SSC Annual Meeting, June 98
Proceedings of the Survey Methods Section

USE OF ADMINISTRATIVE RECORDS IN THE CANADIAN SURVEY OF EMPLOYMENT, PAYROLLS AND HOURS

Eric Rancourt[1] and Michel Hidiroglou[2]

ABSTRACT

The Survey of Employment, Payrolls and Hours is a monthly establishment survey. The survey covers all industries except agriculture, fishing and trapping, private household services, religious organisations and military services. Starting in 1993, two additional data items (payrolls and employment) for Payrolls Deduction accounts from Revenue Canada have become available. A major redesign was undertaken to use this wealth of information in order to reduce the sample size, the associated costs and the response burden, while maintaining the quality of the estimates. To achieve these goals, administrative records were used in a variety of ways. These include the creation of a survey frame for part of the survey; building of an improved design; the use in regression of the independent variables obtained from the payrolls source of Revenue Canada and quality assessment of the surveyed portion. This paper gives an overview of the redesign and presents the methodological aspects of the use of administrative records in the Survey of Employment, Payrolls and Hours.

KEY WORDS : Administrative Data; Complex Enterprise; Model Group; Two-Phase Estimation.

RÉSUMÉ

L'Enquête sur l'emploi, la rémunération et les heures est une enquête mensuelle d'établissements couvrant la majorité des activités industrielles au Canada. Commençant en 1993, deux items de données additionnels (la rémunération et l'emploi) sont devenus disponibles sur les comptes de déduction à la source de Revenu Canada. Un remaniement majeur de l'enquête a été entrepris afin d'utiliser cette source d'information pour réduire la taille de l'échantillon, le coût associé et le fardeau de réponse, tout en préservant la qualité des estimations. Pour réaliser cet objectif, on utilise les dossiers administratifs de plusieurs façons : comme base de sondage pour une partie de l'enquête, en utilisant dans la régression les variables indépendantes obtenues de Revenu Canada et pour l'évaluation de la qualité pour la partie enquêtée. Cet article donne un aperçu du remaniement et présente les aspects méthodologiques de l'utilisation de dossiers administratifs pour l'Enquête sur l'emploi, la rémunération et les heures.

MOTS-CLÉS : Dossiers administratifs; entreprise complexe; groupe modèle; estimation en deux phases.

1. INTRODUCTION

In carrying out surveys, methodologists and managers are always aiming at getting the maximum possible amount of information while attempting to keep costs and response burden at a reasonable level. The ultimate goal is to produce a high quality product with an effective survey design. A candidate for achieving these objectives is administrative data. At Statistics Canada, the interest in using administrative data is present throughout the Bureau as can be seen from the first of the four key strategies in Statistics Canada (1998), which is to use existing information.

There are several ways in which administrative data can be used. Brackstone (1987) categorises them into four areas: Direct tabulations, indirect estimation, survey frames and survey evaluation. In this paper, we will see how the Survey of Employment, Payrolls and Hours (SEPH) is a unique example of three of these. In SEPH, administrative data are used i) as a

[1] Eric Rancourt, Household Survey Methods Division, Statistics Canada, Ottawa (Ontario), Canada K1A 0T6.
[2] Michel Hidiroglou, Business Survey Methods Division, Statistics Canada, Ottawa (Ontario), Canada K1A 0T6.

frame for estimation of Employment and Payrolls; ii) for indirect estimation of Hours and Summarized Earning through regression; and iii) for evaluation of the quality of the establishment survey.

This paper is structured as follows. In Section 2, an overview of the Survey of Employment, Payrolls and Hours is presented. It is followed in section 3 by a description of the nature of the administrative data used in SEPH with explanations on how the redesign was undertaken. The new sample design (after the third phase of the redesign) is discussed in section 4, and estimation is given in section 5. Finally, quality assurance of the data is treated in section 6.

2. THE SURVEY OF EMPLOYMENT, PAYROLLS AND HOURS (SEPH)

The Survey of Employment, Payrolls, and Hours (SEPH) is a monthly survey carried out by Statistics Canada. It collects data on employment, payrolls, working hours, overtime pay and hours, summarised earnings and 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) levels for Canada, and the provinces. The SIC2 are the major fields of activity while SIC3 is a further detailed level. In all, there are close to one million establishments in scope for the survey.

SEPH was designed at the beginning of the 1980's as a stratified sample of establishments (Schiopu-Kratina and Srinath, 1991). At that time, the sample was designed to obtain expected coefficients of variation (CV) of 3% for the estimated number of employees at the provincial-industrial (SIC3) levels. The estimation procedure was strictly based on the survey design weights and the data directly collected from the sampled establishments. That is, no auxiliary information was used in estimation.

This sampling design was recently changed because of the new availability of administrative data from Revenue Canada. The availability of this new administrative data source from Revenue Canada led to a redesign that was spread over three phases, and spanned over a period of four years.

3. ADMINISTRATIVE FILES AND THE REDESIGN OF SEPH

Enterprises are required to remit to Revenue Canada all the amounts of deduction retained from the employees' salaries and wages. Enterprises send their remittances to Revenue Canada for each list of employees, using a global *Payrolls Deduction Account* or by having one for each of their lists of employees. This results in a complete file of all Payrolls Deduction accounts (PD), containing remittances, being available monthly from Revenue Canada. There are two types of PD accounts: automatic (A) and twice monthly (TM). "A" accounts represent once monthly remitters whose average monthly remittances in the previous year were less than \$15, 000. TM accounts represent remitters whose average monthly remittances in the previous year were \$15,000 or greater. TM's remit one to four times a month depending on the type of payrolls used (i.e., monthly, weekly, biweekly, etc.) on an accelerated schedule.

In 1993, Revenue Canada made an agreement with Statistics Canada by which they would be collecting the number of employees and the total amount of payrolls (besides the remittances). Since these two variables were already collected by SEPH, those same variables from Revenue Canada became an excellent source of auxiliary information. Several studies (Lee and Croal 1991; St-Martin 1992), were carried out to evaluate this new information, and it was concluded that SEPH would greatly benefit from using them. Hence the redesign.

The goals of the redesign were multiple. The primary objectives were to:

- Reduce the costs of the survey;
- Reduce response burden;
- Maintain quality of the statistics produced;

And the secondary objectives were to:

- Improve processing and systems;
- Increase the ability to produce small-area estimates.

To achieve these objectives, a plan was designed to extensively use the administrative data. It was implemented in three phases, spanning over a four-year period. More information on this redesign is available in Dolson (1995) and Hidiroglou, Latouche, Armstrong and Gossen (1995). The phases were determined based on the size and complexity of the enterprises on Statistics Canada's Business Register.

The size is the number of employees present on the Business Register; and the complexity is based on the number of establishments and their areas of activity. We illustrate the various structural complexities of the enterprises using definitions and examples. An enterprise is said to be *simple* if it is present in only one province and only one sector of industrial activity (SIC3). An enterprise is said to be *complex* if it is present in more than one province and / or sector of industrial activity (SIC). A simple enterprise can be *single,* with only one establishment as shown in Figures 1 and 2, or *multi,* with more than one establishment as in Figure 3, but with activity in more than one industry or province. In the case of a complex enterprise, it is by definition a multi as shown in Figure 3. Therefore, Figure 2 represents a multi enterprise that it has two establishments; note that it is simple because activity occurs in only one area.

In 1997, a new Business Number BN was introduced for each enterprise to use as a single identification to report to Revenue Canada. For consistency of the arguments, this paper always refers to PDs, but the reader should note that although Phase I and II of the redesign were based on the PDs, phase III is completely based on the new BN.

Figure 1. Simple and Single Enterprise

Enterprise 1

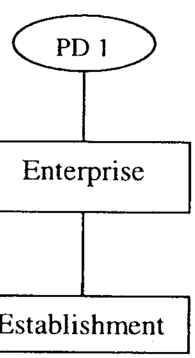


Figure 2. Simple and Multi Enterprise

Enterprise 2

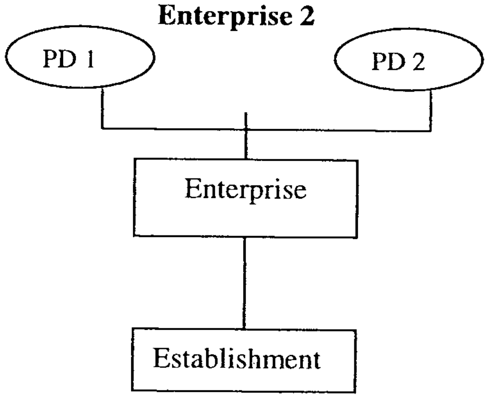
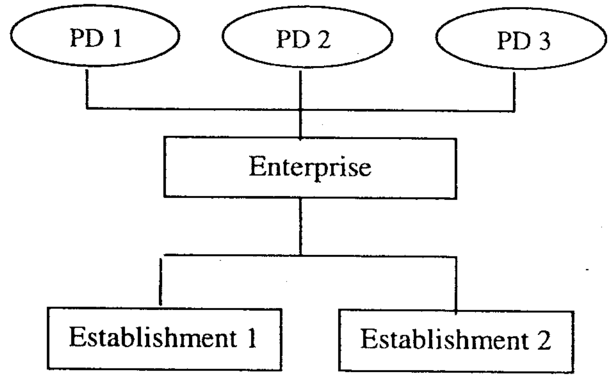


Figure 3. Complex and Multi Enterprise

Enterprise 3



In phase I of the redesign, the population was divided into two portions: one with single enterprises having less than 100 employees (called the small portion) and the other remaining enterprises (called the large portion). The small portion represented about 30 % of the total employment in Canada. In the large portion, the survey remained the same as the one in place before the redesign. However, the small portion was completely based on a sample of PD's to estimate the number of employees and payrolls (See Section 5.1 for more details). Estimation of the other variables (i.e. hours and summarised earnings) was based on regression predictions. For these variables, a sub-sample of PD sample (establishment sample) was selected and directly surveyed. The data collected via the establishment sample were used to predict the variables not available in the small portion. Estimation methods for these variables are presented in Section 5.2.

In phase II, the use of administrative data was extended to all simple enterprises and to the complex enterprises having less than 300 employees. This new and larger small portion represented about 70 % of the total employment in Canada. Although there were some complex enterprises in the new portion, no special treatment was developed for them since they accounted only for a small portion of the employment estimate. In phase II, estimation was based on the same methods as in phase I. They are described in Hidiroglou, Latouche, Armstrong and Gossen (1995).

In phase III, the small portion was extended to all enterprises. Hence only the small portion remains. In phase III, except for the presence of a number of take-all units (see Section 4.3), the estimation process remains essentially the same as in phase I and II. It is described in details in Section 5.

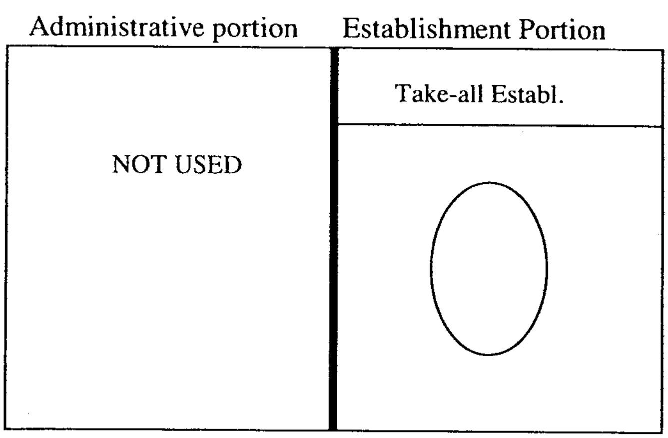
4. SAMPLE DESIGN

In SEPH, the first use of administrative records is as a frame for employment and payrolls. In this section, the design of the new (phase III) survey is presented with particular attention paid to the use of PD's as a frame for producing estimates of employment and payrolls. Aspects of the design, which are also necessary for estimating the regression parameters, are given in Section 5.2.

4.1 Frame

All enterprises, which are present on Statistics Canada's Business Register (BR), constitute the SEPH frame. To be in-scope, an enterprise must be an employer (have at least one employee) and be active. Due to the dynamic nature of the enterprises, the frame is affected by changes between sampling time and estimation time. Changes to the frame are classified into two main types. They are: (1) contact and classification changes, such as telephone contact number and address, activity status, industry or geographical; and (2) structural, changes that occur to the composition of the establishments within the enterprise.

Figure 4. SEPH Frame and samples before the redesign



When structural frame changes are detected during data collection, they are not immediately applied to SEPH during the particular survey cycle. It is the Business Register Division that updates the enterprises on the frame. Frame changes are either brought to the attention of the Business Register by the survey, or by independent sources (feedback from other surveys, or changes in the payrolls deduction account status, such as births or deaths). All of these changes (except for in-sample industry and geography) are applied to the SEPH frame whenever the Business Register has been updated, and it feeds them back to the survey process. However, contact and changes in activity status are immediately reflected in the survey estimates for the particular survey cycle.

Figure 5. SEPH Phase I Frame and samples

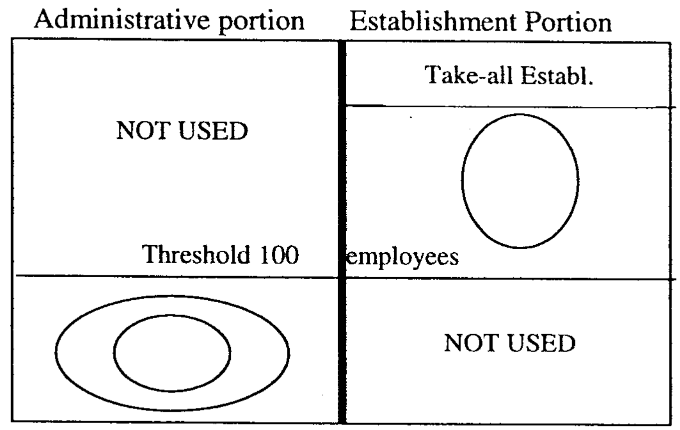


Figure 6. SEPH Phase II Frame and samples

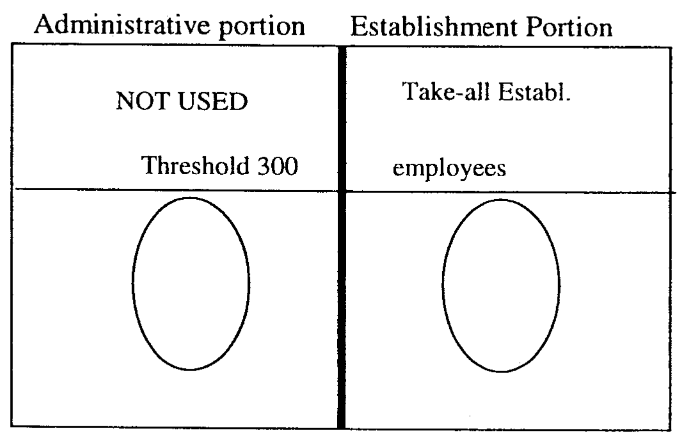
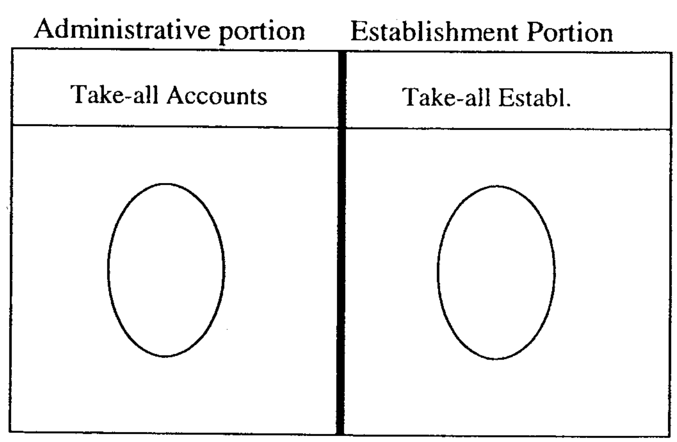


Figure 7. SEPH Phase III Frame and samples



Payroll accounts are linked to all enterprises of the Business Register. The complete SEPH frame can be viewed as being composed of records with PD account data, (the administrative portion) and establishment data, (the establishment portion). Figures 4-7 show the three complete frames as well as samples for each of the phases of the redesign.

In the sample drawn from the administrative portion, only employment and payrolls are reported. In addition to employment and payrolls, the establishment sample drawn from the establishment portion (Business Register) collects hours and summarised earnings, as well as a number of detailed categories (e.g., part time, full time workers).

4.2 Stratification

In the administrative portion, stratification is based on the provinces and the structure of the enterprises. That is, all enterprises with more than one PD are selected (pre-specified) to be in the sample, and the remaining ones are selected within provinces.

In the establishment portion, stratification is based on the Standard Industrial Coding (SIC), the province and the size (number of employees) of the enterprise. Large enterprises whose number of employees is larger than a specified limit are called Take-All units (TA). These units are selected with certainty. They are further grouped into two sets: the extreme ones (TA1) and the others (TA2). The TA1 take-all units are not included in the regression fits, while the TA2 take-all units are included in the regression fits. The remaining units are assigned to strata based on groups of industries, groups of provinces and size (0-19 employees, 20 or more).

4.3 Regression groups

To ensure that the regression is most efficient, the population is divided into sub-populations within which the fits can be achieved. These sub-populations called *regression groups* are built from unions of strata. The regression groups are used at the estimation stage where they form the level at which regressions are performed. All units in the administrative and the establishment portions are assigned to a regression group. The establishment and administrative samples are partitioned into G_E groups for the establishment portion and G_A groups for the administrative portion. The samples are then

$$s_E = \bigcup_{g=1}^{G_E} s_{E,g} \text{ for the establishment sample and}$$

$$s_A = \bigcup_{g=1}^{G_A} s_{A,g} \text{ for the administrative sample.}$$

This partition of both the establishment and administrative universes into regression groups is based on combinations of industries (SIC3) and geography (regions). It should be noted that industrial activity is more discriminatory than geographic. As well, industries do not necessarily have the same breakdowns by region.

Since they are large (even extreme), the TA1 take-all units are not assigned to regression groups, as they could affect the regressions otherwise. Rather each of them is assigned to its own unique regression group.

4.4 Allocation in the establishment portion

Allocation of the establishment sample is performed in two steps. First, take-all units are determined and then the sample is allocated to the remaining portion.

4.4.1 Take-all units

The first step of allocation in SEPH is to determine the number of TA units of each type: Type meaning whether or not they participate in the regression fits. This is achieved by using a procedure known as the *sigma-gap* method. This procedure is carried out as follows. Let $x_{(k)}$ be the k^{th} ordered value. A unit k is declared a take-all unit with respect to variable x if

1) $x_{(k)} > x_M$ where x_M is the median; and

2) $x_{(k)} - x_{(k-1)} > a\sigma_x$ where σ_x is the standard deviation and a is a specified constant.

3) All units such that $x_{(k')} > x_{(k)}$ are also take-all units.

The sigma-gap method that is used in the survey is a slightly modified version of the above method in that the 5% largest records are not used to compute the sigma (σ) value. In SEPH, $a = 7$ for TA1 units and $a = 0.7$ for TA2 units. Further details are available in Lim and Daoust (1997).

4.4.2 Non take-all units

For the non take-all units, allocation is designed to achieve targeted coefficients of variation (CV's) at the provincial and industrial levels. A raking procedure is applied to obtain the required industry by province marginal CV's. The sample size within these industry by province strata are computed to satisfy these CV's. Each province by industry stratum is further stratified into size groups. The allocation to the size groups (from the industry by province levels) is obtained by means of power allocation. The allocation steps for SEPH are described in details in Hidiroglou, Latouche, Armstrong and Gossen (1995), Lim and Rancourt (1996) and Lim and Daoust (1997). An interesting feature of the method lies in that it makes use of sampling variance from both establishment and administrative sources (Section 5.2).

4.5 Allocation in the administrative portion

In the administrative portion, the sample was originally allocated to satisfy processing constraints in terms of volume. To that end, a smaller fraction of units was selected in the large provinces. Yukon and the North Western Territories have a sampling fraction of 100%. The Quebec and Ontario sampling fraction is 10% and that of the other provinces is 20%.

4.6 Sampling

The *administrative* sample is designed as a panel. Although deaths are deleted and births are added, no rotation of this sample takes place. The sampling of the PD accounts uses a variation of Bernoulli sampling. The sample contains all PD account numbers having a pre-specified check digit sampling (the last 2 digits of the PD number are used to control the sampling). Data Total employment and total payrolls are captured by Revenue Canada. The number of sampled PD forms is in the neighbourhood of 200,000 per month. In the future, this number will be increased to cover the full PD universe, representing some 1,000,000 accounts.

The establishment sample is made up of approximately 10,000 establishments per month and a rotation of approximately one twelfth of the sampled units occurs each month.

5. ESTIMATION

In this section, we describe estimation methods used in phase III of the SEPH redesign. The estimation methods in Phases I can be found respectively in Dolson (1994a), Dolson (1994b), and Hidiroglou (1995), while those for Phase II are in Hidiroglou, Latouche, Armstrong and Gossen (1995).

5.1 Direct estimation

In this section, we provide the estimation procedure for administrative records. That is, for the case of estimation of employment and payrolls.

Let $U_A = \{1,\dots,k,\dots,N\}$ be the administrative population. We are interested in estimation of $X_A = \sum_U x_k$, the total for variable X. Since the sample is selected as Bernoulli trials with probability π_k, an estimator of X_A could be $\hat{X}_A^{(0)} = \sum_{h=1}^{H} \sum_{s_{Ah}} \frac{x_k}{\pi_k}$.

A more efficient estimator that uses both population and realised sample counts is

$$\hat{X}_A^{(1)} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{s_{Ah}} x_k$$

Since the population counts are available for size groups within the provinces for the administrative portion, a better estimator is the post-stratified estimator

$$\hat{X}_A^{(s)} = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{p=1}^{P} \frac{N_p}{\hat{N}_p} \sum_{s_{Ah}} x_k, \text{ where } \hat{N}_p = \sum_p \frac{N_h}{n_h}.$$

This is the estimator used to estimate the totals for the variables available from the administrative records.

5.2 Indirect estimation

The establishment sample is used for obtaining a model-assisted approach for the estimation of totals for variables not collected by the PD sample. Models involving linear regression are estimated within each regression group. Total employees and total payrolls for the month are the independent variables, while total hours and summarised earnings are the dependent variables. Ratios of total hours by category of employees to total hours are also estimated. Total employees and total payrolls reported on the administrative sample are used to construct totals of auxiliary variables.

Estimated parameter values from the regression are used to predict total hours for each unit in the model group. That is, for each unit the missing variable is imputed using mass imputation, as described for example in Kovar and Whitridge (1995). However, the estimation process can be analysed without having to consider it to be an imputation process.

The estimation process may be viewed as a two-step regression procedure with stratification being applied at the two levels of sampling. In the first step, the auxiliary population totals are derived. For a given domain U_d, the totals (for employment and payrolls) are obtained through estimation from the administrative sample s_A as

$$\hat{X}_A(d) = \Sigma_{s_A} \tilde{w}_{Ak}\, x_{A,k}(d),$$

where $\tilde{w}_{Ak} = \{N_p / \hat{N}_P\} w_{Ak}$ and $w_{Ak} = N_h / n_h$. This estimator is in fact the post-stratified estimator.

In the second step, the estimated regression coefficients are obtained from the fit between the available variables on the establishment sample, for each regression group. This yields the following estimator for the predicted y variables for domain U_d regressed on the x variables of the establishment sample, $\mathbf{x}_{E,k}$

$$\hat{B}_{E,g} = \left(\Sigma_{s_{E,g}} w_{E,k} \mathbf{x}_{E,k} \mathbf{x}'_{E,k} / \hat{\sigma}^2_{E,k}\right)^{-1} \Sigma_{s_{E,g}} w_{E,k} \mathbf{x}_{E,k}\, y_k / \hat{\sigma}^2_{E,k}.$$

The values $\mathbf{x}_{E,k}$ are the data corresponding to the $\mathbf{x}_{A,k}$ data from the administrative source, and $\hat{\sigma}^2_{Ek}$ is a variance factor with homogeneous residuals. The dependent variables in these fits are either hours or summarized earnings. The maximum number of independent variables included in the fit is two: payrolls and/or number of employees. No intercept term was included as the presence of an intercept was found to increase the possibility of obtaining negative regression weights.

Finally, the predicted variables are produced by multiplying x'_{Ak} by $\hat{B}_g$, leading to

$$\hat{Y}(d) = \sum_{g=1}^{G} \sum_{s_{Ag}} \tilde{w}_{Ak}\, x'_{A,k}(d)\, \hat{\mathbf{B}}_{Eg}.$$

The estimated variance for $\hat{Y}(d)$ is estimated as a product estimator. The estimated variance is made up of two components: one due to the post-stratified estimator for the administrative sample, and the other due to the predicted regression fit (yielding a synthetic estimator) from the establishment sample. Work on these estimators can be found in Hidiroglou, Latouche, Armstrong and Gossen (1995) and Lim and Daoust (1997). Again, the process can be viewed as one of imputation. In this case, instead of having two sampling sources (the administrative and the establishment), there is a sampling source (the administrative sample) and an

imputation source of variation (from the establishment sample where regression is calculated for imputation). See Deville and Särndal (1994), for estimation in presence of regression imputation.

6. DATA QUALITY

The third way in which administrative records have been used in SEPH is for assessment of data quality. This work is two-fold: Firstly, care must be taken in processing administrative records due to their particular nature. Secondly, administrative data are used as quality diagnostics through comparisons both at the macro and micro levels.

6.1 Processing issues

Production of monthly administrative data

Since SEPH is a monthly survey, it would be desirable that the administrative source report their data in a monthly fashion as well. However, as stated in Section 3, the TM Revenue Canada data can represent any pay period (weekly, bi-weekly, etc.) within a reference month, or even across months that include the reference month. Hence the remitted data is transformed to monthly data. Extensive sets of edit rules and transformations were developed to process the TM data. The resulting system is a complete survey processing system that includes outlier detection, editing, imputation and re-weighting. Table 1 shows the distribution of the various types of pay periods used by enterprises.

Table 1. Distribution of the pay periods

Frequency	Amount	Percentage
Weekly	32500	13.0%
Bi-weekly	48750	19.5%
Monthly	162500	65.0%
Other	6250	2.5%

As can be seen, 35% of the records need to be adjusted to produce a complete file of usable monthly data.

Data from external source

Data provided by the external source (Revenue Canada) to SEPH are similar but may not exactly conform, to the concepts and definitions used by the survey. This implies that several of the data processing steps cannot easily be controlled because they are out of scope to the SEPH system. First, since data collection is carried out outside Statistics Canada, no edits (of our own) can be applied at that point. Secondly, our objectives and those of the administrative record producer will not necessarily be the same. As a result, changing priorities may not always be reflected by these data. The key to avoiding these problems (or minimising them) is to be in constant contact with the administration.

Because the control over the data collection procedures cannot be as extensive as desired, supplementary validation steps must be established. In SEPH, in addition to the usual edit-imputation-outlier detection-processing stream, there is also a module aimed at detecting unusual and suspect "big changes" from one month to the next for the administrative data. This is performed univariately on a series of variables, and a general score is assigned to each unit, based on the univariate differences. Those with a score beyond a limit are reviewed and corrected if necessary.

Concordance of regression groups

In the two-step regression estimation approach used in SEPH, it is assumed that the regression models in the administrative and establishment portion are the same for both portions. At the start of the redesign, studies supported this hypothesis. However, this may change with time, and it should be re-assessed periodically.

6.2 Quality monitoring

This activity was performed at two levels: the macro and micro levels.

At the macro level, the administrative employment and payrolls totals provided a convenient source of information to assess the quality of the corresponding estimation produced from the establishment portion.

At the micro level, a number of studies could be performed during the redesign, and they should still be carried out on a regular basis to monitor the quality of the establishment sample.

The first comparison study was the micro match, where the establishment data were compared to the administrative data (through the enterprise) at the unit level. The greatest differences stemming out of the comparison were investigated. Some of the respondents had to be contacted to verify, and in some cases revise the data. Comparisons of imputed data versus administrative records were also very informative on the quality of the processing methods used. It led to several improvements of the imputation system.

7. CONCLUSION

In this paper, we have presented three uses of administrative records for the Survey of Employment, Payrolls and Hours. Administrative records have contributed in significantly improving the efficiency of the survey through a major three-phase redesign. The new survey now uses administrative records as the basis for estimation of employment and payrolls and as a benchmarking tool for the other variables. Major savings were made by using this available wealth of information.

In the future, more and more administrative records will become available, and the design presented here will hopefully serve as a model framework for other applications.

8. REFERENCES

Brackstone, G. J. (1990). Issues in the use of administrative records for statistical purposes. *Survey Methodology*, 16, 29-43.

Dale, V. (1998). Survey of Employment, Payrolls and Hours (SEPH): An overview of phase III systems and processes. Benchmarking and Administrative Data Section document, Labour Division, Statistics Canada.

Deville, J.-C. and Särndal C.-E. (1994). Variance estimation for the regression imputed Horvitz-Thompson estimator. *Journal of Official Statistics*, 10, 381-394.

Dolson, D. (1994a). A summary of the estimation methodology of the Survey of Employment, Payrolls and Hours (January 1983 to February 1994). Working Paper BSMD-94-003E, Methodology Branch, Statistics Canada.

Dolson, D. (1994b). On using a very current source of frame information in Canada's establishment based employment survey, *Proceedings of the First Eurostat Workshop on Techniques of Enterprise Panels*, Luxembourg, 280-292.

Dolson, D. (1995). On redesigning Canada's establishment based employment survey. *Proceedings of the International Conference on Establishment Surveys*, 793-798.

Gervais-Simard, L. (1996). Microdata matching: A tool for evaluating and improving the quality of survey data. *Proceedings of Statistics Canada Symposium 96: Nonsampling errors*, Statistics Canada, 167-177.

Gossen, M. and Latouche, M. (1992). Post-stratification to reduce sample bias in an establishment survey. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 209-214.

Hidiroglou, M.A (1995). Sampling and estimation for stage one of the Canadian Survey of employment, payrolls and hours survey redesign. *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 123-128.

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 1995 Annual Research Conference*, US Bureau of the Census, 171-197.

Kovar, J.G., and Whitridge, P.J. (1995). Imputation of business survey data. In *Business Survey Methods*, Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., Colledge, M.J. and Kott, P.S. eds, 403-423, J. Wiley and Sons.

Lee, H. and Croal, J. (1991). Use of PD-7 data for estimation in the Survey of Employment, Payrolls and Hours. Working Paper BSMD-91-009E, Methodology Branch, Statistics Canada.

Lim, A. and Daoust, P. (1997). Allocation of the sample for the stage III of the redesign of the Survey of Employment, Payrolls and Hours (SEPH). Internal document. Business Survey Methods Division, Statistics Canada.

Lim, A. and Rancourt, E. (1996). Allocation of the sample for the stage II of the redesign of the Survey of Employment, Payrolls and Hours (SEPH). Internal document. Business Survey Methods Division, Statistics Canada.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer-Verlag.

Statistics Canada (1998). *1998-1999 estimates: A report on plans and priorities*, Section III.

St-Martin, P. (1992). Application of the General Regression Technique to improve estimates in an establishment survey. *Proceedings of the 1992 Annual Research Conference*, US Bureau of the Census, 225-244.