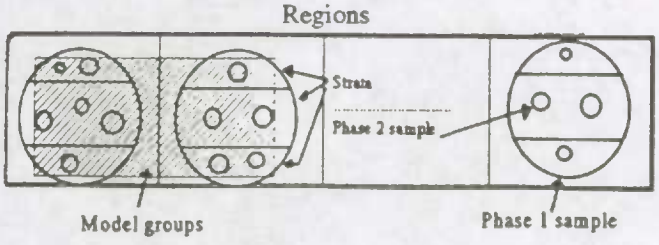
$$n_1 = \sum_{p_1=1}^{P_1} n_{1p_1} .$$

The first-phase inclusion probabilities, within each region, will be denoted as π_{1k} (n_{1r}/N_r for SEPH) for $k \in s_{1r}$. The poststratified auxiliary information can be represented

by the vector $x'_{1k,A}$ = (0,...,1,...,0), a "P_1" dimensional vector. The first-phase poststratified weights for $k \in s_{1p_1}$ are given by: $\tilde{w}_{1k} = w_{1k} N_{p_1} / \hat{N}_{p_1}$ where $\hat{N}_{1p_1} = \sum_{s_{1p_1}} w_{1k}$ and $w_{1k} = 1 / \pi_{1k}$. Denote the vector of auxiliary information for units in the second-phase sample s_2 as $x_{2k,A}$ (payrolls, employment). The full vector of x-variables observed is therefore $x_{k,A} = (x'_{1k,A}, x_{2k,A})$. The estimator of total $X_{2A}(d) = \sum_{U_d} x_{2k,A}$ for domain U_d using the ADMINISTRATIVE data U_d is $\hat{X}_{2A}(d) = \sum_{s_1} \tilde{w}_{1k} x_{2k,A}(d)$ where the components of $x_{2k,A}(d)$ are zero if $k \notin s_d$ and $x_{1k,A}$ otherwise.

3.2 Phase Two Estimation

The second-phase inclusion probabilities, given s_1, are represented as $\pi_{k|s_1}$ (for SEPH this is n_{2rh} / n_{1rh}) for $k \in s_{1rh}$. Data collected by the second-phase sample will be denoted as $y_{k,E}$ and $x_{2k,E}$. The sample s_1 is partitioned into groups where the regression between $y_{k,E}$ and $x_{k,E}$ is good. Each group U_{p_2} = 1, ..., P_2, is called a model group. For SEPH, $k \in s_2$, model groups are sets of SIC3 by province combinations. The regression is carried out within each of these model groups, by regressing the independent variable $y_{k,E}$ (summarized earnings, or hours) on $x_{2k,E}$ (payrolls, employment). The following picture is a representation of the interaction between the sampling and estimation processes:



The estimator of total $Y(d) = \sum_{U_d} y_k$, for domain U_d, will be denoted as $\hat{Y}(d)$. The following steps are used for estimating the total $Y(d)$.

1. Compute overall weights reflecting the sampling at each phase and the auxiliary data at phase one

$$w^*_{1k} = w^*_k \, g_{1k} \tag{3.1}$$

where $g_{1k} = N_{p_1} / \hat{N}_{p_1}$ and $\hat{N}_{p_1} = \sum_{s_{1p_1}} w_{1k}$ is the first-phase poststratification blow up, with the overall sampling weight being

$$w^*_k = \frac{1}{\pi'_k} = \left[\frac{N_r}{n_{1r}}\right] \left[\frac{n_{1rh}}{n_{2rh}}\right], \; k \in s_{2rh}.$$

2. For each model group p_2, regress $y_{k,E}$ (with phase 2) on $x_{2k,E}$ where $k \in s_{2p_2}$, obtaining the estimated regression vector for model group p_2 as

$$\hat{B}_{s_{2p_2}} = \left[\sum_{s_{2p_2}} \frac{\tilde{w}^*_{1k} x_{2k,E} x'_{2k,E}}{\sigma^2_k}\right]^{-1} \sum_{s_{2p_2}} \frac{\tilde{w}_{1k}{}^* x_{2k,E} \, y_{k,E}}{\sigma^2_k} \tag{3.2}$$

where σ^2_k is a scaling factor for each observation $k \in s_{2p_2}$. Predict y_k for phase one units belonging to model group p_2 as: $\hat{y}_{k,A} = x'_{2k,A} \hat{B}_{s_{2p_2}}$ for $k \in s_{1p_2}$. Assign to each predicted value $\hat{y}_{k,A}$ a domain value $\hat{y}_{k,A}(d)$:

$$\hat{y}_{k,A}(d) = \begin{cases} \hat{y}_{k,A} & \text{if } k \in U_d \\ 0 & \text{otherwise} \end{cases}$$

This can also be viewed as assigning to each $k \in s_{2p_2}$ a domain value $x_{2k,A}(d)$ where

$$x_{2k,A}(d) = \begin{cases} x_{2k,A} & \text{if } k \in U_d \\ 0 & \text{otherwise} \end{cases}$$

resulting in $\hat{y}_{k,A}(d) = x'_{2k,A}(d) \hat{B}_{2s_{p_2}}$ for $k \in s_{1p_2}$ (part of phase one sample belonging to model group p_2).

3. The estimator of total for a given domain U_d is obtained as

$$\hat{Y}_{SYN}(d) = \sum_{r=1}^{R} \sum_{h=1}^{H_{1r}} \sum_{s_{1rhd}} \tilde{w}_{1k} \, \hat{y}_{k,A} \tag{3.3}$$

with $\tilde{w}_{1k} = w_{1k} \, g_{1k}$, s_{1rhd} is the part of the sample s_1 (first-phase) belonging to region r, stratum h, and domain U_d (which may be equal to h).

We now go on to write $\hat{Y}_{SYN}(d)$ as the sum of the product of the overall sampling weight w^*_k and factors that incorporate the auxiliary information. These factors will be labeled as g-factors. Now,

$$\hat{Y}_{SYN}(d) = \sum_{r=1}^{R} \sum_{h=1}^{H_{1r}} \sum_{p_2=1}^{P_2} \left[\sum_{s_{1rhp_2d}} \tilde{w}_{1k} x'_{2k,A} \hat{B}_{s_{2p_2}}\right]$$

$$= \sum_{p_2=1}^{P_2} \left[\sum_{s_{1p_2d}} w_{1k} \, g_{1k} x'_{2k,A} \hat{B}_{s_{2p_2}}\right] \tag{3.4}$$

where $s_{1rhp_2d} = s_{1rhd} \cap s_{p_2}$ $\quad s_{1p_2d} = \bigcup_{r=1}^{R} \bigcup_{h=1}^{H_{1r}} s_{1rhp_2d}$ and $g_{1k} = N_{p_1} / \hat{N}_{1p_1}$ for $k \in U_{p_1}$. Therefore, $\hat{Y}_{SYN}(d)$ can be

expressed as the sum of its model group components $\hat{Y}_{SYN,p_2}(d)$, that is, $\hat{Y}_{SYN}(d) = \sum_{p_2=1}^{P_2} \hat{Y}_{SYN,p_2}(d)$ with

$$\hat{Y}_{SYN,p_2}(d) = \sum_{s_{2p_2}} \left[\sum_{s_{1p_2d}} \tilde{w}_{1k} x'_{2k,A} \right] \tag{3.5}$$

The g-factored version for $\hat{Y}_{SYN,p_2}(d)$ is given by $\hat{Y}_{SYN,p_2}(d) = \sum_{s_{2p_2}} w_k^* g_{1k} g_{2k}(d) y_{k,E}$ where the g-factor arising from the second-phase of sampling is

$$g_{2k}(d) = \left(\sum_{s_{1p_2d}} \tilde{w}_{1k} x_{2k,A} \right)' \left(\sum_{s_{2p_2}} \frac{\tilde{w}_{1k}^* x_{2k,E} x'_{2k,E}}{\sigma_k^2} \right)^{-1} \frac{x_{2k,E}}{\sigma_k^2} \tag{3.6}$$

for $k \epsilon s_{2p_2}$ and $k \epsilon U_d$ (the domain of interest). Hence, for each domain U_d, there will be n_{2rh} different g-factors for phase 2, in stratum s_{2rh}. These g-factors share the same $\left(\sum_{s_{1p_2d}} w_{1k} g_{1k} x_{2k,A} \right)'$ value, but differ in the $\left(\sum_{s_{2p_2}} \tilde{w}_{1k}^* \frac{x_{2k,E} x'_{2k,E}}{\sigma_k^2} \right)^{-1} \frac{x_{2k,E}}{\sigma_k^2}$ value. Hence, the estimator of total for a domain U_d can be written as

$$\hat{Y}_{SYN}(d) = \sum_{p_2=1}^{P_2} \sum_{s_{2p_2}} w_k^* g_{1k} g_{2pk}(d) y_{k,E} = \sum_{p_2=1}^{P_2} \sum_{s_{2p_2}} \tilde{w}_k^*(d) y_{k,E} \tag{3.7}$$

where the domain dependent g-factors from phase 2 are given by (3.6) and $\tilde{w}_k^*(d) = w_k^* g_{1k} g_{2k}(d)$ for $k \in s_{2p_1} \cap s_{2p_2}$. The estimated variance across all strata is $v\left[\hat{Y}_{SYN}(d)\right] = \sum_{r=1}^{R} \sum_{h=1}^{H_{1r}} v\left[\hat{Y}_{SYN,rh}(d)\right]$. Here, $v\left[\hat{Y}_{SYN,rh}(d)\right]$ is the variance estimate at the stratum level. It is given by:

$$v[\hat{Y}_{SYN,rh}(d)] = N_r^2 \frac{1-f_{1r}}{n_{1r}} \omega_{1rh} \left[(1-\delta_{rh}) S^2_{ts_{2rh}}(d) + \frac{n_{1r}}{n_{1r}-1} (\bar{t}_{s_{2rh}}(d) - \hat{t}_{U_{2r}}(d))^2 \right] + N_r^2 \omega_{1rh}^2 \frac{1-f_{2rh}}{n_{2rh}} \sum_{s_{2rh}} \frac{(g_{2k}(d) e_{2k,E} - \overline{ge}_{2rh}(d))^2}{n_{2rh}-1} \tag{3.8}$$

where $f_{1r} = \frac{n_{1r}}{N_r}$; $f_{2rh} = \frac{n_{2rh}}{n_{1rh}}$;

$e_{2k,E} = y_{k,E} - x'_{2k,E} \hat{B}_{s_{2p_2}}$ for $k \epsilon s_{2p_2}$;

$\overline{ge}_{2rh}(d) = \frac{1}{n_{2rh}} \sum_{s_{2rh}} g_{2k}(d) e_{2k,E}$;

$\delta_{rh} = \frac{1}{n_{2rh}} \left[\frac{n_{1r} - n_{1rh}}{n_{1r}-1} \right]$; $\omega_{1rh} = \frac{n_{1rh}}{n_{1r}}$;

$\hat{t}_k(d) = g_{1k} e_{1k,E}(d)$; $\bar{t}_{s_{2rh}(d)} = \frac{1}{n_{2rh}} \sum_{s_{2rh}} \hat{t}_k(d)$;

$e_{1k,E}(d) = y_{k,E}(d) - \bar{y}_{p_1,E}(d)$ for $k \epsilon U_{p_1}$;

$\bar{y}_{p_1,E d} = \frac{1}{\sum_{s_{2p_1}} w_k^*} \sum_{s_{2p_1}} w_k^* y_{k,E}(d)$;

$\bar{t}_{U_{2r}}(d) = \sum_{h=1}^{H_{1r}} \omega_{2rh} \bar{t}_{s_{2rh}}$; and

$S^2_{ts_{2rh}}(d) = \frac{1}{n_{2rh}-1} \sum_{s_{2rh}} (\hat{t}_k(d) - \bar{t}_{s_{2rh}}(d))^2$.

Remark 3.1: $\hat{Y}_{SYN}(d)$ is a pure synthetic estimator. No correction involving the regression residuals were involved. Such an estimator is unbiased only if the model group relationships hold within each domain as well. An approximately unbiased version, conditional on the realized domain sample size, was given by Särndal and Hidiroglou (1989). The form of this estimator in the SEPH context would be:

$$\hat{Y}_{UNB}(d) = \sum_{r=1}^{R} \sum_{h=1}^{H_{1r}} \left[\sum_{s_{1rhd}} \tilde{w}_{1k} \hat{y}_{k,A} + \left(\frac{n_{1rd}}{\hat{n}_{1rd}} \right) \sum_{s_{2hrd}} \tilde{w}_{1k}^* g_{1k} \frac{e_{2k,E}}{w_k^*} \right]$$

where $\hat{n}_{1rd} = \sum_{s_{2rd}} 1/\pi_{k|s_{1rh}}$; $s_{1rd} = \bigcup_{h=1}^{H_{1r}} s_{1rhd}$. It should be remarked that $\hat{Y}_{UNB}(d)$ can "blow up" if $1 \le \hat{n}_{1rd} << n_{1rd}$. Särndal and Hidiroglou (1989) have provided a way to account for this. Note that if $\hat{n}_{1rd} = 0$, we use the $\hat{Y}_{SYN}(d)$ estimator.

4. CONCLUSION

The sampling and estimation methodology for stage one of the SEPH redesign has been discussed. This methodology has been operational since March 1994. It has significantly reduced response burden, as well as operational costs. Furthermore, many systems modules have been updated as a result of this redesign, yielding improvements to the overall processing of this survey. The next two stages will complete the full redesign for SEPH.

REFERENCES

Cuthill, I.E. (1989), "The Statistics Canada Business Register", *Proceedings of the Fifth Annual Research Conference*, Washington, D.C.: US Bureau of the Census, 69-86.

Dolson, D. (1994), "On Using a Very Current Source of Frame Information in Canada's Establishment-based Employment Survey", presented at *First Eurostat Workshop on Techniques of Enterprise Panels*, Feb. 21-23,1994, Luxembourg.

Hidiroglou, M.A., Latouche, M., Armstrong, B., and Gossen, M. (1995), "Improving Survey Information Using Administrative Records: The Case of the Canadian Employment Survey", *Proceedings of the Annual Research Conference*, 171-197.

Särndal, C.E., and Hidiroglou, M.A. (1989), "Small Domain Estimation: a conditional analysis", *Journal of the American Statistical Association*, 84, 266-275.

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

Schiopu-Kratina, I. and Srinath, K.P.(1991), "Sample rotation and estimation in the Survey of Employment, Payrolls and Hours," *Survey Methodology*, 17, 79-90.