

68N0003XPB

no. 69

c. 2



**Project to Improve Provincial  
Economic Statistics**

**Projet d'amélioration des statistiques  
économiques provinciales**

**Sampling Designs for  
The Unified Enterprise  
Surveys:**

**Les plans d'échantillonnage  
des enquêtes unifiées sur  
les entreprises :**

**The Early Years**

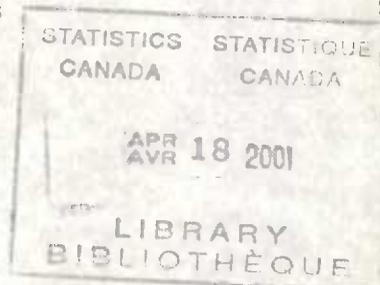
**Les premières années**

Technical Series

Série technique

Number 69

Numéro 69



Internet: [www.statcan.ca](http://www.statcan.ca)

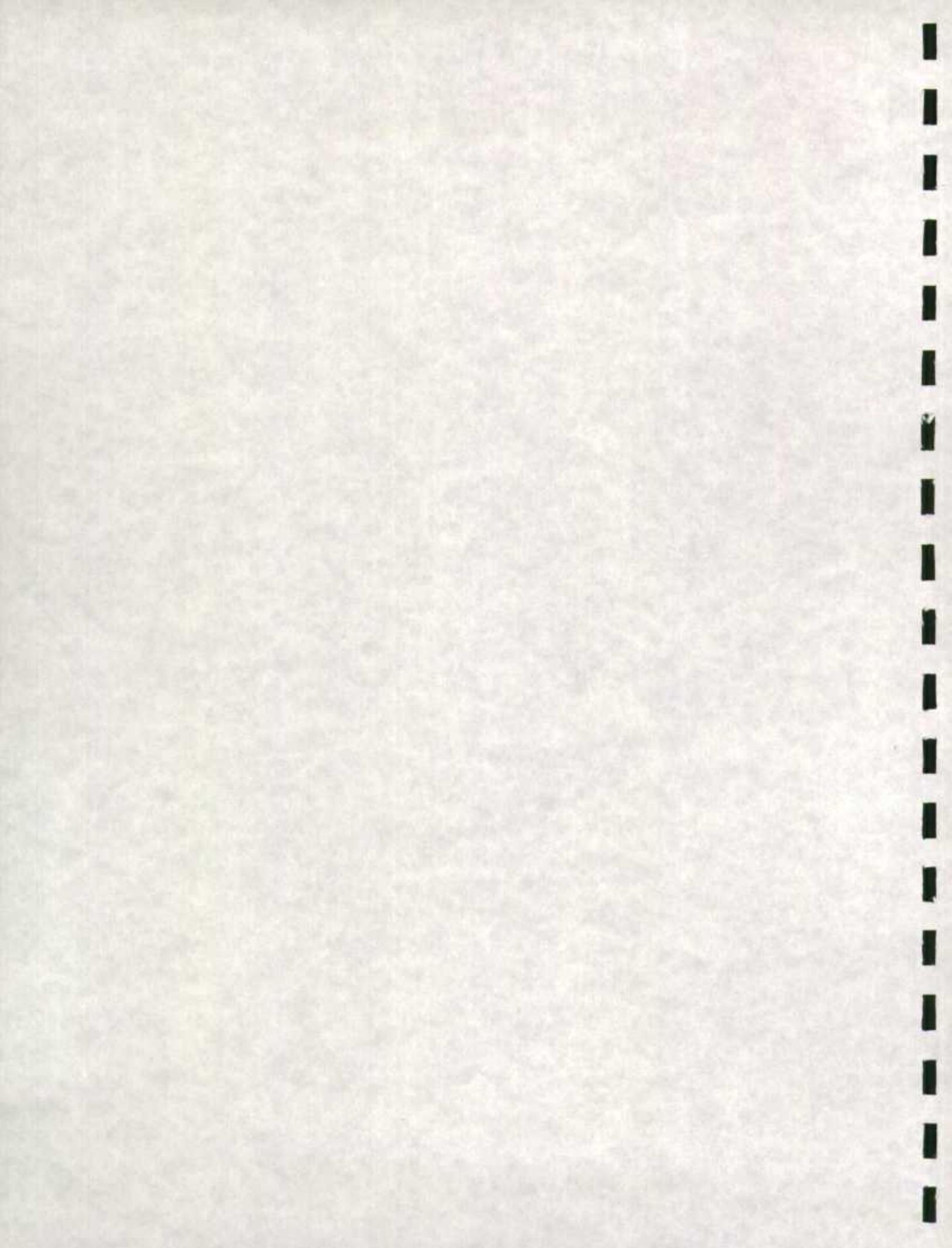
Intranet: <http://pipes>



Statistics  
Canada

Statistique  
Canada

**Canada**



# Sampling Designs for The Unified Enterprise Surveys:

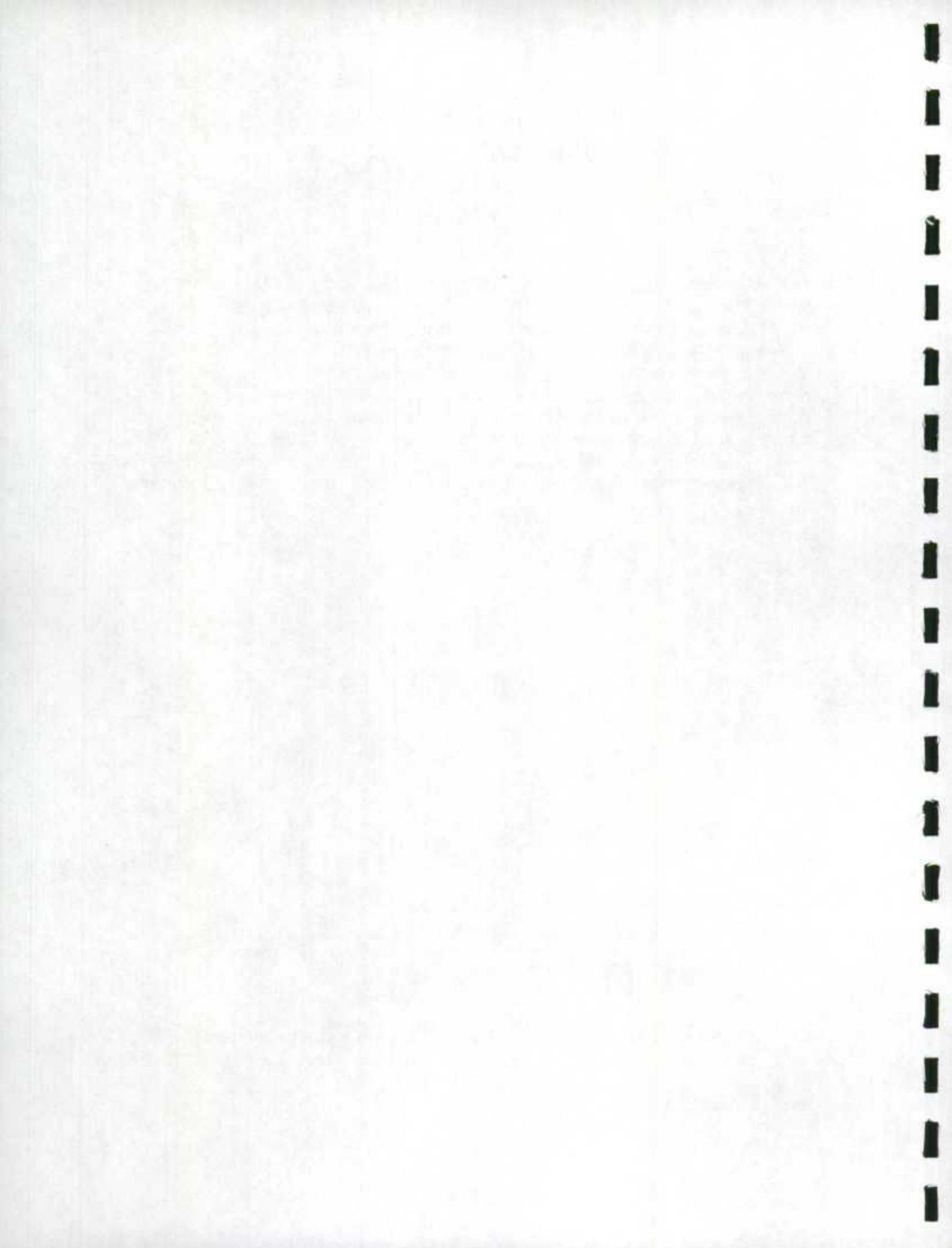
## The Early Years

The new Unified Enterprise Survey (UES) integrates many of Statistics Canada annual surveys. This survey has been modified annually since its implementation in 1997. Some survey concepts and methods used in the sampling cycles remain substantially the same from year to year. These include the use of a single frame, a two-phase approach, the stratification algorithm, the use of random numbers for selection and the sample allocation, but some have been modified. Among these, a new method of determining the coefficient of variation for multi-purpose surveys has been developed for reference year 1999. Also, for UES 2000, new strategies have been developed; one to meet different industry specifications in term of precision and the second for controlling response burden.

### Note of appreciation

*Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.*

For further information on the materials covered in this paper, please contact Bonnie Bercik (613) 951-6790 or Diane Proulx (613) 951-7192  
Fax: (613) 951-0411



# Sampling Designs for The Unified Enterprise Surveys:

## The Early Years

Michelle Simard, Claude Girard, Marie-Noëlle Parent, Jocelyn Smith  
Business Survey Methods Division

March 2001

### Table of Contents

Sections	Page
1 Introduction	3
2 Survey designs	3
3 The frame	5
3.1 Evolution of the frame coverage for UES	6
4 UES concepts	6
4.1 Cell concept	6
4.2 Sampling unit	7
4.3 Simple and complex enterprises	7
5 Stratification	8
5.1 Thresholds (Royce-Maranda) and Stratification	8
5.2 Stratification	9
5.3 Must-take units	9
6 Sample allocation	9
6.1 Sample inflation rates	9
7 Sample selection	11
8 Network sampling	11
8.1 Inclusion probabilities and design weight	12
9 Cell CV determination	13
9.1 UES 1997 and 1998	13
9.2 UES 1999: Raking ratio	13
9.2.1 First application	13
9.2.2 Second application	14
10 UES 2000: New proposal	15
10.1 Multiplication factor	15
10.2 Sample size	16
10.3 Power allocation	16
10.3.1 Introducing a Provincial exponent	16
10.4 UES 2000 CV allocation	17

11	Insignificant cell identification	17
11.1	Definition of criteria	18
11.2	Preliminary results	18
12	First-phase sample: Auxiliary information	20
13	Future direction	20
14	Acknowledgements	20
15	References	20

## Appendixes

1	Front-End files	22
2	Target Population, samples and exclusion thresholds	23
3	Sample size for UES 1997	25
4	Sample size for UES 1998	26
5	Sample size for UES 1999	27
6	Sample size for UES 2000	28

## List of tables and figures

		Page
Table 1	List of Surveys covered by UES - Survey and administrative data	4
Table 2	List of Surveys covered by UES - Administrative data only	4
Table 3	Evolution into the UES frame	6
Table 4	Illustration of a cell	7
Table 5	Actual values used for the exclusion	8
Table 6	Non-response rates for UES in percentage	10
Table 7	Frame imperfection rates for UES in percentage	10
Table 8	Sector definition	13
Table 9	Results of testing different percentile	18
Table 10	Insignificant cells by surveys	19
Table 11	Insignificant cells by provinces	19
Figure 1	Statistical entities available on the Business Register	5
Figure 2	Simple Structured Enterprise	5
Figure 3	Complex Structured Enterprise	6
Figure 4	Sampling Unit for UES	7
Figure 5	Strata defined within a cell	9
Figure 6	Before and after network sampling	12
Figure 7A	First step: allocating the global CV into sectorial CVs	14
Figure 7B	Second step: Sectorial raking	14

## **1 INTRODUCTION**

The Project to Improve Provincial Economic Statistics (PIPES) is beginning its fourth reference year for the year 2000. This project was initiated in 1996, when three provinces signed an agreement with the Canadian government to harmonise their provincial sales tax with the national sales tax (the Goods and Services Taxes (GST)). With this agreement, the Canadian government is collecting the Harmonised Sales Tax (HST) during the year and at the end of each fiscal year, the appropriate share is allocated to the provinces. Statistics Canada obtained the mandate to produce reliable provincial estimates to be used in the determination of the provincial share. However, the infrastructure and the survey programs could not support nor produce the new requirements.

To meet the new requirements, Statistics Canada has been carrying out a major redesign of its annual business programs, some of its household programs as well as a re-engineering of the System of National Accounts and the provincial input-output tables. One of the major improvements is the implementation of a new annual business survey that ultimately will become the vehicle for producing annual estimates for all industries at the provincial and industrial required details: the Unified Enterprise Survey (UES). The UES was created and developed based on consistent, quality and methodologically sound approaches.

The purpose of this paper is to describe the sampling features of the UES through the years. The sample design has gone through some major modifications since its beginning in 1997. Even though the objective of the survey always remained to produce reliable estimates at the provincial level, every year, the arrival of new industries affected the industrial level requirements. As well, on-going programs being integrated into the sample had historical requirements, which did not necessarily go hand in hand with PIPES' objective of controlling response burden. The relative importance of some industry compared to others as well as the ever-present provincial dimension made the sample allocation process a central issue for all managers. For all these reasons, the samples went through an evaluation and a redesign every year, which are presented here.

The paper is organised in the following way. Except for section 12, all sections discuss the sampling features of the second-phase sample of the UES for different reference years. The first sections present the industries and the basic feature of the sample, the survey frames as well as the concepts used throughout the UES. The sampling parameters are described in subsequent sections. These include sample stratification, allocation and selection. Network sampling is explained in section 8. Section 9 and 10 describe new methods used to derive target coefficient of variations (CVs) among the surveys and the provinces. Section 11 discusses a new approach to identify and treat insignificant cells in order to alleviate response burden. Section 12 presents briefly the first-phase sample where auxiliary information is gathered for the estimation strategy. Finally, future directions and acknowledgements are presented in the last sections.

## **2 SURVEY DESIGNS**

The main feature of the design is the use of a two-phase approach. Two samples are selected every year. The first-phase sample is used to gather auxiliary information or data, i.e. administrative records, which will be used in the estimation strategy to produce more efficient estimators. Section 12 presents some details of the auxiliary information selected in this first sample. The second-phase is the main sample and is used to obtain survey data through questionnaires. Financial and non-financial

information are produced to meet the estimation requirements with this sample. The first-phase is a complete overlap of the second-phase and the same sampling units are used in both phases.

In 1997, the first reference year, 7 pilot industries, not being surveyed for a long time, were surveyed through the UES. 16 more industries were integrated in 1998, including the Wholesale industry. For the following years, two other important surveys have been integrated: the Retail industry in 1999 and the Manufacturing industry in 2000. However, for UES 1998 and 2000, two industries were not surveyed in the second-phase, i.e. only the first-phase sample was selected. For these two industries only administrative records will be used to produce the financial and the non-financial estimates. Tables 1 and 2 show all survey industries in UES and their activity years under the UES platform.

**Table 1:** List of Surveys covered by UES - Survey and administrative data

<u>Survey name</u>	<u>NAICS covered</u>	<u>Activity years</u>
Accounting & Bookkeeping	5412	1998, 99, 00
Aquaculture	1125	1997,98, 99, 2000
Book Publishers	51113 (Activity-based)	2000
Couriers	492	1997, 98, 99, 2000
Database & Directory Publishers	51114	1998, 99, 2000
Employment Services	5613	1998, 2000
Food Services & Drinking Places	722	1997, 98, 99, 2000
Geomatics Services	54136, 54137	1998, 99, 2000
Lessors Services	5311, 531310	1997, 98, 99
Management	5416	1998, 99, 2000
Manufacture	31, 32, 33 (including 113,1153)	2000
Mine supplementary	2122,2123	2000
Newspaper Publishers	51111	1998, 99, 2000
Non-Store Retail	454 but not 454310	1999, 2000
Non-Store Retail (other)	454310	2000
Real estate agents	5312, 531320, 531390	1997, 98, 99, 2000
Repair & Maintenance (auto)	8111	1999
Repair & Maintenance (other)	8112, 8113, 8114	1999, 2000
Retail stores	44, 45 but not 454	1999, 2000
Specialized Designs	5414, 54132	1998, 99, 2000
Specialized Publishers	51119	1998, 99
Testing Laboratories	54138	1998, 99, 2000
Translation Services	54193	1998, 99, 2000
Wholesale	41	1998, 99, 2000

**Table 2:** List of Surveys covered by UES - Administrative data only

<u>Survey name</u>	<u>NAICS covered</u>	<u>Survey years<sup>1</sup></u>	<u>Tax years</u>
Construction	23	1997, 1999	1998, 2000
Taxi & Limousines	4853	1997, 1999	1998, 2000

<sup>1</sup> Indicates for which year(s) a questionnaire was sent or is to be sent under UES

### 3 THE FRAME

All sampling processes begins with the creation of a frame, which in UES, is a list of all units in the target population. The frame is produced by the Business Register (BR), Statistics Canada's database system used to manage, update and produce all necessary sampling information for business surveys. The BR was re-designed in the mid-eighties. It is a complex database based on information provided by administrative files. The BR creates and manages population, sample and collection files for every survey using it. The UES front-end files are discussed in Appendix 1. The UES frame covers the universe of Canadian businesses, employers as well as non-employers. It contains information about each of the businesses as well as links to other statistical or administrative entities. Some of the available information on the BR includes: provincial and industrial coding, size variables (revenue and number of employees), address, legal name, unique statistical identifier, Business Number (BN) for linking the business entity to the administrative files, birth and death dates, complexity structure indicators, coverage information, etc.

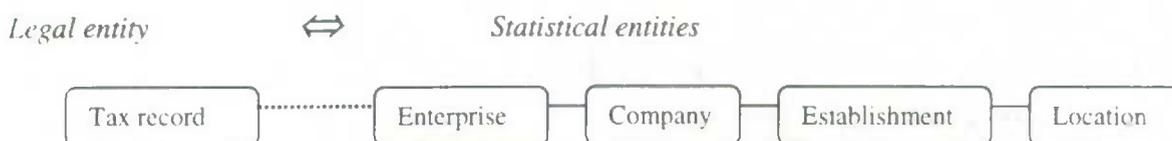
Any one of the four available levels of statistical entities, i.e. enterprise, company, establishment and location residing on the BR, can be used as sampling units. This is shown in Figure 1. The creation of these statistical entities is based on a set of standard rules and criteria. The first two levels are often used when the objective of the survey is to produce financial information, i.e. consolidated revenues and expenses, as well as the financial statement of the business, to name a few. The latter two are most often used when the objective of the survey is to produce operational or production type of information, i.e. type of activity, salary, number of units produced, etc. More details can be found in Cuthill (1990).

Figure 1: Statistical entities available on the Business Register

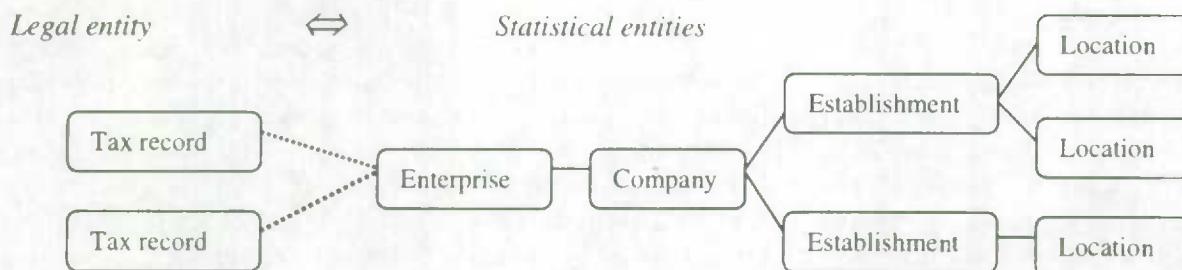


An important level in UES is the legal entity level. Canada Customs and Revenue Agency (CCRA) defines this level in collaboration with Statistics Canada and it is derived, based on the legal structure in which the business defines and operates itself. The legal entity is central in the context of UES for the data acquisition and estimation strategy because it is at this level that the auxiliary information, i.e. administrative records, is available. The two statistical levels used in UES are the enterprise and the establishment. The establishment level is defined as a physical production entity operating in one province and in one industry. Its industrial code is assigned using the North American Industrial Classification System (NAICS), at the 6-digit level. The enterprise is defined as the administrative entity managing the establishment(s). It is at this level that financial statements are produced. The enterprise does not have a provincial classification assigned to it. However, it is assigned the dominant industrial classification of its corresponding establishments. The dominance rules are based on the revenue variable. Figures 2 and 3 show two different types of enterprise as well as the relationship between the entities.

Figure 2: Simple Structured Enterprise



**Figure 3: Complex Structured Enterprise**



### 3.1 Evolution of the frame coverage for UES

Since its beginning, the UES' coverage has increased. There were different types of units loaded onto the frame each year, which made the target population increase in number. The types are presented in Table 3. From July 1997 until October 1999, almost 1.2 million records of non-employer units were added to the frame coverage. These represent a major change and have significant impact for the analyst and the users since every year the target population changes. In the future, there is no new load expected to be introduced onto the BR except for the normal birthing of units.

**Table 3: Evolution into the UES frame**

	Employers <sup>2</sup>	Non-Employers		
		With GST Sales above \$ 30 000.	With GST Sales below \$ 30 000.	No GST Account
<b>Incorporated</b>	APRIL 1987	October 1997 May 1998 October 1998 October 1999	May 1998 October 1998 October 1999	October 1998 October 1999
<b>Unincorporated</b>	APRIL 1987	October 1997 May 1998 October 1998 October 1999	Out of scope for now	Out of scope for now

## 4 UES CONCEPTS

### 4.1 Cell concept

The cell is an essential concept for UES. It is a way to delimit the target population into pieces. It is based on two important dimensions for UES, also used for stratification. One of these is the provincial dimension while the other corresponds to the industrial dimension.

#### UES 1997:

For UES 1997, the cell is a grouping of all establishments operating within the same province/territory and within the same 6-digit NAICS.

<sup>2</sup> Employers and non-employers are defined using payroll information available from CCRA.

Since UES 1998:

For all subsequent years, a cell is a grouping of all establishments operating within the same province/territory and within the same aggregated NAICS level. It can be seen as a 2-dimension table where the rows are the industrial aggregation and the columns are the 10 provinces and the three territories as shown in Table 4. The aggregation is either 4-digit NAICS or 5-digit NAICS.

**Table 4:** Illustration of a cell

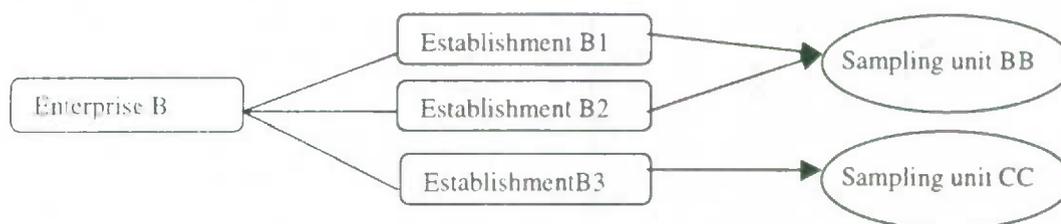
NAICS / Prov.	Prov. 1	Prov.2	Prov. j	Prov.13
NAICS 2311	Cell <sub>11</sub>	Cell <sub>12</sub>	Cell <sub>1j</sub>	Cell <sub>113</sub>
NAICS 23221	Cell <sub>21</sub>	Cell <sub>22</sub>	Cell <sub>2j</sub>	Cell <sub>213</sub>
...	Cell <sub>i1</sub>	Cell <sub>i2</sub>	Cell <sub>ij</sub>	Cell <sub>i13</sub>
NAICS 7224	Cell <sub>n1</sub>	Cell <sub>n2</sub>	Cell <sub>nj</sub>	Cell <sub>n13</sub>

#### 4.2 Sampling unit

The sampling unit is closely related to the concept of cell. Since cells provide a mutually exclusive coverage of the population, an enterprise may have establishments in different cells. Consequently, cells split an enterprise into clusters of establishments that are defined as sampling units. More precisely, a sampling unit is made of all establishments within the same cell that belong to the same enterprise. All sampling information available at the establishment level are aggregated to the sampling unit level, for example, the revenue value which will be used as the size stratification is summed up at this level. The value will be used in the third stratification level as described in section 5.2.

As shown in Figure 4, Enterprise B is linked to three establishments B1, B2 and B3. Since two of the establishments, B1 and B2, operate within the same industrial aggregation and province, they belong to a single sampling unit BB. Similarly B3 yields the sampling unit CC. Note that in UES 1997, the industrial aggregation would be done at the 6-digits NAICS and from UES 1998 and subsequent years, the industrial level would be the 4-digit NAICS or 5-digit NAICS depending on the industries.

**Figure 4:** Sampling Units for UES



#### 4.3 Simple and complex enterprises

An important aspect of an enterprise is its complexity. An enterprise is defined as complex if any of the following criteria is satisfied:

- It is multi-legal (that is, it has more than one tax record in the legal structure of the BR), or;
- It has establishments operating in different provinces, or;
- It has establishments operating in different 6-digit NAICS.

An enterprise that is not complex according to the previous definition is defined as simple. Note that these definitions are not standard throughout Statistics Canada, but only used for UES. This distinction is used to determine to what extent survey and tax data are used at the estimation stage. Since the estimation method used some calibration techniques, it requires both sources of data to be at the same level. Only simple units meet this requirement

## 5 STRATIFICATION

### 5.1 Thresholds (Royce-Maranda) and Stratification

To reduce response burden on small enterprises, UES used a delimitation to partition the population into two pieces, i.e. the portion for which units will be eligible to receive a questionnaire and the portion for which units will not be eligible to receive a questionnaire. The first portion will be estimated with survey and if applicable, administrative data. The second portion, known as the take-none portion will be estimated with administrative data only. This latter portion is comprised of small enterprises.

#### UES 1997:

In UES 1997, there were thresholds, applied after the size stratification level was defined and after the sample was selected. They were used to delimit two portions.

#### **For Construction and Food Industries:**

- Simple enterprises selected in the sample, which have establishment revenue of \$50K or less.
- Complex enterprises selected in the sample, which have enterprise revenue of \$1 million or less.

#### **For Aquaculture, Taxi-Limousine, Courier, Lessors and Real Estate Agents Industries:**

- Simple enterprises selected in the sample, which have establishment revenue of \$150K or less.
- Complex enterprises selected in the sample, which have enterprise revenue of \$1 million or less.

#### Since UES 1998:

More elaborate threshold values were calculated with the same purpose to reduce response burden on small enterprises. They have been implemented consistently for all industries. The main difference is that these thresholds are now used as lower stratification boundaries. Furthermore, the values of these thresholds are calculated in such a way to ensure that a maximum of 5 % of the total revenue of the cell is estimated with tax only, i.e. with the take-none portion. These are known as the Royce-Maranda (R-M) thresholds. Table 5 presents the actual values used for each year.

**Table 5: Actual values used for exclusion**

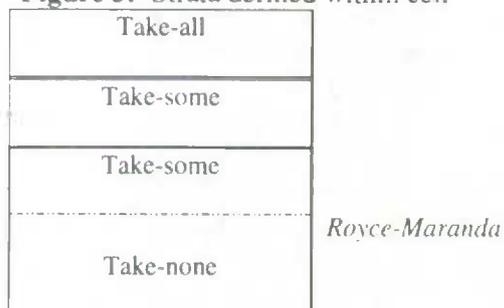
UES 1998 - 1999	UES 2000
30 000	30 000
45 844	45 000
70 057	70 000
107 056	105 000
163 597	160 000
250 000	250 000

In Appendix 2, there are illustrations of the frames, the two samples and the exclusion thresholds.

## 5.2 Stratification

There are three basic levels of stratification used in UES. The first two are used to meet the estimation requirements; the last one makes the design more efficient. First, the sampling units are grouped into cells, defined by provinces and industrial aggregations. These are the primary and secondary strata. The third stratification level uses a size measure, usually revenue available on the frame. Since UES 1998, the RM are then applied to delineate the portion above and below the thresholds for each cell. The portion above will then be divided into three relatively homogeneous groups based on the units' size. The resulting groups are one take-all stratum and two take-some strata above RM and one take-none below RM as presented in Figure 5. The method used for the stratification is the Lavallée-Hidiroglou (L-H) algorithm (1988). This iterative method consists of determining optimal stratum boundaries as to minimise the total sample size for a given or targeted CV. The minimisation process is done in function of an allocation method chosen by the statistician. The method used in the UES design is described in the section 6. The actual CV determination, i.e. how to determine the target CV of each cell, will be the topic of section 9.

Figure 5: Strata defined within cell



## 5.3 Must-take units

In addition to the take-all units defined by the L-H algorithm, there is another type of unit, which are selected with certainty, they are known as the must-take units. In contrast to take-all units, they are not identified by being above some optimal values as derived by the L-H algorithm, but rather above one single value applied to all industries the same way. In practice, this means that all sampling units belonging to a complex enterprise with an enterprise revenue larger than \$25 millions will be selected. Operationally, they are put aside before the L-H algorithm but are still included in the calculation of the target CV.

## 6 SAMPLE ALLOCATION

The sample size in each stratum is determined by the L-H algorithm, which requires for input, a target CV at the cell level, i.e., the first two stratification dimensions. The sample size allocation is completed with a proportional square root of the revenue total between the 3 size strata.

### 6.1 Sample inflation rates

After obtaining the sample size for each stratum, they are inflated to account for non-response and frame imperfections such as deaths and misclassifications. In UES 1997, the non-response inflation rate was 20% for all industries. In UES 1998, the rate was 40% for almost all, with the exception of

Real Estate Agents and Employment with 15%; Wholesale with 20%; Food with 25%; and Management with 30%. As for UES 1999, the rates varied more by industry. Table 6 presents the non-response rates by industries. The inflation rates for frame imperfections are industry-specific. Table 7 presents the inflation rates for the three reference years.

**Table 6:** Non-response rates for UES in percentage

Survey-NAICS level	UES 1999	UES 2000
Accounting & Bookkeeping	40	45
Aquaculture	40	45
Book Publishers	n/a	40
Construction	15	n/a
Courier	40	45
Database publishers	40	65
Food	25	55
Geomatics	40	40
Lessors	35	n/a
Management	40	55
Mine supplementary	n/a	40
Newspaper	40	30
Non-Stores Retail	40	40
Real Estate	25	55
Repair & Maintenance (auto)	40	n/a
Repair & Maintenance (other)	40	50
Retail stores	40	40
Specialized Design	40	50
Specialized Publishers	40	n/a
Taxi-Limousine	40	n/a
Testing Labs	40	70
Translation	40	45
Wholesale	20	40

**Table 7:** Frame imperfection rates for UES in percentage

Survey-NAICS level	UES 1997	UES 1998	UES 1999	UES 2000
Accounting & Bookkeeping	n/a	25	25	20
Aquaculture	10.2	40	40	15
Book Publishers	n/a	n/a	n/a	25
Construction	9.4	n/a	22	n/a
Courier	7.2	20	20	15
Database publishers	n/a	15	15	30
Employment	n/a	20	n/a	30
Food	15.9	10 - 25	10 - 20	10
Geomatics	n/a	10 - 20	25	25
Lessors	5.7	10 - 25	15	n/a
Management	n/a	25	40	35
Mine supplementary	n/a	n/a	n/a	15
Newspaper	n/a	15	15	15

<b>Non-Stores Retail</b>	n/a	n/a	15	15
<b>Real Estate Agents</b>	5.7	15 - 35	40	25
<b>Repair &amp; Maintenance (auto)</b>	n/a	n/a	25	n/a
<b>Repair &amp; Maintenance (other)</b>	n/a	n/a	25	35
<b>Retail stores</b>	n/a	n/a	15	15
<b>Specialized Design</b>	n/a	15	30	20
<b>Specialized Publishers</b>	n/a	15	15	n/a
<b>Taxi-Limousine</b>	5.2	n/a	15	n/a
<b>Testing Labs</b>	n/a	20	50	50
<b>Translation</b>	n/a	15	15	25
<b>Wholesale except 4191</b>	n/a	35	50	15
<b>Wholesale - 4191</b>	n/a	50	35	15

## 7 SAMPLE SELECTION

Prior to sampling selection, a random number (RN) on the interval  $[0,1)$  is assigned to all sampling units. Then, units are ordered by their RN and the selection is made starting with the smallest RN as to select as many sampling units as needed to meet the pre-determined sample sizes in each stratum. This is similar to a Simple Random Sample (SRS) design within each stratum. Note that the same RN are used for the first-phase, thus the complete overlap.

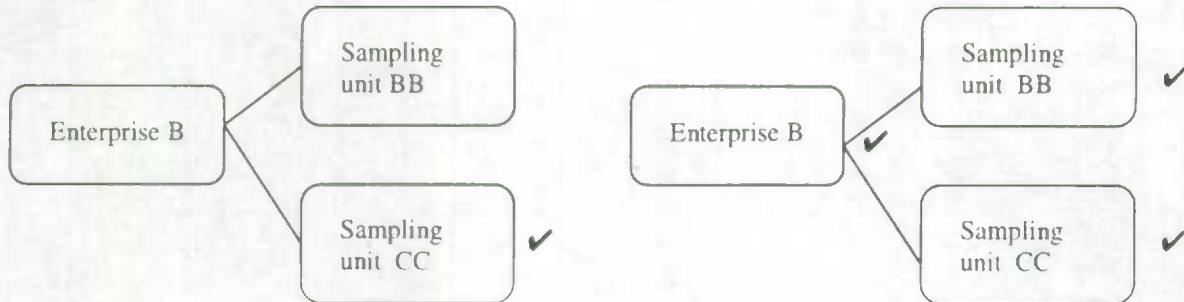
## 8 NETWORK SAMPLING

One of the objectives of the UES is to ensure coherence between the enterprise and its associated establishments' estimates. The sampling challenge remained in the selection of all establishments associated with a selected enterprise without over-sampling. Note that only complex units are problematic, since for the simple ones, enterprise and establishment are exactly the same. To meet all objectives and based on the population's characteristics, an approach known as network sampling, was chosen. This technique is achieved as follows: once an initial sample of sampling units has been obtained using Stratified Simple Random Sampling (STSRs), adjustments are made using network sampling. Network sampling is a feature by which originally non-selected sampling units get to join the initial sample if a linked (through a network) unit was selected in the original sample. In UES, two units are linked if they belong to the same enterprise and operate in the same field of activity. See Simard and Hidioglou (1999) for more details.

Network sampling wasn't introduced in UES as a means to increase sampling efficiency, but merely to complete the "sampling picture" of a selected complex enterprise. This is different in spirit to a closely related sampling strategy known as adaptive sampling. Read Thompson and Seber (1996) for more details. Adaptive sampling is essentially about adding units while in the field that will result in a more effective sample than what was originally drawn. Network sampling was a feature for UES 1997, UES 1999 and for UES 2000. In UES 1998, a census approach for all multi-provincial units was used instead.

Using the example given in Figure 4, suppose that, initially, only sampling unit CC had been selected into the sample (left-hand side of Figure 6). Network sampling will result in having units BB; CC and the associated enterprise B included in the sample (right-hand side of Figure 6).

**Figure 6:** Before and after network sampling



### 8.1 Inclusion probabilities and design weight

Network sampling has significant impacts on inclusion probabilities of units and conversely on the design weights. Note that by definition network sampling can only be applied on complex enterprises. For units of complex enterprises, final inclusion probabilities are often different than initial design inclusion probabilities due to network sampling. Under network sampling, the probability that a given unit ultimately joins the sample is not only a function of its initial design inclusion probability but also of the initial design inclusion probabilities of all units of the network it's in. Note that initial design probabilities are defined based on the initial sampling design, that is, before the network is applied. In the case of UES, since traditional STSRS is used, the design probabilities are equal to  $n_h / N_h$  for each stratum  $h$ .

Let  $\pi_i$  and  $\pi_i^*$  stand for the initial and the final inclusion probability, i.e. after network sampling of unit  $i$ , respectively. If unit  $i$  is related by network sampling to units  $i_1, \dots, i_n$  and  $S^*$  is the final sample then

$$\begin{aligned} \pi_i^* &= \text{Prob}(i \in S^*) = 1 - \text{Prob}(i \notin S^*) = 1 - \text{Prob}(i \notin S, i_1 \notin S, \dots, i_n \notin S) \\ &= 1 - \text{Prob}(i \notin S) \times \text{Prob}(i_1 \notin S) \times \dots \times \text{Prob}(i_n \notin S) = 1 - \prod_{k \in \text{Net}(i)} (1 - \pi_k) \end{aligned} \quad (1)$$

where  $\text{Net}(i)$  is the list of units related to unit  $i$  by network sampling (including  $i$  itself). The resulting network weight  $w_i^* = \pi_i^{*-1}$  can then be used in a Horvitz-Thompson type of estimator:

$$\hat{Y}_{\text{net}} = \sum_{S^*} w_i^* y_i \quad (2)$$

One other method explores the possibility of using a weight-share approach to calculate the design weight. The weight-share method facilitates the calculations of a variance estimator. Read all details about the two estimators and the variance estimation techniques in Girard and Simard (2000).

This concludes the steps required for the selection of units in the sample. Appendices 3 to 6 present the final sample sizes by UES surveys from 1997 to 2000, respectively. The next sections present exactly how the sample is shared between the surveys through the CV allocation.

## 9 CELL CV DETERMINATION

With the arrival of on-going surveys into the UES platform, the sampling design became a crucial step for the whole sampling process. The HST allocation formula requires estimates of equally reliable and detailed precision for all provinces. For the surveys being integrated, the provincial dimension is not as important as they often require very detailed estimates for their different industrial aggregations. The sampling allocation process had to meet both requirements, the provincial and the industrial ones, and response burden had to be reduced as well. This task became a challenge.

### 9.1 UES 1997 and 1998

For UES 1997 and 1998, a common CV value was targeted for each cell, no matter what province or portion of industry it covered. In 1997, the common target CV was 15 % and in 1998, it was 7.5%. Read Simard and Laniel (1998) for more information about these sampling designs. That strategy, however did not take into account the relative importance of the cell.

### 9.2 UES 1999: Raking ratio

The fact that the designs did not take into account the relative importance of each cell lead to a new strategy for UES 1999: a two-step raking ratio application. The raking ratio is an iterative process, which distributes optimal values into cells of any table based on fixed marginal values. The two applications are described in the next two sub-sections.

#### 9.2.1 First Application

The first application is to distribute the overall CV,  $CV_t$ , i.e. national CV all industries over sector levels as shown in Table 8. It is not efficient to constrain the CV to be the same for all sectors, since sectors are not of equal importance. A power allocation was then used in order to minimise large CV differences between small and large sectors as described in Bankier (1988). For a given sector  $s$ ,  $CV_s$  are calculated to be inversely proportional to the revenue total ( $GBI_s$ ) as shown in (3).

$$CV_s = \frac{CV_t GBI_t}{GBI_s^{s \exp} \sqrt{\sum_s GBI_s^{2-2s \exp}}} \quad (3)$$

This comes from the fact that variance is  $V_s = CV_s^2 GBI_s^2$  and that one wants  $\sum_s V_s = V_t$

For UES 1999, the overall target CV was set at 0.75% and the power ( $s \exp.$ ) used for the industrial allocation was 0.25.

**Table 8:** Sector definition

Sector	Surveys	Division
11	Aquaculture	Agriculture
23	Construction	Manufacturing, Construction and Energy
41	Wholesale	Distributive trade
44-45	Retail	Distributive trade
48-49	Taxi-Limousine, Courier	Transportation
51	Database, Specialized and Newspaper Publishers	Services Industry

53	Lessors, Real Estate Agents	Services Industry
54	Accounting, Geomatics, Management, Specialized Design, Testing Labs, Translation	Services Industry
56	Employment	Services Industry
72	Food	Services Industry
81	Repair and Maintenance	Services Industry

### 9.2.2 Second Application

The second application will be completed by sector and the runs are done independently within sector. For example, say there are 10 sectors in one reference year. The first application is done only one time with all sectors together and the results are ten sectorial CVs. Then 10 second applications will be run.

The second application is done in two steps. In this application, the sector CVs calculated in the first raking is allocated to each respective sector' cells, as defined in 4.1. Figures 7A and 7B show the relationship between the two applications. It consists of taking each sector independently and applying the raking for the industrial aggregations belonging to the given sector.

Figure 7 A: First step: allocating the global CV into sectorial CVs

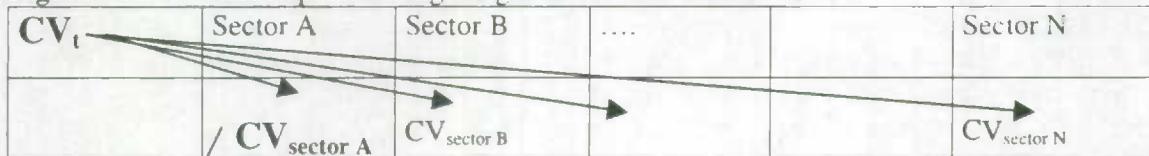


Figure 7 B: Second step: Sectorial raking

$CV_{sector A}$	Prov. 1 $CV_{prov 1}$	Prov. 2 $CV_{prov 2}$				Prov. 13 $CV_{prov 13}$
NAICS4 A1 $CV_{stratum A1}$						
NAICS4 A2 $CV_{stratum A1}$				cell <sub>s,p</sub> $CV_{cell}$		
.						
NAICS4 AJ $CV_{stratum AJ}$						

The first step is to distribute the overall sectorial CV,  $CV_{sector}$ , over the province and the industrial stratification level (stratification = NAICS4 or NAICS5) in the margins. It is not efficient to constrain the CV to be the same for all sectors since sectors are not of equal importance. The same power allocation described in (3) was used with the same power, i.e. 0.25.

As for the provincial CV,  $CV_{prov}$ , a constraint was applied so that all provinces have the same provincial CV, which is also calculated based on the overall target CV. This was done to satisfy the provincial driven constraints for HST. For a given province p,  $CV_p$  was determined by equation (4):

$$CV_p = \frac{CV_s GBI_s}{\sqrt{\sum_p GBI_p^2}} \quad (4)$$

Once the marginal CV's are calculated, the second step: the raking algorithm proceeds to determine the cell level CV's. Initial cell CV's,  $CV_{s*p}^{(0)}$ , are computed using the following equation:

$$CV_{s*p}^{(0)} = 1/2 * \frac{CV_s GBI_s}{\sqrt{\sum_p GBI_{s*p}^2}} + 1/2 * \frac{CV_p GBI_p}{GBI_{s*p}^{s \text{ exp}} \sqrt{\sum_s GBI_{s*p}^{2-2s \text{ exp}}}} \quad (5)$$

These CV's are then used in the raking ratio algorithm to determine a CV for each cell  $s*p$  with the following equation:

$$CV_{s*p}^{(i)} = \frac{CV_{s*p}^{(i-1)} CV_s CV_p GBI_s GBI_p}{\sqrt{\sum_s (CV_{s*p}^{(i-1)})^2 GBI_{s*p}^2} * \sqrt{\sum_p (CV_{s*p}^{(i-1)})^2 GBI_{s*p}^2}} \quad (6)$$

$i$  is the current iteration of the raking algorithm.

The allocation process used give little importance to some cells, and thus their target CV ended up being too high for some surveys. Since a representative sample in each cell was needed, it was decided to set some CV ceiling. All target CVs above 15% were reduced to 15%. Also, a list of important cells was produced by each survey managers along with the desired ceilings ranging from 5% to 15%. These adjustments changed about 40% of all targeted CVs, which is quite significant in terms of changes.

## 10 UES 2000: NEW PROPOSAL

The major change occurring for 2000 is the improvement of the sample allocation, i.e. the CV allocation. Due to the over-use of the CV ceilings at the end of the process, some other options were evaluated. Three are presented in the next sub-sections.

### 10.1 Multiplication factor

One of the options is to use a multiplication factor into the CV formula, which would make some CVs systematically bigger or smaller than others. This is described in Daoust and Lim (1986). The Survey of Employment, Payrolls and Hours (SEPH) underwent a major redesign and this was the method chosen. However for SEPH, since the same level of precision was not required across all the provinces, the provincial CV in small provinces were aimed to be 3 times greater than those for big provinces as shown in (7):

$$CV_p^* = \frac{CV_s GBI_s}{\sqrt{\sum_p (GBI_p^*)^2}} \quad (7)$$

$$\text{GBI}_p^* = \begin{cases} 3 * \text{GBI}_p & \text{for the territories, Atlantic and Prairie provinces} \\ \text{GBI}_p & \text{otherwise} \end{cases}$$
$$\text{CV}_p = \begin{cases} 3 * \text{CV}_p^* & \text{for the territories, Atlantic and Prairie provinces} \\ \text{CV}_p^* & \text{otherwise} \end{cases}$$

Implementing this option for UES did not have satisfactory results. First, UES contains many surveys and what can be perceived as the large province will vary from survey to survey. For example, Aquaculture has an important business activity in Prince Edward Island whereas Wholesale, Ontario is perceived as more important. There is a need to identify large provinces in every survey and that proved to be tedious. Also, even though larger provinces will have smaller CVs as desired, the other provinces will have higher cell CVs and most of the time these went over the maximum target CVs. This is not an improvement since most CVs will still end up requiring a change.

## **10.2 Sample size**

The second option is more complicated and has many operational and system implications. Basically, it is to set a sample size, as is done for many surveys, instead of setting a target CV. Looking back to the sample sizes obtained in previous UES occasions, one could determine the expected sample size for a given survey or sector. The raking would then allocate that sample size to the different cells in that survey.

It is important at this point that the allocation process remains essentially the same for UES to gain stability in time. It also has to be pointed out that the raking ratio as it was implemented in 1999 worked well and only minor changes seem necessary for future years. This option implies substantial modification to the systems and programs already in place. For these reasons, this option was not evaluated viable for now.

## **10.3 Power allocation**

This option is to use the same approach as the industry power allocation but for the province dimension. The main concern is that the allocation was not adequate for provinces of greater importance. A loss of efficiency resulted from the fact that the importance of each province was not considered in the allocation process. This stems from the assumption stated in the beginning of PIPES, which mentioned equal provincial precision. However, it was agreed that the equality constraint could never really get applied completely. Managers agreed that the constraint should be released slightly. It is now an almost equal CV that is driving the sample allocation. This being decided, the relative importance of province could now be considered.

### **10.3.1 Introducing a Provincial Exponent**

A provincial exponent will cause the CVs to be inversely related to the provincial GBI. This inverse relationship is desirable since for most sectors there is a considerable variation in the overall provincial GBIs. Thus, it is illogical to give the same importance to a cell in a province with 0.05% of the overall sector GBI, as to a cell in a province with 45% of the overall sector GBI.

To introduce a provincial exponent the following changes are made.

$$CV_p = \frac{CV_s GBI_s}{\sqrt{\sum_p GBI_p^2}} \quad (4) \quad \text{is replaced with} \quad CV_p = \frac{CV_s GBI_s}{GBI_p^{p \exp} \sqrt{\sum_p GBI_p^{2-2p \exp}}} \quad (8)$$

$$1/2 * \frac{CV_n GBI_n}{\sqrt{\sum_p GBI_{n*p}^2}} \quad (5) \quad \text{is replaced with} \quad 1/2 * \frac{CV_n GBI_n}{GBI_{n*p}^{p \exp} \sqrt{\sum_p GBI_{n*p}^{2-2p \exp}}} \quad (9)$$

Different combination of exponents were studied and in the end, the final provincial exponent used was 0.10 in order to keep overall provincial CVs from varying too much; and thereby, respecting the HST restriction.

#### 10.4 UES 2000 CV allocation

In UES 2000, the two-step application described in section 9.2 was used. The overall target CV was set at 0.65 %, the industrial exponent in the first application was 0.25. As for the second application, the industrial exponent was 0.20 and the provincial one 0.10.

### 11 INSIGNIFICANT CELL IDENTIFICATION

In the past, all cells with less than 15 sampling units were automatically censuses. Some concerns were raised about this approach. This number was fixed because of the L-H algorithm was set to create 3 strata, which becomes impossible from a population with less than 15 sampling units. However this increased the response burden for a very minimal gain of efficiency in the estimates. To further reduce response burden, a new feature was introduced for UES 2000. In addition to the R-M thresholds, there was some criteria used to defined and identified insignificant cells. Once these were identified, those cells did not automatically have a census as was the case for previous years..

This leads to a two-fold definition of an insignificant cell. First, small in term of units belonging to a cell, i.e. below a certain k number of units, the cells are identified insignificant. Second, small in term of economic importance of the cell relative to the global NAICS activity (NA) and/or relative to the global province activity (PA).

For a given sector, two ratios, PA and NA are defined as follows:

PA= total GBI in the cell/ total GBI in a given province.

NA= total GBI in the cell/ total GBI in a given stratification NAICS.

Using those three criteria: four situations can arise:

- A - It has less than k sampling units and both PA and NA are small. The cell is insignificant for both criteria. No questionnaire should be sent.
- B - It has at least k sampling units but both PA and NA are small. The cell has sufficient sampling units, but it has poor economic activity. Proceed with sampling, but do not use any improvement factors such as over-sampling rates.

- C - It has less than k sampling units but at least one or both of PA and NA are evaluated sufficiently big. The cell does not have sufficient sampling units, but an important activity. Proceed with a census.
- D - It has at least k sampling units and at least one or both of PA and NA are evaluated sufficiently big. The cell is significant for all criteria. Proceed with regular sampling.

### 11.1 Definition of the criteria

The challenges remained in the definition of the three parameters, namely i) k, ii) the cut-off value for PA (PAC) and iii) the cut-of value for NA (NAC) and within what level would the comparison be made. Since the raking is firstly done at the sector level, (see Table 9), it was decided to use that same level. For a given sector, a cell is “compared” to all other cells within the same sector. It is important to note that some sectors are very important while others are not. For example, Aquaculture only has one stratification NAICS belonging to its sector whereas Wholesale has about 25. Though convenient, the same parameters PAC and NAC can not be used for all sectors. For example, let’s say 15 % is set as the cut-off value. And let’s say Aquaculture and Wholesale have equal proportion of revenue for each cell. Aquaculture with only one dimension will not failed the cut-off value because the cell proportion is 100% of the total automatically. On the other hand, Wholesale which has 25 dimensions (4% for each cell) will mathematically have all failures even though the total absolute revenue can be larger for every cell than Aquaculture.

This is not a problem for parameter k. The parameter k is set to 15 as in previous years. To determine PAC and NAC, an approach using percentile was proposed. Even though the cut-off value would be different, the percentile used would be the same. Using the same percentile for all sectors gives certain coherence in the method and the resulting value is “adapted” to each sector. Different values of percentiles were studied in order to find the best compromise.

### 11.2 Preliminary results

For UES 2000 population test file, there are about 400 000 sampling units distributed between about 1200 cells, all eligible to go through the sampling process. Table 9 shows some results of testing different values of percentiles. The first number is a count of cells and the one within parenthesis is the corresponding counts of sampling units.

**Table 9: Results of testing different percentile**

Scenario	Case A	Case B	Case C	Case D
NAC and PAC= Percentile 40%	154 (798)	44 (2356)	191 (1295)	658 (375 320)
NAC and PAC= Percentile 35%	132 (646)	26 (1147)	213 (1447)	676 (376 529)
NAC and PAC= Percentile 30%	114 (521)	10 (345)	231 (1572)	692 (377 331)
NAC and PAC= Percentile 25%	95 (371)	3 (78)	250 (1722)	699 (377598)
NAC and PAC= Percentile 20%	76 (220)	0 (0)	269 (1873)	702 (377676)

Table 10 presents the distribution for the 114 cells defined as insignificant, i.e. case A, using the chosen percentile; 30%. Table 11 presents the distribution by provinces.

**Table 10:** Insignificant cells by surveys

Survey	Number in case A	Total cells in survey	Ratio (in %)
Accounting & Bookkeeping	0	13	0
Aquaculture	0	12	0
Courier	4	26	15.4
Employment	0	13	0
Food	3	52	5.8
Geomatics	4	24	16.7
Lessors	3	62	4.8
Management	0	13	0
Mine supplementary	3	24	12.5
Newspaper	0	12	0
Real Estate Agents	6	38	15.8
Retail stores	29	301	9.6
Repair & Maintenance (auto)	3	35	8.6
Repair & Maintenance (other)	4	38	10.5
Specialized Design	11	54	20.4
Testing Labs	1	13	7.7
Translation	3	13	23.1
Wholesale	40	304	13.2
<b>Total</b>	<b>114</b>	<b>1047</b>	<b>10.9</b>

**Table 11:** Insignificant cells by provinces

Province	Number in case A	Total cells in province	Ratio (in %)
Newfoundland	18	86	20.9
Prince-Edward-Island	20	81	24.7
Nova Scotia	5	86	5.8
New Brunswick	6	85	7.1
Quebec	0	87	0
Ontario	0	86	0
Manitoba	1	86	1.2
Saskatchewan	3	86	3.5
Alberta	1	86	1.2
British-Columbia	0	86	0
Yukon	21	75	28
North-western Territories	27	74	36.5
Nunavut	12	43	27.9
<b>Total Canada</b>	<b>114</b>	<b>1047</b>	<b>10.9</b>

The actual implementation during production of this procedure involves an additional step. It was agreed to allow each survey manager to revise the cell identified as insignificant for a final approval, before not sending any questionnaires.

## **12 FIRST-PHASE SAMPLE: AUXILIARY INFORMATION**

The first-phase sample has exactly the same sampling characteristics as described in the previous sections except they yield larger sample size with the setting of smaller global CVs. There are two minor differences i) the type of information obtained for the selected units, i.e. administrative information via tax records instead of questionnaires and ii) an additional stratification level, namely the type of tax record. There are two types of tax records provided by CCRA. Records from the incorporated business known as T2 and Records from the non-incorporated business, known as T1. Since 1998, with the introduction of the General Index for Financial Information (GIFI), the T2 sample is actually a census. The tax records are used for the simple units only as auxiliary information in the chosen calibration technique.

## **13 FUTURE DIRECTION**

Being in full development mode since its implementation in 1997, UES entered a stabilisation period in 2000. After Wholesale in 1998, Retail in 1999 and Manufacture in 2000, no more major annual surveys will be integrated in the near future. The next step for the UES surveys is the possible introduction of a rotation pattern for 2001 as another way to control response burden.

## **14 ACKNOWLEDGEMENTS**

The authors would like to thank Normand Laniel and Claude Poirier for their guidance in the project and Guylaine Dubreuil for her work in the early designs. They also would like to thank Joe Kresovic and Stuart Pursey for their useful comments.

## **15 REFERENCES**

- Bankier, M.D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician* **42**, 174-177.
- Cuthill, I. (1990). The Statistics Canada's Business Register. Internal document.
- Daoust, Pierre, Lim, Ann (1997). Allocation of the Sample for the Stage III Redesign of the Survey of Employment, Payrolls and Hours (SEPH). Internal Working Document, Statistics Canada.
- Girard, C and Simard, M.(2000) Network Sampling: An application in a Major Survey. ICES-II,
- Lavallée, P. and Hidiroglou, M.A. (1988). On the stratification of skewed population. *Survey Methodology*, no.14, pp 33-45
- Simard, M. and Hidiroglou, M. (1999). Estimation For Annual Business Surveys Based On Two-Phase Network Sampling.. *SSC Proceedings of the Survey methods Section*, pp 11-19.

Simard, M. and Laniel, N. (1998). Échantillonnage et Estimation pour l'enquête unifiée sur les entreprises. *SSC Proceedings of the Survey methods Section*, pp 77-82.

Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. John Wiley and Sons.

## **Appendix 1: Front-End files**

The front-end is comprised of three basic files from the BR: the Survey Universe File(SUF), the Sample Control File(SCF) and the Survey Interface File (SIF). The SUF is the starting point of any business survey hooked to the BR. It delineates, for a given effective date, the target population of the survey. For UES, the SUF covers all BR units, which were ever alive during the reference year. The ever-alive concept was changed from <alive at least one day> (1997 and 1998) to <alive more than one day> (1999 and 2000).

### **Survey Universe File (SUF):**

The SUF is the first frozen picture of the target population. It is the central image for CAPEX, IOFD, ASM and all in-scope UES industries. In the series of programs for the sampling process, the first ones are used to determine all population attributes from the BR-SUF. From those programs based on the SUF, we obtain: 1) the links between statistical entities, i.e. enterprise and establishments, (for allocation, combined reports and coherence analysis); 2) the links between statistical entities and administrative records, i.e. the T2, T1 (for all the uses of tax data records: E and I, allocation, mapping and the two-phase estimation approach); 3) the determination of complex-simple and multi-legal enterprises, according to UES definition, all used in subsequent functions.

At the beginning of PIPES it was decided to centralise all survey populations into one SUF, as this is one of PIPES basic objectives of integrating annual business surveys: a common frame. All parties involved have to decide the actual date of creation of the SUF. It is from that first picture in time 1) that we delimit which units are must-take, (\$25 million); 2) that we cluster establishments into sampling units; 3) that we assign probabilities of selection and stratification attributes; 4) that we obtain the population count in each stratum, 5) that we apply network sampling and 6) that we select units receiving questionnaires to ensure a representative sample, just to name a few.

### **Post-Survey Universe Files (post-SUF)**

Knowing that the SUF suffers from coverage errors and classifications, methodology proposed a post-stratification technique in the estimation approach, which basically means that later on, a second frozen picture in time of our target population is produced. This is the post-SUF, which includes subsequent updates, all obtain from sources independent to the survey (birth additions, death removals, simple-complex switch and NAICS corrections) still representing the same reference year. This was introduced to allow more time to BRD and survey managers for frame correction. This second picture can be seen as the last picture for which establishments and enterprise statistics will be produced. This is performed through final weighting of the selected units at the estimation stage.

### **Sample Control File (SCF)**

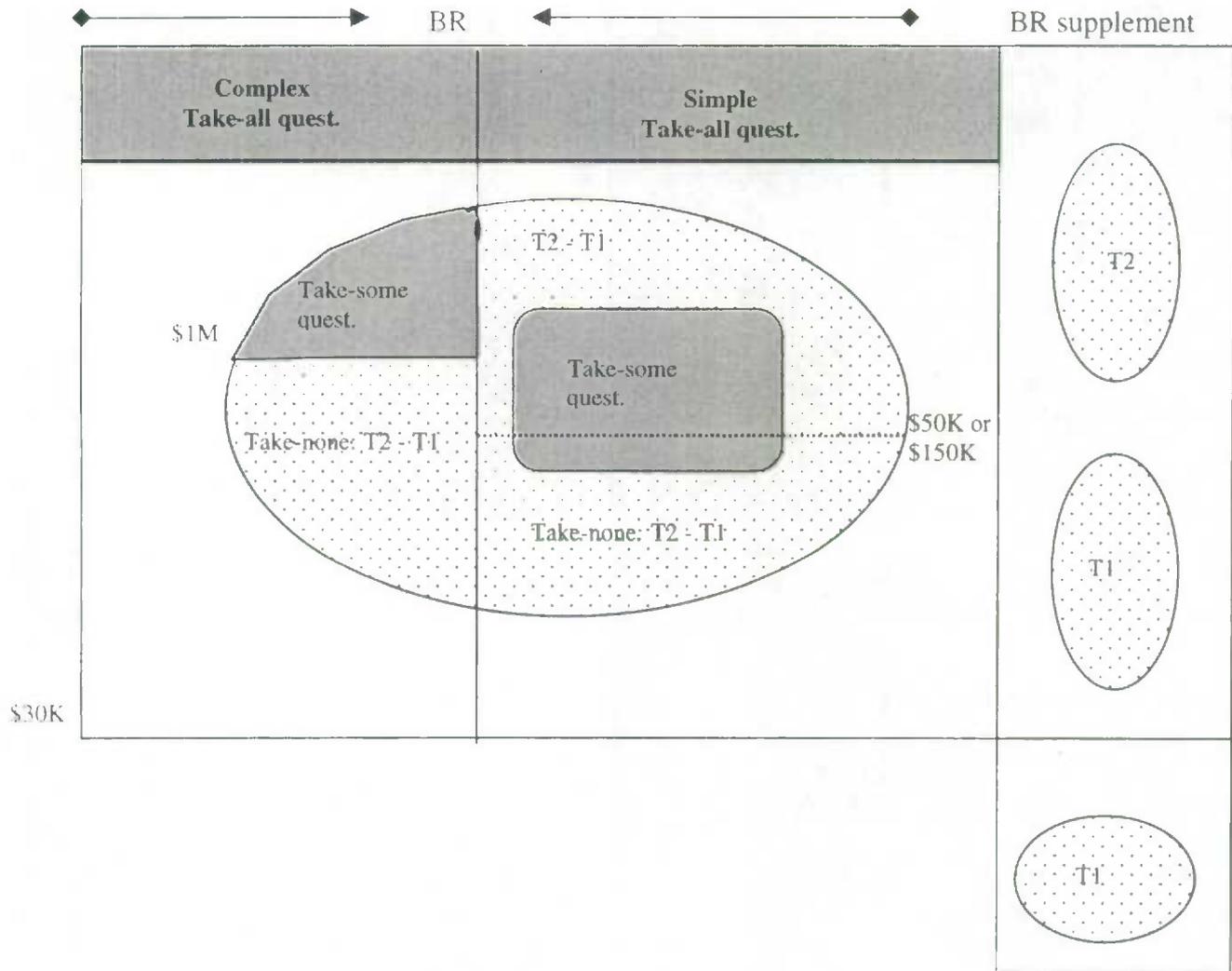
The SCF is the exact same picture of the population as the SUF, with the additions of all flags created in the sampling programs as described above plus the indication whether the units were selected to receive a questionnaire or not. All sampling information is recorded on the SCF. If there were rotation patterns, it would be recorded on the SCF. There are two SCFs created. The SCF-establishment contains CAPEX, ASM, UES and head office sampling information. The SCF-enterprise contains IOFD sampling information. Both SCFs representing the enterprise and establishment population for the same effective date for coherence analysis purposes. The SUF and SCF contain dead units, as these are part of the UES target population.

### **Survey Interface File (SIF)**

The SIF is conceptually different, as it contains the most recent information of each selected business entity. It retrieves from the CFDB the most recent collection information for SCF selected units. Each project manager is responsible for its survey and has to go through customisation and the creation of collection entities (CE). Some units are combined into one CE and combined reports are created. All this information is put on the SIF. It also contains collection information, coverage statement, address, legal names, which are not on the SUF, nor on the SCF. It is important for respondent relations to send the most up-to-date information on the respondent.

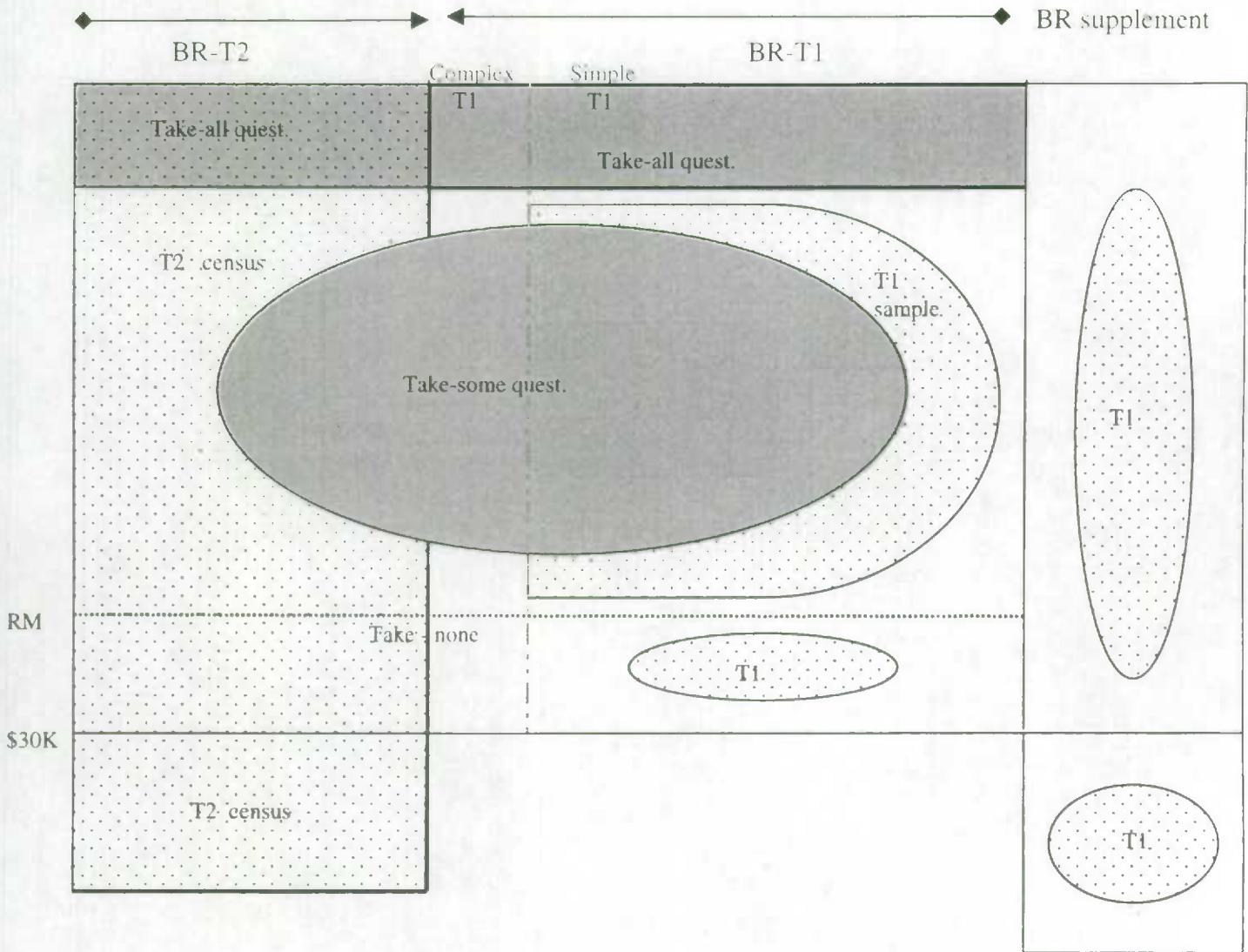
Appendix 2: Target Population, Samples and Exclusion thresholds.

UES 1997:



**UES 1998-1999 -2000**

UES 1998 -UES 1999



### Appendix 3: Sample size for UES 1997

#### Counts of establishments

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
AQUACUL	n/a	n/a	n/a	n/a	238	121
CONSTRU	n/a	n/a	n/a	n/a	110078	8167
COURIER	n/a	n/a	n/a	n/a	998	326
FOODSER	n/a	n/a	n/a	n/a	54235	4242
LESSORS	n/a	n/a	n/a	n/a	15676	1681
REALEST	n/a	n/a	n/a	n/a	4223	581
TAXILIM	n/a	n/a	n/a	n/a	828	321
<b>TOTAL</b>	n/a	n/a	n/a	n/a	<b>186276</b>	<b>15439</b>

Please note the followings for the next appendices.

*These will either be counts of establishments or counts of sampling units.*

*These only included units above the exclusion thresholds.*

*These only included the second-phase units.*

*This is the notation used.*

*Pop (C) : Number of complex in the population*

*Sample (C) : Number of complex in the sample*

*Pop (S) : Number of simple in the population*

*Sample (S) : Number of simple in the sample*

*Pop (T) : Total number in the population*

*Sample (T) : Total number in the sample*

Appendix 4: Sample size for UES 1998

Counts of establishments

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
ACCBOOK	337	294	12720	721	13057	1015
AQUACUL	36	34	315	175	351	209
COURIER	n/a	n/a	n/a	n/a	2445	599
DATAPUB	n/a	n/a	n/a	n/a	102	81
EMPLOYM	233	176	1649	344	1882	520
FOODSER	3570	3052	47854	3143	51424	6195
GEOMATI	n/a	n/a	n/a	n/a	1883	405
LESSORS	814	494	38561	2680	39375	3174
MANAGEM	436	185	26849	1357	27285	1542
NEWPAPE	249	226	460	152	709	378
REALEST	114	88	19783	1822	19897	1910
SPDESIG	n/a	n/a	n/a	n/a	5504	1090
SPECPUB	n/a	n/a	n/a	n/a	146	91
TESTLAB	132	76	2331	293	2463	369
TRANSLA	n/a	n/a	n/a	n/a	1147	193
WHOLESA	11464	9849	41571	5678	53035	15527
TOTAL	17385	14474	192093	16365	220705	33298

**Appendix 5: Sample size for UES 1999**

**Counts of establishments**

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
ACCBOOK	324	297	13820	664	14144	961
AQUACUL	41	39	311	173	352	212
BOOKPUB	41	38	380	123	421	161
CONSTRU	1319	998	118572	7056	119891	8054
COURIER	218	218	2715	385	2933	603
DATAPUB	20	20	100	100	120	120
FOODSER	3850	3323	50430	2810	54280	6133
GEOMATI	39	38	2161	437	2200	475
LESSORS	943	627	41573	2596	42516	3223
MANAGEM	425	268	34820	1934	35245	2202
NEWPAPE	334	332	457	148	791	480
NRSTORE	234	234	3308	3306	3542	3540
PERIODI	178	146	659	158	837	304
REALEST	77	49	24481	1700	24558	1749
RETSTORE	17858	17749	94591	11640	112449	29389
RMAUTOM	472	377	23210	1599	23682	1976
RMOTHER	458	322	16161	1153	16619	1475
SPDESIG	43	42	6341	1094	6384	1136
SPEC PUB	0	0	148	148	148	148
TAXILIM	10	9	5972	419	5982	428
TESTLAB	134	111	3253	420	3387	531
TRANSLA	3	3	1135	142	1138	145
WHOLESA	11936	11366	44545	5767	56481	17133
<b>TOTAL</b>	<b>38957</b>	<b>36606</b>	<b>489143</b>	<b>43972</b>	<b>528100</b>	<b>80578</b>

**Counts of sampling units**

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
ACCBOOK	125	100	13800	661	13925	761
AQUACUL	17	15	302	167	319	182
BOOKPUB	23	20	379	122	402	142
CONSTRU	909	626	118440	7041	119349	7667
COURIER	141	141	2708	378	2849	519
DATAPUB	17	17	99	99	116	116
FOODSER	680	310	49418	2422	50098	2732
GEOMATI	34	33	2139	418	2173	451
LESSORS	599	316	41172	2270	41771	2586
MANAGEM	331	194	34799	1924	35130	2118
NEWPAPE	67	65	421	118	488	183
NRSTORE	144	144	3296	3294	3440	3438
PERIODI	88	61	655	155	743	216
REALEST	54	29	24428	1685	24482	1714
RETSTOR	3708	3607	91948	9826	95656	13433
RMAUTOM	164	87	23146	1578	23310	1665
RMOTHER	269	174	16138	1137	16407	1311
SPDESIG	41	40	6338	1092	6379	1132
SPEC PUB	0	0	148	148	148	148
TAXILIM	8	7	5972	419	5980	426
TESTLAB	95	72	3249	418	3344	490
TRANSLA	3	3	1135	142	1138	145
WHOLESA	7948	7477	43548	5388	51496	12865
<b>TOTAL</b>	<b>15465</b>	<b>13538</b>	<b>483678</b>	<b>40902</b>	<b>499143</b>	<b>54440</b>

**Appendix 6: Sample size for UES 2000**

**Counts of establishments**

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
ACCBOOK	400	372	14497	624	14897	996
AQUACUL	19	19	348	173	367	192
BOOKPUB	37	36	1422	1422	1459	1458
COURIER	241	241	2739	323	2980	564
DATAPUB	14	14	76	76	90	90
EMPLOYM	260	242	1788	532	2048	774
FOODSER	3884	3426	51639	3650	55523	7076
GEOMATI	33	32	2311	553	2344	585
LESSORS	1058	739	44797	3172	45855	3911
MANAGEM	402	267	38658	2222	39060	2489
MINESUP	480	431	567	233	1047	664
NEWPAPE	285	282	449	131	734	413
NRSTORE	543	507	3956	2921	4499	3428
REALEST	84	65	27685	2508	27769	2573
RETSTOR	18000	16950	102267	11669	120267	28619
RMAUTOM	520	429	22747	2233	23267	2662
RMOTHER	457	378	15962	2219	16419	2597
SPDESIG	38	37	6786	1234	6824	1271
TESTLAB	112	94	3075	660	3187	754
TRANSLA	3	3	1239	257	1242	260
WHOLESA	11498	10824	45535	4238	57033	15062
<b>TOTAL</b>	<b>38368</b>	<b>35388</b>	<b>388543</b>	<b>41050</b>	<b>426911</b>	<b>76438</b>

**Counts of sampling units.**

SURVEY	Pop (C)	Sample(C)	Pop(S)	Sample(S)	Pop(T)	Sample(T)
ACCBOOK	124	99	14476	611	14600	710
AQUACUL	15	15	322	148	337	163
BOOKPUB	21	20	1421	1421	1442	1441
COURIER	127	127	2734	318	2861	445
DATAPUB	12	12	75	75	87	87
EMPLOYM	130	115	1763	521	1893	636
FOODSER	669	341	50623	3266	51292	3607
GEOMATI	27	26	2289	533	2316	559
LESSORS	643	369	44394	2837	45037	3206
MANAGEM	318	206	38646	2217	38964	2423
MINESUP	221	181	554	224	775	405
NEWPAPE	64	61	415	102	479	163
NRSTORE	268	233	3918	2899	4186	3132
REALEST	59	41	27632	2480	27691	2521
RETSTOR	3823	2922	99436	9658	103259	12580
RMAUTOM	181	104	22674	2203	22855	2307
RMOTHER	268	205	15935	2197	16203	2402
SPDESIG	37	36	6784	1233	6821	1269
TESTLAB	82	65	3067	652	3149	717
TRANSLA	3	3	1239	257	1242	260
WHOLESA	7129	6611	44391	3874	51520	10485
<b>TOTAL</b>	<b>14221</b>	<b>11792</b>	<b>382788</b>	<b>37726</b>	<b>397009</b>	<b>49518</b>

## Technical Series - Index

April 4, 2001

PIPES has a series of technical paper reprints primarily for internal purposes. A list of the reprints currently available is presented below. For copies, contact Bonnie Bercik at (613) 951-6790 or Diane Proulx at (613) 951-7192, fax number (613) 951-0411 or write to Statistics Canada, 13<sup>th</sup> Floor, Jean Talon Building, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

1. Unified Enterprise Statistics Program – Project to Improve Provincial Economic Statistics – May 5, 1997 – PIPES Project Managers.
2. PIPES Evaluation Framework – September 15, 1997 – Philip Smith.
3. Report on the Unified Enterprise Survey & Reporting Arrangements Business Consultations – August 1997 – Guy Gellatly, Larry Murphy and Junior Smith.
4. Update on PIPES Progress: Notes for a Briefing for Federal and Provincial Finance Officials, Halifax, Nova Scotia, March 12, 1997 – Philip Smith.
5. An Overview of The Project to Improve Provincial Economic Statistics – November 1997 – George Beelen, Francine Hardy and Don Royce.
6. Using Databases to Design, Generate and Store Business Questionnaires at Statistics Canada – November 5, 1997 – Alana M. Boltwood.
7. The How and Why of Business Statistics – January 1999 – Elise Mennie. *(Not for external dissemination)*
8. An update on PIPES Fifteen Months into the Project – April 24, 1998 – Philip Smith.
9. Key Provider Manager (KPM) – 1997-98 Annual Report – May 1998 – Vicki Crompton.
10. A Framework for Planning Unified Enterprise Survey Data Collection – October 28, 1998 – Alana Boltwood.
11. Impact of the PIPES Funding on the Services Division Programme and Achievements in 1997-98 – April 1998 – Gordon Baldwin. *(Not for external dissemination)*
12. PIPES Organization and Decision-Making Structure – August 17, 1998 – Philip Smith. *(Not for external dissemination)*
13. The Central Goal of PIPES – November 17, 1997 – Philip Smith.
14. The Terminology and Framework of the Unified Enterprise Questionnaire – Revised March 1999 – Philip Smith.
15. Realizing and Measuring Quality Improvements in Provincial Economic Accounts – August 1998 – Philip Smith.
16. Annual Report 1997-98 – Ombudsman for Small Business Response Burden – July 1998, – Michael Issa. *(Not for external dissemination)*
17. Decision Making in PIPES – October 1, 1998 – Philip Smith.
18. Task Force on Electronic Data Reporting – April, 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. PIPES Information Package – October 1998 – Philip Smith.
20. UES and the Non-Business Sectors – September 17, 1997 – Art Ridgeway.
21. CATS User Guide – April 1998 – Janet Howatson. *(Not for external dissemination)* Not available

22. Report on Collection and Data Capture Operation OID for UES 1997 – September 3, 1998 – Anne Ladouceur. *(Not for external dissemination)*
23. SDD Contribution to PIPES 1998-1999 – September 1998 – Shirley Dolan.
24. The Harmonized Sales Tax Revenue Allocation Formula – August 1998 – Karen Hall. *(Not for external dissemination)*
25. Task Group on Data Acquisition Strategy Report – July 22, 1998 – François Maranda and Don Royce.
26. Roles and Responsibilities in the Unified Enterprise Statistics Program – December 15, 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan and Philip Smith. *(Not for external dissemination)*
27. Paper on the Project to Improve Provincial Economic Statistics from the Joint IASS/IAOS Conference – July 22, 1998 – Don Royce.
28. Respondent Relations Task Force – March 5, 1999 – Wayne Smith.
29. Response Analysis Follow-up Survey – March 1999 – Kristen Underwood.
30. Data Sharing Information Package – March 1999 – John Crysdale. *(Not for external dissemination)*
31. Coherence Analysis – Case Study from the Key Provider Manager Program – April 23, 1999 – Rachel Bernier and Julie Mandeville.
32. Evaluation of Collection Support Material used during the 1997 Unified Enterprise Survey – November 16, 1998 – Yvele Paquette.
33. Waiver Information Package – May 1999 – John Crysdale. *(Not for external dissemination)*
34. The PIPES Plan for 1999-00 – June 14, 1999 – Philip Smith. *(Not for external dissemination)*
35. BTS + Forum Post-conference Actions – April 1999 – Cornwall Conference Participants.
36. Report of the Task Force on Sources of Business Information – March 1999 – Vicki Crompton and Mark Marcogliese.
37. Field 5 Task Force Report on Improving Generic Boards – August 1999 – Mel Jones.
38. Study of Business Survey Questionnaires – June 1999 – Jason Gilmore.
39. Complexity Scale for Business Questionnaires – June 1999 – Jason Gilmore.
40. Update on PIPES – September 1999 – Philip Smith.
41. Exclusion Thresholds & Sampling Practices for Business Surveys – Implementation Strategy – September 1999 – Implementation Strategy Team.
42. Use of Tax Data in the Production of Provincial Economic Statistics – October 1999 – Peter Bissett.
43. Data Quality Survey 1996 – March 1999 – Ed Bunko. *(Not for external dissemination)*
44. Estimates of Information Cost to Business Respondents, 1998 – September 16, 1999 – Linda Grant and Michael Issa. *(Not for external dissemination)*
45. Data Security Task Force – January 1999 – Mark Steski. *(Not for external dissemination)*
46. Key Provider Manager Program – 1998-99 Program Report – June 1999 – Monique Gaudreau.
47. Treatment of Head offices and other support units in the UES – October 1999 – Charles Delorme.

48. UES Walkthrough – December 6, 1999 – Philip Smith.
49. The UES Strategy to Combine Enterprise, Establishment and Legal-Entity Data – November 29, 1999 – Philip Smith.
50. The Data Quality of the 1997 Unified Enterprise Survey (UES) Pilot – January 10, 2000 – Stuart Pursey.
51. Options for the Survey of Household Spending – January 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. *(Not for external dissemination)* AWAITING APPROVAL BY MIKE SHERIDAN
52. PIPES Budget Targets for 2001-02 and Beyond – March 28, 2000 – Project Management Team. *(Not for external dissemination)*
53. Unincorporated Business Statistics from the Tax Estimates Program – March 8, 2000 – Philip Smith.
54. PIPES and the Use of Statistics for Administrative Purposes – April 12, 2000 – Philip Smith.
55. Information Manual for Interviewers on the 1999 Survey of the Construction Industry – February 2000 – Claude Grenier. *(Not for external dissemination)*
56. The Conceptual and Practical Evolution of the Unified Enterprise Survey Integrated Questionnaire – May 2000 – Mario A. Vella and Annette Laurent.
57. Statistics Canada's Broad Strategy for Business Statistics – May 25, 2000 – Philip Smith.
58. Calendarizing Business Survey Data – June 2000 – Peter S. Tarassoff.
59. Report of the Task Force on Disclosure Avoidance Practices for Business Data – October 2000 – Daphne Bennett, John Crysdale, Ziad Ghanem, Pat Grainger, Tony Labillois, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitridge and Elaine Wilson. *(Not for external dissemination)*
60. Coherence Analysis of Large Enterprise Data – June 2000 – Monique Gaudreau and Jim Tebrake.
61. An Overview of the 1998 GIFI Database – July 26, 2000 – Jennifer Sarah Jones.
62. Head Offices and Other Ancillary Units in the Unified Enterprise Survey – May 2000 – Malika Hamdad and Michel Tessier.
63. Exploring the Potential of EDA: Illustrated with examples from the Canadian Food Services Industry Annual Survey – June 2000 – Jim Tebrake, Daphne Bennett and Larry Murphy.
64. The Challenge of Collecting Quality Data from Large Enterprises: Lessons learned from the Key Provider Manager Approach – June 2000 – Monique Gaudreau and Janet Hughes.
65. A New Approach to Processing for the Unified Enterprise Statistics Program – An Edit and Imputation Example – June 2000 – Colleen Martin and Jean-François Laroche.
66. Data Allocation – A new survey process for business surveys – June 2000 – Charles Delorme.
67. Harmonizing Survey Content: Integrating the Annual Survey of Manufactures into the Unified Enterprise Survey – June 2000 – John Crysdale.
68. Challenges Involved in Using Tax Data for Statistical Purposes – November 27, 2000 – Philip Smith.
69. Sampling Designs for The Unified Enterprise Surveys: The Early Years – March 2001 – Michelle Simard, Claude Girard, Marie-Noëlle Parent and Jocelyn Smith.

## **Other documents**

PIPES Project Structure – Revised October 2000

Confidentiality of Statistical Data – April 1995 – D. Binder and L. Desramaux

The Business Register Hellerman Autocoder Project – November 1999 – Business Register Division  
(English version only)

The Why of Business Surveys – March 2000

64. Le défi que représente la collecte de données de qualité auprès des grandes entreprises : Enseignements tirés de la stratégie des gestionnaires des répondants clés – juin 2000 – Monique Gaudreau et Janet Hughes.
65. Une nouvelle stratégie de traitement des données pour le programme unifié des statistiques auprès des entreprises – Un exemple de vérification et d'imputation – juin 2000 – Colleen Martin et Jean-François Laroche.
66. Répartition des données – Un nouveau processus pour les enquêtes-entreprises – juin 2000 – Charles Delorme.
67. Harmonisation du contenu des enquêtes : Intégration de l'Enquête annuelle des manufactures à l'Enquête unifiée auprès des entreprises – juin 2000 – John Crysdale.
68. Défis liés à l'utilisation des données fiscales à des fins statistiques – le 27 novembre 2000 – Philip Smith.
69. Les plans d'échantillonnage des enquêtes unifiées sur les entreprises : Les premières années – mars 2001 – Michelle Simard, Claude Girard, Marie-Noëlle Parent et Jocelyn Smith.

## **Autres documents**

La structure du projet du PASEP – Révisé en octobre 2000

La confidentialité des données statistiques – avril 1995 – D. Binder et L. Desramaux

The Business Register Hellerman Autocoder Project – novembre 1999 – Division du registre des entreprises  
(version anglaise seulement)

Le pourquoi des enquêtes-entreprises – mars 2000

44. Estimations des coûts d'information pour les entreprises répondantes, 1998 – le 16 septembre 1999 – Linda Grant et Michael Issa. (*Diffusion interne seulement*)
45. Groupe de travail de la sécurité des données – janvier 1999 – Mark Steski. (*Diffusion interne seulement*)
46. Programme des gestionnaires des répondants clés – Rapport de programme pour 1998-1999 – juin 1999 – Monique Gaudreau.
47. Traitement des sièges sociaux et autres unités de support dans l'EUE – le 28 octobre 1999 – Charles Delorme.
48. Tour d'horizon de l'EUE – le 6 décembre 1999 – Philip Smith.
49. Stratégie de l'EUE en vue de combiner les données au niveau de l'entreprise, de l'établissement et de l'entité légale – le 29 novembre 1999 – Philip Smith.
50. La qualité des données de l'Enquête-pilote unifiée de 1997 auprès des entreprises (EUE) – le 10 janvier 2000 - Stuart Pursey.
51. Options pour l'Enquête sur les dépenses des ménages – janvier 2000 – Cynthia Baumgarten, David Binder, Louis-Marc Ducharme, Alison Hale, Albert Meguerditchian, Mike Sheridan, M.P. Singh, Philip Smith, Maryanne Webber, Brian Williams, Karen Wilson. (*Diffusion interne seulement*) EN ATTENTE DE L'APPROBATION DE MIKE SHERIDAN
52. Budget projeté du PASEP pour les exercices 2001-2002 et ultérieurs – le 28 mars 2000 – Équipe de gestion de projet. (*Diffusion interne seulement*)
53. Statistiques sur les entreprises non constituées en société du Programme des estimations fiscales – le 8 mars 2000 – Philip Smith.
54. Le PASEP et l'utilisation des statistiques à des fins administratives – le 12 avril 2000 – Philip Smith.
55. Manuel d'information des intervieweurs – Enquête de 1999 sur l'industrie de la construction – février 2000 – Claude Grenier. (*Diffusion interne seulement*)
56. Évolution conceptuelle et pratique du questionnaire intégré de l'Enquête unifiée auprès des entreprises – mai 2000 – Mario A. Vella et Annette Laurent.
57. Stratégie générale de Statistique Canada relativement à la statistique des entreprises – le 25 mai 2000 – Philip Smith.
58. La calendrialisation des données des enquêtes-entreprises – juin 2000 – Peter S. Tarassoff.
59. Rapport du Groupe de travail sur les pratiques de contrôle de la divulgation des données sur les entreprises – octobre 2000 – Daphne Bennett, John Crysdale, Ziad Ghanem, Pat Grainger, Tony Labillois, Serge Lavallée, Jackie Leblanc, Robert Masse, Bruno Pépin, Randall Sheldrick, Leslie Shinder, Patricia Whitridge et Elaine Wilson. (*Diffusion interne seulement*)
60. Analyse de la cohérence des données sur les grandes entreprises – juin 2000 – Monique Gaudreau et Jim Tebrake.
61. Aperçu de la base de données de l'IGRF 1998 – le 26 juillet 2000 – Jennifer Sarah Jones.
62. Sièges sociaux et autres unités de soutien dans le cadre de l'Enquête unifiée auprès des entreprises – mai 2000 – Malika Hamdad et Michel Tessier.
63. Exploration des possibilités de l'EDA : Illustration à partir d'exemples de l'Enquête annuelle sur le secteur des services de restauration au Canada – juin 2000 – Jim Tebrake, Daphne Bennett et Larry Murphy.

21. Le guide de l'utilisateur du SASC – avril 1998 – Janet Howatson. (*Diffusion interne seulement*) Non disponible
22. Compte rendu de la collecte et la saisie de données DOI pour l'EUE de 1997 – le 3 septembre 1998 – Anne Ladouceur. (*Diffusion interne seulement*)
23. Contribution prévue de la DDS au PASEP, 1998-1999 – septembre 1998 – Shirley Dolan.
24. La formule de répartition des recettes de la taxe de vente harmonisée – août 1998 – Karen Hall. (*Diffusion interne seulement*)
25. Groupe de travail sur l'acquisition des données auprès des entreprises – le 22 juillet 1998 – François Maranda et Don Royce.
26. Rôles et responsabilités dans le cadre du Programme unifié des statistiques sur les entreprises – le 15 décembre 1998 – George Andrusiak, Richard Barnabé, Albert Meguerditchian, Ray Ryan et Philip Smith. (*Diffusion interne seulement*)
27. Document sur le Projet d'amélioration des statistiques économiques Provinciales de la conférence mixte de l'AISE/AISO – le 22 juillet 1998 – Don Royce.
28. Groupe de travail sur les relations avec les répondants – le 5 mars 1999 – Wayne Smith.
29. Enquête de suivi et d'analyse des réponses – mars 1999 – Kristen Underwood.
30. Dossier d'information sur le partage des données – mars 1999 – John Crysdale. (*Diffusion interne seulement*)
31. Analyse de cohérence -- Étude de cas du programme des Gestionnaires des répondants clés – le 23 avril 1999 – Rachel Bernier et Julie Mandeville.
32. Évaluation des documents de soutien de la collecte utilisés durant l'Enquête unifiée sur les entreprises de 1997 – le 16 novembre 1998 – Yvele Paquette.
33. Trousse d'information sur les renonciations – mai 1999 – John Crysdale. (*Diffusion interne seulement*)
34. Plan du PASEP pour 1999-2000 – le 14 juin 1999 – Philip Smith. (*Diffusion interne seulement*)
35. Forum SCE + Mesures de suivi de la conférence – avril 1999 – Participants de la conférence de Cornwall.
36. Rapport du groupe d'étude des sources d'information sur les entreprises – mars 1999 – Vicki Crompton et Mark Marcogiiese.
37. Rapport du Groupe de travail du secteur 5 sur l'amélioration de la dotation générique – août 1999 – Mel Jones.
38. Étude des questionnaires des enquêtes auprès des entreprises – juin 1999 – Jason Gilmore.
39. Échelle de complexité des questionnaires sur les entreprises – juin 1999 – Jason Gilmore.
40. Rapport sur l'avancement du PASEP – septembre 1999 – Philip Smith.
41. Seuils d'exclusion et méthodes particulières d'échantillonnage pour les enquêtes-entreprises – Stratégie de mise en oeuvre – septembre 1999 – L'Équipe de la stratégie de la mise en oeuvre.
42. Utilisation des données fiscales pour la production des statistiques économiques provinciales – octobre 1999 – Peter Bissett.
43. Enquête sur la qualité des données de 1996 – mars 1999 – Ed Bunko. (*Diffusion interne seulement*)

Dans le cadre du PASEP, on a réimprimé une série de documents techniques, principalement pour usage interne. Voici la liste des réimpressions disponibles. Pour obtenir des copies communiquez avec Bonnie Bercik au (613) 951-6790 ou Diane Proulx au (613) 951-7192, numéro de télécopieur (613) 951-0411 ou écrire à Statistique Canada, 13<sup>e</sup> étage, Immeuble Jean Talon, Parc Tunney, Ottawa, Ontario, K1A 0T6

1. Programme unifié des statistiques sur les entreprises – Projet d'amélioration des statistiques économiques provinciales – le 5 mai 1997 – Programme de Gestionnaire du PASEP.
2. Cadre d'évaluation du PASEP – le 15 septembre 1997 – Philip Smith.
3. Rapport de l'Enquête unifiée sur les entreprises et les modalités de déclaration – Consultations auprès des entreprises – août 1997 – Guy Gellatly, Larry Murphy et Junior Smith.
4. Bilan de l'évolution du PASEP : Notes d'une séance d'information à l'intention des représentants fédéral et provinciaux des finances, qui a eu lieu à Halifax (Nouvelle-Écosse), le 12 mars 1997 – Philip Smith.
5. Aperçu du Projet d'amélioration des statistiques économiques provinciales – novembre 1997 – George Beelen, Francine Hardy et Don Royce.
6. Des bases de données pour la conception, la génération et le stockage des questionnaires-entreprises à Statistique Canada – le 5 novembre 1997 – Alana M. Boltwood.
7. La statistique des entreprises : sa raison d'être – janvier 1999 – Elise Mennie. (*Diffusion interne seulement*)
8. Bilan du PASEP 15 mois après son lancement – le 24 avril 1998 – Philip Smith.
9. Programme des gestionnaires des répondants clés (GRC) – Rapport annuel pour 1997-1998 – mai 1998 – Vicki Crompton.
10. Un cadre de planification de la collecte des données de l'Enquête unifiée sur les entreprises – le 28 octobre 1998 – Alana Boltwood.
11. Répercussions du financement du PASEP sur le programme et les réalisations de la Division des services en 1997-1998 – avril 1998 – Gordon Baldwin. (*Diffusion interne seulement*)
12. L'organisation et la structure décisionnelle du PASEP – le 17 août 1998 – Philip Smith. (*Diffusion interne seulement*)
13. Les buts principaux du PASEP – le 17 novembre 1997 – Philip Smith.
14. Terminologie et cadre de référence du questionnaire de l'Enquête unifiée sur les entreprises – Révisé en mars 1999 – Philip Smith.
15. Amélioration de la qualité des statistiques économiques provinciales et mesure des changements apportés – août 1998 – Philip Smith.
16. Rapport annuel 1997-1998 – Médiateur – Fardeau de réponse de la petite entreprise – juillet 1998 – Michael Issa. (*Diffusion interne seulement*)
17. Le processus décisionnel du PASEP le 1<sup>er</sup> octobre 1998 – Philip Smith.
18. Groupe de travail sur la déclaration électronique des données (DED) – avril 1998 – George Andrusiak, Monique Gaudreau, Laurie Hill, Anne Ladouceur, Denis Leblanc, Mario Ménard, Guy Parent, Joe Wilkinson, Doug Zinnicker.
19. Trousse d'information sur le PASEP – octobre 1998 – Philip Smith.
20. L'EUE et les secteurs non commerciaux – le 17 septembre 1997 – Art Ridgeway.

**Comptes des unités d'échantillonnage**

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
TENUE DE LIVRES	124	99	14 476	611	14 600	710
AQUACULTURE	15	15	322	148	337	163
ÉDITEURS DE LIVRES	21	20	1 421	1421	1 442	1 441
MESSAGERIES	127	127	2 734	318	2 861	445
ÉDITEURS DE DONNÉES	12	12	75	75	87	87
EMPLOI	130	115	1 763	521	1 893	636
RESTAURATION	669	341	50 623	3 266	51 292	3 607
GÉOMATIQUE	27	26	2 289	533	2 316	559
BAILLEURS	643	369	44 394	2 837	45 037	3 206
GESTION	318	206	38 646	2 217	38 964	2 423
EXTRACTION DE MINÉRAIS	221	181	554	224	775	405
JOURNAUX	64	61	415	102	479	163
DÉTAILLANTS HORS MAGASIN	268	233	3 918	2 899	4 186	3 132
IMMOBILIER	59	41	27 632	2 480	27 691	2 521
COMMERCE DE DÉTAIL	3 823	2 922	99 436	9 658	103 259	12 580
R ET E AUTOS	181	104	22 674	2 203	22 855	2 307
R ET E AUTRES	268	205	15 935	2 197	16 203	2 402
DESIGN SPÉCIALISÉ	37	36	6 784	1 233	6 821	1 269
LAB D'ESSAI	82	65	3 067	652	3 149	717
TRADUCTION	3	3	1 239	257	1 242	260
COMMERCE DE GROS	7 129	6 611	44 391	3 874	51 520	10 485
<b>TOTAL</b>	<b>14 221</b>	<b>11 792</b>	<b>382 788</b>	<b>37 726</b>	<b>397 009</b>	<b>49 518</b>

PÉRIODIQUE	88	61	655	155	743	216
IMMOBILIER	54	29	24 428	1 685	24 482	1 714
COMMERCE DE DÉTAIL	3 708	3 607	91 948	9 826	95 656	13 433
R ET E AUTOS	164	87	23 146	1 578	23 310	1 665
R ET E AUTRES	269	174	16 138	1 137	16 407	1 311
DESIGN SPÉCIALISÉ	41	40	6 338	1 092	6 379	1 132
ÉDITEURS SPÉCIALISÉS	0	0	148	148	148	148
TAXIS- LIM.	8	7	5 972	419	5 980	426
LAB D'ESSAI	95	72	3 249	418	3 344	490
TRADUCTION	3	3	1 135	142	1 138	145
COMMERCE DE GROS	7 948	7 477	4 3548	5 388	51 496	12 865
TOTAL	15 465	13 538	483 678	40 902	499 143	54 440

### Annexe 6 : Taille de l'échantillon pour l'EUE 2000

#### Comptes des établissements

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
TENUE DE LIVRES	400	372	14 497	624	14 897	996
AQUACULTURE	19	19	348	173	367	192
ÉDITEURS DE LIVRES	37	36	1 422	1 422	1 459	1 458
MESSAGERIES	241	241	2 739	323	2 980	564
ÉDITEURS DE DONNÉES	14	14	76	76	90	90
EMPLOI	260	242	1 788	532	2 048	774
RETAURATION	3 884	3 426	51 639	3 650	55 523	7 076
GÉOMATIQUE	33	32	2 311	553	2 344	585
BAILLEURS	1 058	739	44 797	3 172	45 855	3 911
GESTION	402	267	38 658	2 222	39 060	2 489
EXTRACTION MINÉRAIS	480	431	567	233	1 047	664
JOURNAUX	285	282	449	131	734	413
DÉTAILLANTS HORS MAGASIN	543	507	3 956	2 921	4 499	3 428
IMMOBILIER	84	65	27 685	2 508	27 769	2 573
COMMERCE DE DÉTAIL	18 000	16 950	102 267	11 669	120 267	28 619
R ET E AUTOS	520	429	22 747	2 233	23 267	2 662
R ET E AUTRES	457	378	15 962	2 219	16 419	2 597
DESIGN SPÉCIALISÉ	38	37	6 786	1 234	6 824	1 271
LAB D'ESSAI	112	94	3 075	660	3 187	754
TRADUCTION	3	3	1 239	257	1 242	260
COMMERCE DE GROS	11 498	10 824	45 535	4 238	57 033	15 062
TOTAL	38 368	35 388	388 543	41 050	426 911	76 438

**Annexe 5 : Taille de l'échantillon pour l'EUE 1999**

**Comptes des établissements**

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
TENUE DE LIVRES	324	297	13 820	664	14 144	961
AQUACULTURE	41	39	311	173	352	212
ÉDITEURS DE LIVRES	41	38	380	123	421	161
CONSTRUCTION	1 319	998	118 572	7 056	119 891	8 054
MESSAGERIES	218	218	2 715	385	2 933	603
ÉDITEURS DE DONNÉES	20	20	100	100	120	120
RESTAURATION	3 850	3 323	50 430	2 810	54 280	6 133
GÉOMATIQUE	39	38	2 161	437	2 200	475
BAILLEURS	943	627	41 573	2 596	42 516	3 223
GESTION	425	268	34 820	1 934	35 245	2 202
JOURNAUX	334	332	457	148	791	480
DÉTAILLANTS HORS MAGASIN	234	234	3 308	3 306	3 542	3 540
PÉRIODIQUE	178	146	659	158	837	304
IMMOBILIER	77	49	24 481	1 700	24 558	1 749
COMMERCE DE DÉTAIL	17 858	17 749	94 591	11 640	112 449	29 389
R ET E AUTOS	472	377	23 210	1 599	23 682	1 976
R ET E AUTRES	458	322	16 161	1 153	16 619	1 475
DESIGN SPÉCIALISÉ	43	42	6 341	1 094	6 384	1 136
ÉDITEURS SPÉCIALISÉS	0	0	148	148	148	148
TAXIS-LIM.	10	9	5 972	419	5 982	428
LAB D'ESSAI	134	111	3 253	420	3 387	531
TRADUCTION	3	3	1 135	142	1 138	145
COMMERCE DE GROS	11 936	11 366	44 545	5 767	56 481	17 133
<b>TOTAL</b>	<b>38 957</b>	<b>36 606</b>	<b>489 143</b>	<b>43 972</b>	<b>528 100</b>	<b>80 578</b>

**Comptes des unités d'échantillonnage**

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
TENUE DE LIVRES	125	100	13 800	661	13 925	761
AQUACULTURE	17	15	302	167	319	182
ÉDITEURS DE LIVRES	23	20	379	122	402	142
CONSTRUCTION	909	626	118 440	7 041	119 349	7 667
MESSAGERIES	141	141	2 708	378	2 849	519
ÉDITEURS DE DONNÉES	17	17	99	99	116	116
RESTAURATION	680	310	49 418	2 422	50 098	2 732
GÉOMATIQUE	34	33	2 139	418	2 173	451
BAILLEURS	599	316	41 172	2 270	41 771	2 586
GESTION	331	194	34 799	1 924	35 130	2 118
JOURNAUX	67	65	421	118	488	183
DÉTAILLANTS HORS MAGASIN	144	144	3 296	3 294	3 440	3 438

**Annexe 4 : Taille de l'échantillon pour l'EUE 1998**

**Comptes des établissements**

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
TENUE DE LIVRES	337	294	12 720	721	13 057	1015
AQUACULTURE	36	34	315	175	351	209
MESSAGERIES	s.o.	s.o.	s.o.	s.o.	2 445	599
ÉDITEURS DE DONNÉES	s.o.	s.o.	s.o.	s.o.	102	81
EMPLOI	233	176	1 649	344	1 882	520
RESTAURATION	3 570	3 052	47 854	3 143	51 424	6 195
GÉOMATIQUE	s.o.	s.o.	s.o.	s.o.	1 883	405
BAILLEURS	814	494	38 561	2 680	39 375	3 174
GESTION	436	185	26 849	1 357	27 285	1 542
JOURNAUX	249	226	460	152	709	378
AGENTS IMMOBILIERS	114	88	19 783	1 822	19 897	1 910
DESIGN SPÉCIALISÉ	s.o.	s.o.	s.o.	s.o.	5 504	1 090
ÉDITEURS SPÉCIALISÉS	s.o.	s.o.	s.o.	s.o.	146	91
LABO D'ESSAI	132	76	2 331	293	2 463	369
TRADUCTION	s.o.	s.o.	s.o.	s.o.	1 147	193
COMMERCE DE GROS	11 464	9 849	41 571	5 678	53 035	15 527
<b>TOTAL</b>	<b>17 385</b>	<b>14 474</b>	<b>19 2093</b>	<b>16 365</b>	<b>220 705</b>	<b>33 298</b>

### Annexe 3 : Taille de l'échantillon pour l'EUE 1997

#### Comptes des établissements

ENQUÊTE	Pop. (C)	Échantillon (C)	Pop. (S)	Échantillon (S)	Pop. (T)	Échantillon (T)
AQUACULTURE	S.O.	S.O.	S.O.	S.O.	238	121
CONSTRUCTION	S.O.	S.O.	S.O.	S.O.	110 078	8 167
MESSAGERIES	S.O.	S.O.	S.O.	S.O.	998	326
RESTAURATION	S.O.	S.O.	S.O.	S.O.	54 235	4 242
BAILLEURS	S.O.	S.O.	S.O.	S.O.	15 676	1 681
AGENTS IMMOBILIERS	S.O.	S.O.	S.O.	S.O.	4 223	581
TAXIS-LIM.	S.O.	S.O.	S.O.	S.O.	828	321
<b>TOTAL</b>	S.O.	S.O.	S.O.	S.O.	<b>18 6276</b>	<b>15 439</b>

Veuillez noter ce qui suit pour les annexes suivantes.

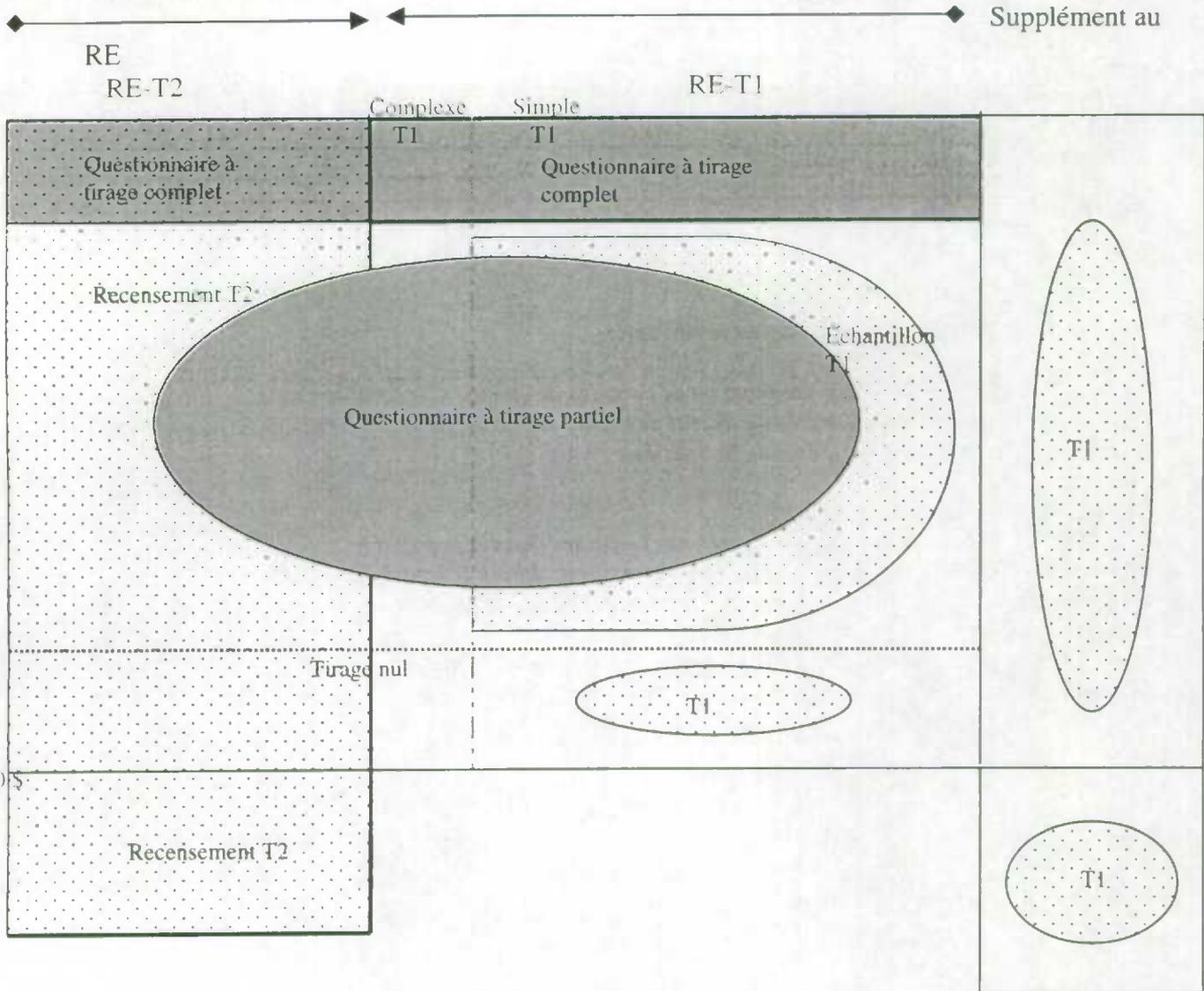
*Il s'agira de comptes des établissements ou de comptes des unités d'échantillonnage.  
Elles ne comprennent que les unités au-dessus des seuils d'exclusion.  
Elles ne comprennent que les unités de la deuxième phase.*

*Voici la cotation utilisée.*

**Pop. (C) :** Nombre d'entreprises complexes dans la population.  
**Échantillon (C) :** Nombre d'entreprises complexes dans l'échantillon.  
**Pop. (S) :** Nombre d'entreprises simples dans la population.  
**Échantillon (S) :** Nombre d'entreprises simples dans l'échantillon.  
**Pop. (T) :** Nombre total d'entreprises dans la population.  
**Échantillon (T) :** Nombre total d'entreprises dans l'échantillon.

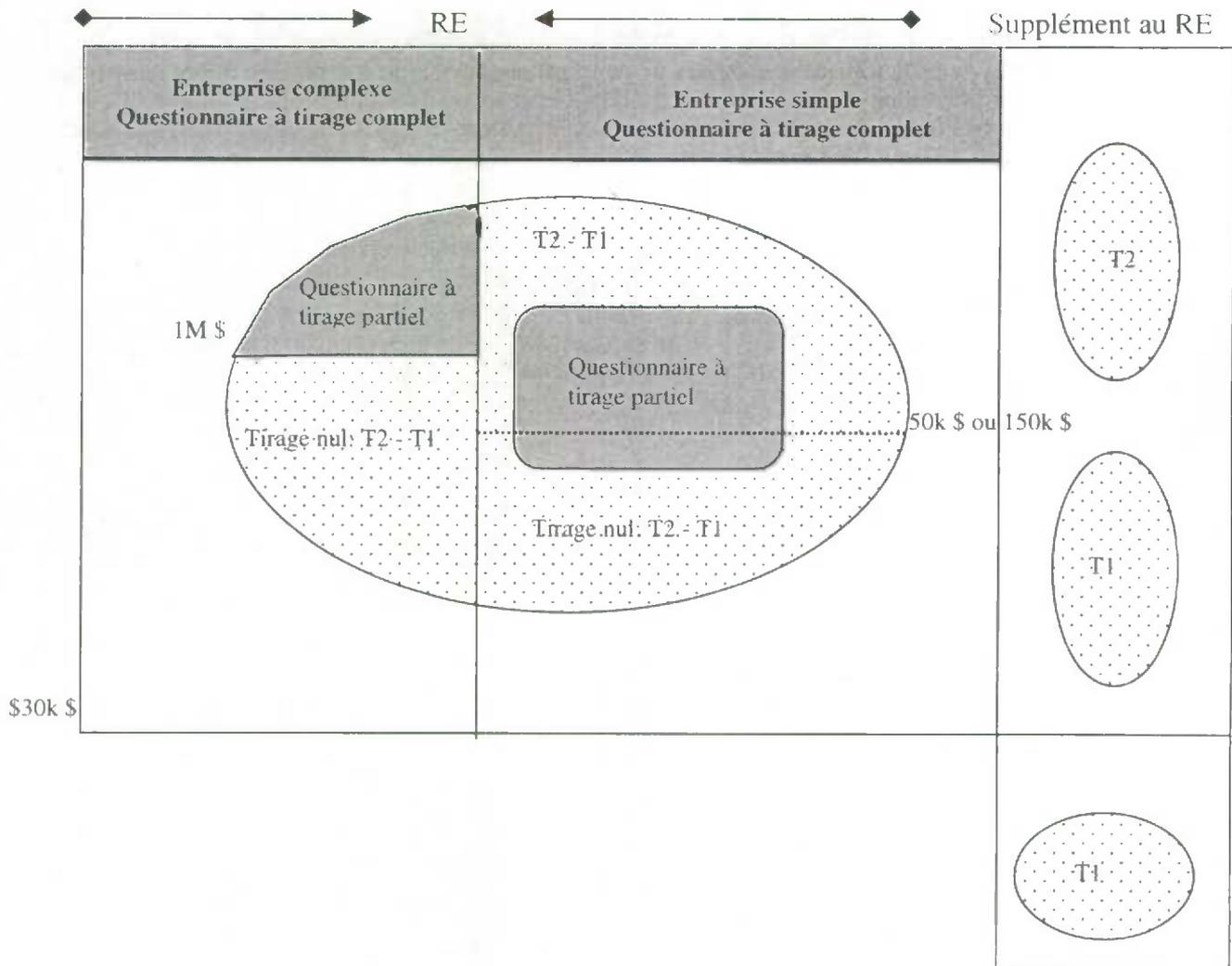
EUE 1998, 1999 et 2000

EUE 1998 – EUE 1999



Annexe 2 : Population cible, échantillons et seuils d'exclusion

EUE 1997 :



cohérence. Le FUE et le FCE renferment les « unités décédées », puisque celles-ci font partie de la population cible de l'EUE.

#### **Fichier de l'interface d'enquête (FIE)**

Le FIE diffère sur le plan conceptuel, puisqu'il renferme les renseignements les plus à jour à l'égard de chaque entité commerciale sélectionnée. Il récupère de la DDRC les renseignements de collecte les plus récents pour les unités sélectionnées du FCE. Chaque gestionnaire de projet est responsable de son enquête et doit opter pour la personnalisation et la création d'entités de collecte (EC). Certaines unités sont combinées en une EC, et des rapports combinés sont établis. Tous ces renseignements sont entrés dans le FIE. Celui-ci renferme également des renseignements sur la collecte, l'état de la couverture, l'adresse, les dénominations sociales, qui n'apparaissent pas dans le FUE, ni dans le FCE. Il est important pour les relations avec le répondant de lui envoyer les renseignements les plus à jour.

## **Annexe 1 : Fichiers de départ**

Les fichiers de départ sont composés de trois fichiers de base du RE : le fichier de l'univers d'enquête (FUE), le fichier de contrôle de l'échantillon (FCE) et le fichier de l'interface d'enquête (FIE). Le FUE est le point de départ de toute enquête auprès des entreprises qui s'appuie sur le RE. Il présente, à une date donnée, la population cible de l'enquête. Pour l'EUE, le FUE couvre toutes les unités du RE, qui étaient toujours actives au cours de l'année de référence. Le concept « toujours active » est passé de <active au moins un jour > (1997 et 1998) à <active plus d'un jour > (1999 et 2000).

### **Fichier de l'univers d'enquête (FUE) :**

Le FUE est un portrait figé de la population cible. Il s'agit de l'image centrale pour CAPEX, la DOFI et l'EAM, de même que pour toutes les industries dans le champ de l'EUE. Les programmes de la série pour le processus d'échantillonnage permettent de déterminer l'ensemble des attributs de la population à partir du RE-FUE. De ces programmes fondés sur le FUE, nous obtenons : 1) les liens entre les entités statistiques, c'est-à-dire l'entreprise et les établissements (pour ce qui est de la répartition, des rapports combinés et de l'analyse de cohérence); 2) les liens entre les entités statistiques et les dossiers administratifs, c'est-à-dire les T2 et les T1 (pour tous les recours aux dossiers de données fiscales : la vérification et imputation, la répartition, la concordance et l'approche de l'estimation à deux phases; 3) la détermination des entreprises simples, des entreprises complexes et des entreprises à entités légales multiples, conformément à la définition de l'EUE, dont on se sert dans les fonctions subséquentes.

Au début du PASEP, on a décidé de centraliser toutes les populations d'enquête en un FUE, vu que le PASEP vise fondamentalement à intégrer les enquêtes annuelles auprès des entreprises en une base commune. Ce sont toutes les parties concernées qui devaient décider de la date réelle d'établissement du FUE. C'est à partir de cette première vue d'ensemble dans le temps 1) que nous définissons les unités à tirage complet (25 millions de dollars); 2) que nous établissons les établissements en grappes dans les unités d'échantillonnage; 3) que nous attribuons des probabilités de sélection et des attributs de stratification; 4) que nous obtenons le compte de la population dans chaque strate; 5) que nous appliquons l'échantillonnage par réseau et 6) que nous sélectionnons les unités qui recevront les questionnaires afin d'obtenir un échantillon représentatif, entre autres choses.

### **Fichiers de l'univers d'enquête post-stratification (post-FUE)**

Comme elle savait que le FUE comportait des erreurs de couverture et de classification, la Méthodologie a proposé une technique de stratification a posteriori, la post-stratification dans le cadre de l'approche de l'estimation, ce qui signifie fondamentalement qu'on obtiendra ultérieurement un deuxième portrait figé dans le temps de la population cible. Il s'agit du FUE a posteriori, qui comprend des mises à jour subséquentes en provenance de toutes de sources indépendantes à l'enquête (ajout des nouvelles entreprises, retrait des entreprises ayant cessé leurs activités, le passage d'une entreprise simple à une entreprise complexe et les corrections apportées au SCIAN) qui représentent toujours la même année de référence. Ainsi, la DRE et les gestionnaires d'enquêtes disposent de plus de temps pour corriger la base. Le deuxième portrait peut être considéré comme le dernier portrait à l'égard duquel des statistiques sur les établissements et l'entreprise seront produites. Il est établi par l'entremise d'une pondération finale des unités sélectionnées à l'étape de l'estimation.

### **Fichier de contrôle de l'échantillon (FCE)**

Le FCE représente la même population que le FUE, en plus de renfermer l'ensemble des indicateurs créés dans les programmes d'échantillonnage, comme il est décrit ci-dessus, et l'indication que les unités doivent ou non recevoir le questionnaire. Tous les renseignements d'échantillonnage sont consignés dans le FCE. En cas de plans de renouvellement, les renseignements seraient consignés dans le FCE. Deux FCE sont créés. Le FCE fondé sur les établissements renferme les renseignements d'échantillonnage du CAPEX, de l'EAM, de l'EUE, de même que du bureau central. Le FCE fondé sur l'entreprise comprend les renseignements d'échantillonnage de la DOFI. Les deux FCE représentent la population de l'entreprise ou des établissements à la même date aux fins de l'analyse de

## **14 REMERCIEMENTS**

Les auteurs aimeraient remercier Normand Laniel et Claude Poirier pour leurs conseils dans le cadre du projet, de même que Guylaine Dubreuil pour sa participation aux premières enquêtes. Ils aimeraient également remercier Joe Kresovic et Stuart Pursey pour leurs commentaires utiles.

## **15 RÉFÉRENCES**

- Bankier, M.D. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. *The American Statistician* **42**, 174-177.
- Cuthill, I. (1990). The Statistics Canada's Business Register. Internal document.
- Daoust, Pierre, Lim, Ann (1997). Allocation of the Sample for the Stage III Redesign of the Survey of Employment, Payrolls and Hours (SEPH). Internal Working Document, Statistics Canada.
- Girard, C and Simard, M.(2000) Network Sampling: An application in a Major Survey. ICES-II,
- Lavallée, P. and Hidirolou, M.A. (1988). On the stratification of skewed population. *Survey Methodology*, no.14, pp 33-45
- Simard, M. and Hidirolou, M. (1999). Estimation For Annual Business Surveys Based On Two-Phase Network Sampling. *SSC Proceedings of the Survey methods Section*, pp 11-19.
- Simard, M. and Laniel, N. (1998). Échantillonnage et Estimation pour l'enquête unifiée sur les entreprises. *SSC Proceedings of the Survey methods Section*, pp 77-82.
- Thompson, S.K. and Seber, G.A.F. (1996). *Adaptive Sampling*. John Wiley and Sons.

**Tableau 11 : Cellules négligeables par province**

Province	Nombre dans le cas A	Nombre de cellules dans la province	Ratio (en %)
Terre-Neuve	18	86	20,9
Île-du-Prince-Édouard	20	81	24,7
Nouvelle-Écosse	5	86	5,8
Nouveau-Brunswick	6	85	7,1
Québec	0	87	0
Ontario	0	86	0
Manitoba	1	86	1,2
Saskatchewan	3	86	3,5
Alberta	1	86	1,2
Colombie-Britannique	0	86	0
Yukon	21	75	28
Territoires du Nord-Ouest	27	74	36,5
Nunavut	12	43	27,9
Total Canada	114	1 047	10,9

L'application de cette procédure durant la production a amené une étape supplémentaire. Il a été convenu que le gestionnaire de chaque enquête examine les cellules définies comme négligeable pour approbation finale, avant de décider de ne pas envoyer de questionnaire.

## 12 ÉCHANTILLON DE PREMIÈRE PHASE : INFORMATION AUXILIAIRE

L'échantillon de première phase possède les mêmes caractéristiques d'échantillonnage que celles dont il a été question dans les sections précédentes, sauf qu'on utilise des c.v. globaux inférieurs pour produire une taille d'échantillon supérieure. Par contre, il existe deux différences mineures : i) le type de renseignements obtenus pour les unités sélectionnées, c'est-à-dire ce sont des dossiers fiscaux, plutôt que des questionnaires et ii) un niveau de stratification supplémentaire, à savoir le type de dossiers fiscaux. L'ADRC fournit deux types de dossiers fiscaux. Il s'agit des dossiers qui s'appliquent aux entreprises constituées en société, les T2, et des dossiers qui concernent les entreprises non constituées en société, les T1. Depuis l'introduction de l'Index général des renseignements financiers (IGRF) en 1998, l'échantillon T2 constitue en fait un recensement. Les dossiers fiscaux sont utilisés uniquement pour les unités simples à titre d'information auxiliaire dans le cadre de la technique de calage choisie.

## 13 ORIENTATION FUTURE

En plein développement depuis sa mise en œuvre en 1997, l'EUE est en période de stabilisation depuis 2000. Après l'enquête sur le commerce de gros en 1998, l'enquête sur le commerce de détail en 1999 et l'enquête sur le secteur de la fabrication en 2000, aucune autre enquête annuelle d'importance ne sera intégrée dans un avenir proche. La prochaine étape pour ce qui est des enquêtes intégrées à l'EUE sera l'adoption possible d'un plan de renouvellement pour 2001 qui permettra de contrôler davantage le fardeau de réponse.

**Tableau 9 :** Résultats de la mise à l'essai de différents centiles

Scénario	Cas A	Cas B	Cas C	Cas D
SAS et SAP= centile 40 %	154 (798)	44 (2 356)	191 (1 295)	658 (375 320)
SAS et SAP= centile 35 %	132 (646)	26 (1 147)	213 (1 447)	676 (376 529)
SAS et SAP= centile 30 %	114 (521)	10 (345)	231 (1 572)	692 (377 331)
SAS et SAP= centile 25 %	95 (371)	3 (78)	250 (1 722)	699 (377 598)
SAS et SAP= centile 20 %	76 (220)	0 (0)	269 (1 873)	702 (377 676)

Le tableau 10 présente la distribution pour les 114 cellules définies comme négligeables, c'est-à-dire le cas A, au moyen du centile choisi, 30 %. Le tableau 11 présente la distribution par province.

**Tableau 10 :** Cellules négligeables par enquête

<u>Enquête</u>	Nombre dans le cas A	Nombre de cellules dans l'enquête	Ratio (en %)
Services de comptabilité et de tenue de livres	0	13	0
Aquaculture	0	12	0
Messageries	4	26	15,4
Services d'emploi	0	13	0
Services de restauration	3	52	5,8
Géomatique	4	24	16,7
Bailleurs	3	62	4,8
Gestion	0	13	0
Extraction de minerais	3	24	12,5
Éditeurs de journaux	0	12	0
Agents immobiliers	6	38	15,8
Commerce de détail	29	301	9,6
Réparation et entretien de véhicules automobiles	3	35	8,6
Réparation et entretien (autres)	4	38	10,5
Services spécialisés de design	11	55	20,4
Laboratoires d'essai	1	13	7,7
Services de traduction	3	13	23,1
Commerce de gros	40	304	13,2
Total	114	1 047	10,9

En appliquant les trois critères, on obtient l'une des quatre situations suivantes :

- A - La cellule compte moins que  $k$  unités d'échantillonnage, et les ratios AP et AS sont petits. La cellule est négligeable selon les deux critères. Aucun questionnaire ne doit être envoyé.
- B - La cellule comporte au moins  $k$  unités d'échantillonnage, mais les ratios AP et AS sont petits. La cellule comporte un nombre suffisant d'unités d'échantillonnage, mais son activité économique est faible. Il faut procéder à l'échantillonnage, sans toutefois utiliser des facteurs d'amélioration, comme les taux de suréchantillonnage.
- C - La cellule comporte moins que  $k$  unités d'échantillonnage, mais au moins un des ratios ou les deux ratios AP et AS sont évalués comme étant suffisamment importants. La cellule ne comporte pas un nombre suffisant d'unités d'échantillonnage, mais son activité économique est importante. Il faut la recenser.
- D - La cellule comporte au moins  $k$  unités d'échantillonnage et au moins un ou les deux ratios AP et AS sont évalués comme étant suffisamment importants. La cellule est importante selon tous les critères. Elle doit faire partie de l'échantillonnage régulier.

### **11.1 Définition des critères**

La définition des trois paramètres pose un problème, à savoir i)  $k$ , ii) la valeur seuil du ratio AP (SAP) et iii) la valeur seuil du ratio AS (SAS), de même que le niveau auquel la comparaison sera établie. Comme le ratissage par le quotient est d'abord appliqué au niveau sectoriel (voir le tableau 9), il a été décidé d'utiliser le même niveau. Pour un secteur donné, une cellule est « comparée » à l'ensemble des autres cellules au sein du même secteur. Il est important de souligner que certains secteurs sont très importants tandis que d'autres le sont moins. C'est le cas par exemple de l'aquaculture qui ne dispose que d'une stratification SCIAN pour son secteur, tandis que le commerce de gros en a environ 25. Même si l'approche peut sembler pratique, une même valeur de SAP et SAS ne peut pas s'appliquer à l'ensemble des secteurs. Par exemple, disons que la valeur seuil est établie à 15 % et que le revenu de l'aquaculture et du commerce de gros est réparti également dans chaque cellule. L'aquaculture, comptant une seule dimension, aura toutes ces cellules recensées parce que la proportion de la cellule représente automatiquement 100 % du total. En revanche, pour le commerce de gros, qui compte 25 dimensions (4 % pour chaque cellule), il sera impossible mathématiquement d'être au-dessus du seuil de 15 %, même si le total de revenu est supérieur dans chaque cellule à celui de l'aquaculture.

Cela ne pose pas de problème pour le paramètre  $k$ . Celui-ci est établi à 15 comme pour les années précédentes. Afin de déterminer le SAP et le SAS, on a proposé d'utiliser une approche qui s'appuie sur les centiles. Même si la valeur est différente, le centile utilisé demeure le même. En utilisant le même centile pour tous les secteurs, on obtient une certaine cohérence dans la méthode et la valeur résultante est « adaptée » à chaque secteur. Différentes valeurs des centiles ont été étudiées afin d'en arriver au meilleur compromis.

### **11.2 Résultats préliminaires**

Pour le fichier-test de population de l'EUE 2000, environ 400 000 unités d'échantillonnage sont réparties dans environ 1 200 cellules, toutes admissibles au processus d'échantillonnage. Le tableau 9 présente certains résultats de la mise à l'essai des différentes valeurs des centiles. Le premier chiffre est un compte des cellules et celui entre parenthèses est le compte correspondant d'unités d'échantillonnage.

Pour introduire un exposant provincial, les changements suivants ont été apportés.

$$CV_p = \frac{CV_s GBI_s}{\sqrt{\sum_p GBI_p^2}} \quad (4) \quad \text{est remplacé par} \quad CV_p = \frac{CV_s GBI_s}{GBI_p^{pexp} \sqrt{\sum_p GBI_p^{2-2pexp}}} \quad (8)$$

$$1/2 * \frac{CV_n GBI_n}{\sqrt{\sum_p GBI_{n*p}^2}} \quad (5) \quad \text{est remplacé par} \quad 1/2 * \frac{CV_n GBI_n}{GBI_{n*p}^{pexp} \sqrt{\sum_p GBI_{n*p}^{2-2pexp}}} \quad (9)$$

On a étudié différentes combinaisons d'exposants pour finalement opter pour un exposant provincial de 0,10 afin d'éviter que les c.v. provinciaux globaux varient trop et, du coup, respecter la restriction relative à la TVH.

#### 10.4 EUE 2000 : Répartition du c.v.

Dans le cadre de l'EUE 2000, on a opté pour l'application en deux étapes telle que définie à la section 9.2. Le c.v. cible global a été établi à 0,65 % et l'exposant industriel dans la première application était de 0,25. Pour la deuxième application, l'exposant industriel était de 0,20 et l'exposant provincial, de 0,10.

### 11 DÉFINITION DES CELLULES NÉGLIGEABLES

Autrefois, toutes les cellules comportant moins de 15 unités d'échantillonnage étaient automatiquement recensées. Quelques préoccupations ont été soulevées au sujet de l'approche. On avait fixé le nombre de 15 parce que l'algorithme de L-H vise à créer trois strates, ce qui est impossible à partir d'une population comptant moins de 15 unités d'échantillonnage. Comme conséquence, cela a augmenté le fardeau de réponse tout en améliorant très légèrement les estimations. Pour réduire davantage le fardeau de réponse, on a introduit une nouvelle caractéristique pour l'EUE 2000. En plus des seuils de R-M, certains critères permettaient de définir et d'identifier les cellules négligeables. Une fois celles-ci repérées, les cellules ne seraient pas automatiquement recensées comme les années précédentes.

Ainsi, on a obtenu une définition à deux dimensions d'une cellule négligeable. D'abord, le cas d'une petite cellule quant au nombre d'unités qu'elle renferme. C'est-à-dire, si plus petit qu'un certain nombre  $k$  d'unités, la cellule est considérée comme négligeable. Puis, une petite cellule quant à l'importance économique de la cellule relativement à l'activité globale du SCIAN (AS) et/ou à l'activité globale de la province (AP).

Pour un secteur donné, on définit deux ratios, AP et AS, comme suit :

AP = revenu total dans la cellule/revenu total dans une province donnée.

AS = revenu total dans la cellule/revenu total dans une stratification donnée du SCIAN.

L'application de cette option pour l'EUE n'a pas donné de résultats satisfaisants. Mentionnons d'abord que l'EUE renferme de nombreuses enquêtes, et que ce qu'on entend par une grande province varie d'une enquête à l'autre. Par exemple, l'aquaculture a des activités commerciales importantes à l'Île-du-Prince-Édouard, tandis que le commerce de gros, c'est en Ontario. Il faut définir ce qu'on entend par grande province dans chaque enquête, ce qui s'est avéré fastidieux. En outre, même si les grandes provinces obtenaient des c.v. inférieurs, les autres provinces moins importantes obtenaient des c.v., qui, dans la plupart des cas, excédaient les c.v. cibles maximums. Il ne s'agit pas là d'une amélioration étant donné que la plupart des c.v. se verront tout de même être modifiés.

## **10.2 Taille d'échantillon**

La deuxième option est plus compliquée et influe à divers égards sur les opérations et le système. Il s'agit fondamentalement d'établir une taille d'échantillon cible, comme on le fait pour de nombreuses enquêtes, plutôt que d'établir un c.v. cible. En examinant les tailles d'échantillon obtenues dans les EUE précédentes, on peut déterminer la taille d'échantillon prévue pour une enquête ou un secteur donné. Un algorithme permettrait alors de répartir la taille d'échantillon aux différentes cellules de l'enquête.

Il est important à ce stade-ci que le processus de répartition demeure essentiellement le même pour l'EUE pour des raisons de stabilité dans le temps. Il faut également souligner que le ratissage par le quotient, comme il a été mis en œuvre en 1999, fonctionne bien et que seuls quelques changements mineurs semblent nécessaires pour les années à venir. Cette deuxième option suppose qu'il faille modifier substantiellement les systèmes et les programmes déjà en place. C'est pourquoi on n'a pas évalué que l'option était viable pour l'instant.

## **10.3 Répartition en puissance**

Il s'agit d'appliquer aux provinces la méthode de répartition en puissance tel qu'appliquée pour les industries. Les principales inquiétudes proviennent du fait que la répartition ne convient pas aux provinces importantes. On attribue une perte d'efficacité au fait que l'importance de chaque province n'a pas été tenue en compte dans le cadre du processus de répartition. Cela découle de l'hypothèse énoncée au début du PASEP, hypothèse selon laquelle le même degré de précision doit s'appliquer à toutes les provinces. On a toutefois convenu que la contrainte de l'égalité ne pourrait jamais réellement être appliquée de façon exhaustive. Les gestionnaires ont accepté que la contrainte soit assouplie légèrement. Un c.v. presque égal est maintenant à la base de la répartition de l'échantillon. Cela étant, on peut tenir compte de l'importance relative de la province.

### **10.3.1 Introduction d'un exposant provincial**

Un exposant provincial fera en sorte que les c.v. seront inversement proportionnels au revenu total provincial. Cette relation inverse est désirable étant donné que pour la plupart des secteurs, le revenu total provincial global varie considérablement. Par conséquent, il est illogique d'attribuer la même importance à une cellule dont le revenu total correspond à 0,05 % de la province et à une autre dont le revenu total se chiffre à 45 % de la province.

$$CV_{s^*p}^{(0)} = 1/2 * \frac{CV_s GBI_s}{\sqrt{\sum_p GBI_{s^*p}^2}} + 1/2 * \frac{CV_p GBI_p}{GBI_{s^*p}^{s \exp} \sqrt{\sum_s GBI_{s^*p}^{2-2s \exp}}} \quad (5)$$

Ces c.v. sont alors utilisés dans l'algorithme du ratissage par le quotient afin de déterminer un c.v. pour chaque cellule  $s^*p$  au moyen de l'équation suivante :

$$CV_{s^*p}^{(i)} = \frac{CV_{s^*p}^{(i-1)} CV_s CV_p GBI_s GBI_p}{\sqrt{\sum_s (CV_{s^*p}^{(i-1)})^2 GBI_{s^*p}^2} * \sqrt{\sum_p (CV_{s^*p}^{(i-1)})^2 GBI_{s^*p}^2}} \quad (6)$$

$i$  est l'itération courante de l'algorithme.

Le processus de répartition utilisé accorde très peu d'importance à certaines cellules, ce qui fait que leur c.v. cible finit par être trop élevé pour certaines enquêtes. Comme on avait besoin d'un échantillon représentatif dans chaque cellule, on a décidé d'établir un plafond. Tous les c.v. cibles au-dessus de 15 % ont été ramenés à 15 %. De plus, les questionnaires de chaque enquête ont produit une liste des cellules importantes et de plafonds désirés qui varient de 5 % à 15 %. Ces ajustements ont modifié environ 40 % de tous les c.v. cibles, ce qui représente un changement considérable.

## 10 EUE 2000 : NOUVELLE PROPOSITION

Le principal changement pour 2000 est l'amélioration de la répartition de l'échantillon, c'est-à-dire la répartition du c.v.. Compte tenu de la sur-utilisation des plafonds de c.v. à la fin du processus, certaines autres options ont été évaluées. Vous en trouverez trois dans les prochaines sous-sections.

### 10.1 Facteur de multiplication

Une des options consiste à intégrer un facteur de multiplication dans la formule des c.v., faisant en sorte que certains c.v. seront systématiquement supérieurs ou inférieurs à d'autres. Il en est question dans Daoust et Lim (1986). L'Enquête sur l'emploi, la rémunération et les heures (EERH) a fait l'objet d'un remaniement important, au cours duquel on a privilégié cette méthode. Cependant pour l'EERH, comme le même niveau de précision n'était pas requis dans toutes les provinces, le c.v. provincial dans les petites provinces devait être trois fois supérieur à celui des grandes provinces, comme il est indiqué en (7) :

$$CV_p^* = \frac{CV_s GBI_s}{\sqrt{\sum_p (GBI^*)^2}} \quad (7)$$

$$GBI_p^* = \begin{cases} 3 * GBI_p & \text{pour les territoires, les provinces de l'Atlantique et des Prairies,} \\ GBI_p & \text{autrement} \end{cases}$$

$$CV_p = \begin{cases} 3 * CV_p^* & \text{pour les territoires, les provinces de l'Atlantique et des Prairies,} \\ CV_p^* & \text{autrement} \end{cases}$$

10 secteurs. La première application n'est effectuée qu'une fois : tous les secteurs sont regroupés, et on obtient 10 c.v. sectoriels. Par la suite, on exécutera 10 deuxièmes applications.

La deuxième application est effectuée en deux étapes. Les c.v. sectoriels calculés dans le premier ratissage sont attribués aux cellules de chaque secteur respectif, comme il est défini en 4.1. Les figures 7A et 7B présentent la relation entre les deux applications. Il s'agit de prendre chaque secteur séparément et d'appliquer le ratissage aux agrégations d'industries qui appartiennent au secteur donné.

Figure 7 A : Première étape : Répartition du c.v. global en c.v. sectoriels

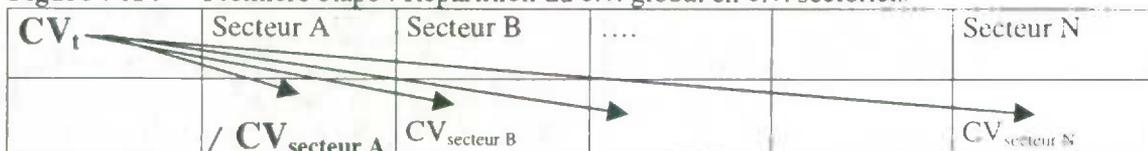


Figure 7 B : Deuxième étape : Ratissage sectoriel

$CV_{\text{secteur A}}$	Prov. 1 $CV_{\text{prov 1}}$	Prov. 2 $CV_{\text{prov 2}}$					Prov. 13 $CV_{\text{prov 13}}$
SCIAN4 A1 $CV_{\text{strate A1}}$							
SCIAN4 A2 $CV_{\text{strate A1}}$				cellule <sub>s+p</sub> $CV_{\text{cellule}}$			
SCIAN4 AJ $CV_{\text{strate AJ}}$							

La première étape consiste à répartir le c.v. sectoriel global, soit le  $c.v.\text{-sectoriel}$ , à l'ensemble de la province, et au niveau de stratification industrielle (stratification = SCIAN4 ou SCIAN5) dans chacune des marges. Il n'est pas souhaitable que le c.v. soit le même pour tous les secteurs, vu que les secteurs n'ont pas la même importance. On a appliqué la même méthode de répartition en puissance définie en (3) avec la même puissance, c'est-à-dire 0,25.

Pour le c.v. provincial,  $c.v.\text{-prov}$ , on a appliqué une contrainte de sorte que toutes les provinces aient le même c.v. provincial, qui est également calculé en fonction du c.v. cible global. On a procédé ainsi afin de satisfaire aux contraintes provinciales pour la TVH. Pour une province donnée p, le c.v.<sub>p</sub> a été établi selon l'équation (4) :

$$CV_p = \frac{CV_t GBI_t}{\sqrt{\sum_p GBI_p^2}} \quad (4)$$

Une fois les c.v. marginaux calculés, l'algorithme à la deuxième étape permet de déterminer les c.v. au niveau de la cellule. Les c.v. initiaux au niveau de la cellule, soit les  $c.v.\text{-}^{(0)}_{s+p}$ , sont calculés au moyen de l'équation suivante :

## 9.2 EUE 1999 : Ratissage par le quotient

Parce que les plans ne tenaient pas compte de l'importance relative de chaque cellule, on a adopté une nouvelle stratégie pour l'EUE 1999 : une application du ratissage par le quotient en deux étapes. Le ratissage par le quotient est un processus itératif, qui permet de distribuer des valeurs optimales dans les cellules de n'importe quel tableau en fonction de valeurs marginales fixes. Les deux applications sont décrites dans les deux prochaines sous-sections.

### 9.2.1 Première application

La première application consiste à distribuer le c.v. global, soit  $c.v.$ , c'est-à-dire le c.v. national de toutes les industries dans tous les secteurs, présenté au tableau 8. Il n'est pas souhaitable que le c.v. soit le même pour tous les secteurs, vu que ceux-ci n'ont pas la même importance. On s'est donc servi d'une méthode de répartition en puissance pour minimiser les grandes différences de c.v. entre les petits et les grands secteurs, tel que décrit dans Bankier (1988). Pour un secteur donné  $s$ , les c.v. sont calculés inversement proportionnels au total du revenu ( $RBE_s$ ), comme il est indiqué en (3). Dans les formules subséquentes, l'expression GBI représente le revenu total.

$$CV_s = \frac{CV_t GBI_t}{GBI_s^{s \exp} \sqrt{\sum_s GBI_s^{2-2s \exp}}} \quad (3)$$

Ceci parce que la variance est  $V_s = CV_s^2 GBI_s^2$  et que l'on veut obtenir  $\sum_s V_s = V_t$

Pour l'EUE 1999, le c.v. cible global se chiffrait à 0,75 %, et la puissance ( $s \exp.$ ) utilisée pour la répartition des industries était de 0,25.

Tableau 8 : Définition des secteurs

Secteur	Enquêtes	Division
11	Aquaculture	Agriculture
23	Construction	Fabrication, construction et énergie
41	Commerce de gros	Statistique du commerce
44-45	Commerce de détail	Statistique du commerce
48-49	Taxis-limousines, Messageries	Transports
51	Base de données, éditeurs spécialisés et éditeurs de journaux	Industries de service
53	Agents et courtiers immobiliers	Industries de service
54	Comptabilité, géomatique, gestion, design spécialisé, laboratoires d'essai, traduction	Industries de service
56	Emploi	Industries de service
72	Restauration	Industries de service
81	Réparation et entretien	Industries de service

### 9.2.2 Deuxième application

La deuxième application sera effectuée par secteur, et les exécutions se font de manière indépendante à l'intérieur de chaque secteur. Par exemple, disons que pour une année de référence on compte

avant que le réseau ne soit appliqué. Dans le cas de l'EUE, comme on se sert de l'EASS conventionnel, les probabilités d'inclusion dans le plan sont égales à  $n_h / N_h$  pour chaque strate  $h$ .

Disons que  $\pi_i$  et  $\pi_i^*$  représentent la probabilité d'inclusion initiale et la probabilité d'inclusion finale, c'est-à-dire après l'échantillonnage par réseau de l'unité  $i$ , respectivement. Si l'unité  $i$  est liée, au moyen de l'échantillonnage par réseau, aux unités  $i_1, \dots, i_n$  et que  $S^*$  est l'échantillonnage final, alors

$$\begin{aligned} \pi_i^* &= \text{Prob}(i \in S^*) = 1 - \text{Prob}(i \notin S^*) = 1 - \text{Prob}(i \notin S, i_1 \notin S, \dots, i_n \notin S) \\ &= 1 - \text{Prob}(i \notin S) \times \text{Prob}(i_1 \notin S) \times \dots \times \text{Prob}(i_n \notin S) = 1 - \prod_{k \in \text{Net}(i)} (1 - \pi_k) \end{aligned} \quad (1)$$

où  $\text{Net}(i)$  est la liste des unités liées à l'unité  $i$  par l'échantillonnage par réseau (y compris  $i$ ). On peut alors utiliser le poids de réseau résultant  $w_i^* = \pi_i^{*-1}$  dans un type d'estimateur Horvitz-Thompson :

$$\hat{Y}_{\text{net}} = \sum_{S^*} w_i^* y_i \quad (2)$$

On peut envisager la possibilité de recourir à la méthode du poids partagé pour calculer les poids de sondage. La méthode du poids partagé facilite le calcul de l'estimateur de la variance. Consultez Girard et Simard (2000) pour les détails sur ces deux estimateurs et les techniques d'estimation de la variance.

Voilà qui conclut les étapes requises pour la sélection des unités de l'échantillon. Les annexes 3 à 6 présentent les tailles d'échantillon finales par enquête de l'EUE de 1997 à 2000, respectivement. Les prochaines sections présentent la façon dont l'échantillon est partagé entre les enquêtes au moyen de la répartition du CV.

## 9 DÉTERMINATION DU CV DES CELLULES

Avec l'arrivée des enquêtes continues dans la plate-forme de l'EUE, la répartition de l'échantillon est devenue une étape cruciale pour l'ensemble du processus d'échantillonnage. Pour appliquer la formule de répartition de la TVH, il faut des estimations d'une précision également fiable et détaillée pour l'ensemble des provinces. Dans le cas des enquêtes à intégrer, la dimension provinciale n'est pas aussi importante puisque les enquêtes requièrent souvent des estimations très détaillées pour leurs différentes agrégations d'industries. Le processus de répartition de l'échantillonnage devait donc satisfaire aux deux exigences, celle qui s'applique à la province et celle qui concerne l'industrie, et le fardeau de réponse devait également être réduit. Cette tâche est devenue un défi.

### 9.1 EUE 1997 et EUE 1998

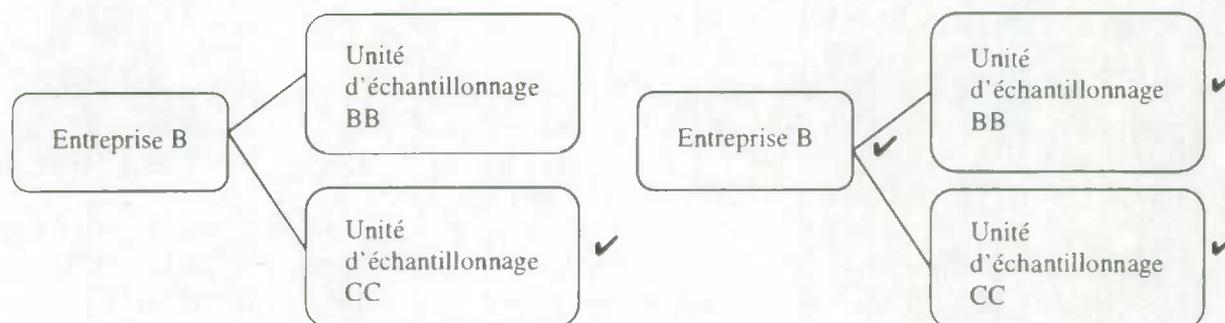
Pour l'EUE 1997 et l'EUE 1998, on a ciblé une valeur de c.v. commune pour chaque cellule, peu importe la province ou la portion de l'industrie couverte. En 1997, le c.v. cible commun était de 15 % et en 1998, il était de 7,5 %. Pour plus de renseignements sur ces plans d'échantillonnage, consultez Simard et Laniel (1998). Cette stratégie ne tenait toutefois pas compte de l'importance relative de la cellule.

seules les unités complexes posent problème, compte tenu du fait que pour les unités simples, l'entreprise et l'établissement sont exactement les mêmes. Pour atteindre tous les objectifs et en fonction des caractéristiques de la population, on a adopté une approche appelée échantillonnage par réseau. Cette technique est appliquée comme suit : une fois qu'on a obtenu un échantillon initial au moyen de l'échantillonnage aléatoire simple stratifié (EASS), on effectue des ajustements avec l'échantillonnage par réseau. L'échantillonnage par réseau est une méthode qui permet d'ajouter à l'échantillon initial des unités d'échantillonnage non sélectionnées au départ à condition qu'une unité liée (par l'entremise d'un réseau) ait été sélectionnée dans l'échantillon initial. Dans l'EUE, on lie deux unités si elles appartiennent à la même entreprise et ont des activités dans le même domaine. Pour plus de détails, consultez Simard et Hidioglou (1999).

L'échantillonnage par réseau n'a pas été introduit dans l'EUE afin d'accroître l'efficacité de l'échantillonnage, mais simplement pour compléter le « portrait d'échantillonnage » d'une entreprise complexe sélectionnée. Cela diffère d'une méthode d'échantillonnage similaire appelée échantillonnage par grappe adapté. Lire Thompson et Seber (1996) pour plus de détails à ce sujet. L'échantillonnage par grappe adapté consiste essentiellement en l'ajout d'unités sur le terrain, ce qui accroît l'efficacité de l'échantillon par rapport à l'échantillon initial. L'échantillonnage par réseau était une caractéristique de l'EUE 1997, de l'EUE 1999 et de l'EUE 2000. Dans l'EUE 1998, on s'est plutôt servi d'une approche de recensement pour l'ensemble des unités multiprovinciales.

À partir de l'exemple qui se trouve à la figure 4, supposons qu'au départ seule l'unité d'échantillonnage CC a été sélectionnée dans l'échantillon (côté gauche de la figure 6). Grâce à l'échantillonnage par réseau, les unités BB, CC et l'entreprise associée B sont incluses dans l'échantillon (côté droit de la figure 6).

Figure 6 : Avant et après l'échantillonnage par réseau



### 8.1 Probabilités d'inclusion et poids de sondage

L'échantillonnage par réseau influe considérablement sur les probabilités d'inclusion des unités et inversement sur les poids de sondage. Soulignons que par définition l'échantillonnage par réseau ne peut s'appliquer qu'aux entreprises complexes. Dans ce cas, les probabilités d'inclusion finales diffèrent souvent des probabilités d'inclusion dans le plan initial. Dans le cadre de l'échantillonnage par réseau, la probabilité qu'une unité donnée se retrouve ultimement dans l'échantillon ne dépend pas uniquement de sa probabilité d'inclusion dans le plan initial, mais aussi des probabilités d'inclusion dans le plan initial de l'ensemble des unités du réseau dans laquelle elle se situe. Soulignons que les probabilités d'inclusion dans le plan initial sont définies en fonction du plan d'échantillonnage initial, autrement dit,

**Tableau 7 : Taux des imperfections de la base pour l'EUE en pourcentage**

Enquête — Niveau du SCIAN	EUE 1997	EUE 1998	EUE 1999	EUE 2000
Services de comptabilité et de tenue de livres	s.o.	25	25	20
Aquaculture	10,2	40	40	15
Éditeurs de livres	s.o.	s.o.	s.o.	25
Construction	9,4	s.o.	22	s.o.
Messageries	7,2	20	20	15
Éditeurs de bases de données	s.o.	15	15	30
Services d'emploi	s.o.	20	s.o.	30
Services de restauration	14,9	10 – 25	10 – 20	10
Géomatique	s.o.	10 – 25	25	25
Bailleurs	5,7	10 – 25	15	s.o.
Gestion	s.o.	25	40	35
Extraction de minerais	s.o.	s.o.	s.o.	15
Éditeurs de journaux	s.o.	15	15	15
Détaillants hors-magasins	s.o.	s.o.	15	15
Agents immobiliers	5,7	15 – 35	40	25
Réparation et entretien de véhicules automobiles	s.o.	s.o.	25	s.o.
Réparation et entretien (autres)	s.o.	s.o.	25	35
Commerce de détail	s.o.	s.o.	15	15
Services spécialisés de design	s.o.	15	30	20
Éditeurs spécialisés	s.o.	15	15	s.o.
Taxis et limousines	5,2	s.o.	15	s.o.
Laboratoires d'essai	s.o.	20	50	50
Services de traduction	s.o.	15	15	25
Commerce de gros, sauf 4191	s.o.	35	50	15
Commerce de gros – 4191	s.o.	50	35	15

## 7 SÉLECTION DE L'ÉCHANTILLON

Avant de procéder à la sélection de l'échantillon, on attribue un numéro aléatoire (NA) dans l'intervalle  $[0,1)$  à l'ensemble des unités d'échantillonnage. Puis, on ordonne selon leur NA et on procède à la sélection de l'échantillon en commençant par le plus petit NA de façon à prélever le nombre d'unités d'échantillon nécessaires pour satisfaire les tailles d'échantillon prédéterminées dans chaque strate. Ceci est similaire à un échantillon aléatoire simple (EAS) dans chaque strate. Soulignons que les mêmes NA sont utilisés pour la sélection de la première phase, assurant ainsi le chevauchement complet.

## 8 ÉCHANTILLONNAGE PAR RÉSEAU

L'un des objectifs de l'EUE est d'assurer une cohérence entre les estimations au niveau de l'entreprise et celui de ces établissements correspondants. Le défi consiste donc à sélectionner tous les établissements associés à une entreprise sélectionnée sans qu'il y ait suréchantillonnage. Soulignons que

## 6.1 Taux d'inflation de l'échantillon

Une fois la taille d'échantillon obtenue pour chaque strate, les échantillons sont gonflés de façon à tenir compte des imperfections de la base, telles les entreprises mortes et les classifications erronées et de la non-réponse. Dans l'EUE 1997, le taux d'inflation de la non-réponse était de 20 % pour l'ensemble des industries. Dans l'EUE 1998, le taux s'élevait à 40 % dans la plupart des cas, sauf pour les agents immobiliers et l'emploi (15 %), le commerce de gros (20 %), la restauration (25 %) et la gestion (30 %). Comme pour l'EUE 1999, les taux varient davantage selon les industries. Le tableau 6 présente les taux de non-réponse par industrie. Les taux d'inflation qui s'appliquent aux imperfections de la base sont indiqués par industrie. Le tableau 7 présente les taux d'inflation pour les trois années de référence.

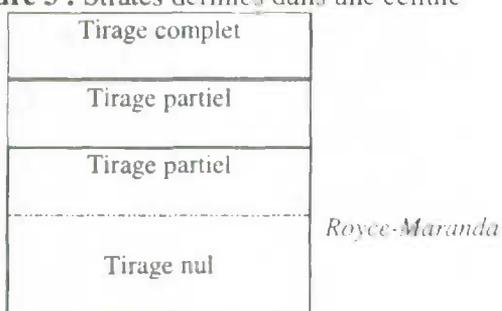
Tableau 6 : Taux de non-réponse pour l'EUE en pourcentage

Enquête — Niveau du SCIAN	EUE 1999	EUE 2000
Services de comptabilité et de tenue de livres	40	45
Aquaculture	40	45
Éditeurs de livres	s.o.	40
Construction	15	s.o.
Messageries	40	45
Éditeurs de bases de données	40	65
Services de restauration	25	55
Géomatique	40	40
Bailleurs	35	s.o.
Gestion	40	55
Extraction de minerais	s.o.	40
Éditeurs de journaux	40	30
Détaillants hors-magasins	40	40
Agents immobiliers	25	55
Réparation et entretien de véhicules automobiles	40	s.o.
Réparation et entretien (autres)	40	50
Commerce de détail	40	40
Services spécialisés de design	40	50
Éditeurs spécialisés	40	s.o.
Taxis-Limousines	40	s.o.
Laboratoires d'essai	40	70
Traduction	40	45
Commerce de gros	20	40

## 5.2 Stratification

Dans l'EUE, on utilise trois niveaux de stratification de base. Les deux premiers sont utilisés afin de satisfaire aux exigences en matière d'estimation; le dernier augmente l'efficacité du plan. D'abord, les unités d'échantillonnage sont regroupées en cellules, définies selon les provinces et les agrégations d'industries. Il s'agit des strates primaires et secondaires. Le troisième niveau de stratification s'appuie sur une mesure de taille, habituellement le revenu qui se trouve dans la base. Depuis l'EUE 1998, on applique le seuil de R-M pour représenter les portions qui se situent au-dessus et en-dessous des seuils pour chaque cellule. La portion supérieure est alors divisée en trois groupes relativement homogènes en fonction de la taille des unités. Les groupes résultants sont : une strate à tirage complet et deux strates à tirage partiel au-dessus du seuil de R-M, et une strate à tirage nul en-dessous du seuil de R-M, comme il est présenté à la figure 5. Pour la stratification, on se sert de l'algorithme de Lavallée-Hidiroglou (L-H) (1988). Cette méthode itérative consiste à déterminer les limites optimales des strates de façon à minimiser la taille de l'échantillon total pour un c.v. donné ou ciblé. Le processus de minimisation est effectué en fonction d'une méthode de répartition établie par le statisticien. La méthode utilisée dans le cas du plan de l'EUE est décrite à la section 6. Il sera question à la section 9 de la détermination réelle du c.v., c'est-à-dire de la façon de déterminer le c.v. cible de chaque cellule.

Figure 5 : Strates définies dans une cellule



## 5.3 Unités à tirage obligatoire

En plus des unités appartenant aux strates à tirage complet définies par l'algorithme L-H, un autre type d'unité est sélectionné avec certitude, les unités à tirage obligatoire. Contrairement aux unités appartenant aux strates à tirage complet, on ne définit pas ces unités comme étant au-dessus de valeurs optimales, mais plutôt comme étant au-dessus d'une seule valeur qui s'applique à toutes les industries uniformément. Dans la pratique, cela signifie qu'on sélectionne toutes les unités d'échantillonnage appartenant à une entreprise complexe dont le revenu d'entreprise est supérieur à 25 millions de dollars. Sur le plan opérationnel, les unités sont mises de côté avant l'application de l'algorithme L-H, mais sont tout de même incluses dans le calcul du c.v. cible.

## 6 RÉPARTITION DE L'ÉCHANTILLON

La taille de l'échantillon de chaque strate est définie par l'algorithme L-H, qui requiert en entrée, un c.v. cible au niveau de la cellule, c'est-à-dire les deux premières dimensions de stratification. La répartition de la taille de l'échantillon se fait de façon proportionnelle à la racine carrée du revenu total entre les trois strates.

## 5 STRATIFICATION

### 5.1 Seuils (Royce-Maranda) et stratification

Pour réduire le fardeau de réponse imposé aux petites entreprises, l'EUE a utilisé une délimitation pour fractionner la population en deux : la portion dont les unités seront éligibles à recevoir un questionnaire et la portion dont les unités ne le seront pas. La première portion sera estimée au moyen de données d'enquête et, s'il y a lieu, de données administratives. La deuxième portion, appelée la portion à tirage nul, sera estimée uniquement à partir de données administratives. Cette dernière portion comprend les petites entreprises.

#### EUE 1997 :

Dans l'EUE 1997, on appliquait des seuils après que le niveau de stratification de taille défini et après que l'échantillon ait été prélevé. Ils permettaient de délimiter les deux portions.

#### **Pour les industries de la construction et de la restauration :**

- Les entreprises simples sélectionnées dans l'échantillon dont le revenu d'établissement est de 50 000 \$ ou moins.
- Les entreprises complexes sélectionnées dans l'échantillon dont le revenu d'entreprise s'élève à un million de dollars ou moins.

#### **Pour les industries de l'aquaculture, des taxis-limousines, des messageries, des agents et courtiers immobiliers :**

- Les entreprises simples sélectionnées dans l'échantillon dont le revenu d'établissement se chiffre à 150 000 \$ ou moins.
- Les entreprises complexes sélectionnées dans l'échantillon dont le revenu d'entreprise est de un million de dollars ou moins.

#### Depuis l'EUE 1998 :

On a calculé des valeurs de seuils plus élaborés, toujours dans le but de réduire le fardeau de réponse imposé aux petites entreprises. On les a appliqués uniformément à l'ensemble des industries. La principale différence est que ces seuils sont maintenant utilisés comme bornes inférieures de stratification. Par ailleurs, les valeurs de ces seuils sont calculées de façon à garantir qu'au maximum 5 % du revenu total de la cellule est estimé uniquement à partir des données fiscales, c'est-à-dire la portion des unités à tirage nul. Il s'agit des seuils de Royce-Maranda (R-M). Le tableau 5 présente les valeurs réelles utilisées pour chaque année.

**Tableau 5 : Valeurs utilisées pour les exclusions**

EUE 1998-1999	EUE 2000
30 000	30 000
45 844	45 000
70 057	70 000
107 056	105 000
163 597	160 000
250 000	250 000

À l'annexe 2, on illustre les bases, les deux échantillons et les seuils d'exclusion.

## 4.2 Unité d'échantillonnage

L'unité d'échantillonnage est étroitement liée au concept de cellule. Comme les cellules fournissent une couverture mutuellement exclusive de la population, une entreprise peut avoir des établissements dans différentes cellules. Par conséquent, les cellules permettent de diviser une entreprise en grappes d'établissements qui sont définies comme des unités d'échantillonnage. Plus précisément, une unité d'échantillonnage est composée de l'ensemble des établissements qui appartiennent à la même entreprise et au sein de la même cellule. Tous les renseignements d'échantillonnage dont on dispose au niveau de l'établissement sont agrégés au niveau de l'unité d'échantillonnage, par exemple, la valeur du revenu qui servira à la stratification de la taille est additionnée à ce niveau. On se servira de ces valeurs pour le troisième niveau de stratification, tel que décrit à la section 5.2.

Comme le présente la figure 4, l'entreprise B est liée à trois établissements : B1, B2 et B3. Comme deux des établissements, soit B1 et B2, ont des activités au sein des mêmes agrégation d'industries et province, ils appartiennent à une seule unité d'échantillonnage, soit BB. De même, l'établissement B3 correspond à l'unité d'échantillonnage CC. Soulignons que dans le cadre de l'EUE de 1997, l'agrégation d'industries représentera un code à six chiffres du SCIAN et qu'à partir de l'EUE de 1998, le niveau de l'industrie représente un code à quatre chiffres ou un code à cinq chiffres du SCIAN, selon les branches d'activité.

Figure 4 : Unités d'échantillonnage pour l'EUE



## 4.3 Entreprises simples et complexes

Un aspect important de l'entreprise est sa complexité. On la définit comme complexe si l'un des critères suivants s'applique :

- une entreprise qui a des entités légales multiples, autrement dit, une entreprise qui a plus d'un dossier administratif dans sa structure juridique du RE;
- une entreprise qui a des établissements ayant des activités dans plusieurs provinces;
- une entreprise qui a des établissements ayant des activités dans plusieurs industries à six chiffres selon le SCIAN.

Une entreprise qui n'est pas complexe conformément à la définition précédente est définie comme étant simple. Soulignons que ces définitions ne sont pas des définitions types à Statistique Canada; il s'agit plutôt de définitions qui ne s'appliquent qu'à l'EUE. Cette distinction permet de déterminer dans quelle mesure on utilise les données d'enquête et les données administratives à l'étape de l'estimation. Comme la méthode de l'estimation s'appuie en partie sur des techniques de calage, les deux sources de données se doivent d'être au même niveau. Conséquemment, seules les unités simples répondent à cette exigence.

**Tableau 3 :** Évolution de la base de sondage de l'EUE

	Avec salariés <sup>2</sup>	Sans salariés		
		Ventes TPS supérieures à 30 000 \$	Ventes TPS inférieures à 30 000 \$	Aucun compte de TPS
<b>Entreprises constituées en société</b>	AVRIL 1987	Octobre 1997 Mai 1998 Octobre 1998 Octobre 1999	Mai 1998 Octobre 1998 Octobre 1999	Octobre 1998 Octobre 1999
<b>Entreprises non constituées en société</b>	AVRIL 1987	Octobre 1997 Mai 1998 Octobre 1998 Octobre 1999	Hors du champ pour l'instant	Hors du champ pour l'instant

#### 4 CONCEPTS S'APPLIQUANT À L'EUE

##### 4.1 Concept de cellules

La cellule est un concept essentiel pour l'EUE. Il s'agit d'une façon de fractionner la population cible. Elle se fonde sur deux dimensions importantes pour l'EUE dont on se sert également pour la stratification. Il s'agit de la dimension qui se rapporte à la province et de la dimension qui concerne l'industrie.

EUE 1997 :

Pour l'EUE 1997, la cellule consiste en un groupement de tous les établissements ayant des activités dans la même province ou le même territoire et portant le même code à six chiffres du SCIAN.

Depuis l'EUE 1998 :

Pour toutes les années suivantes, une cellule est un groupement de tous les établissements ayant des activités dans la même province ou territoire et au même niveau d'agrégation du SCIAN. On peut l'illustrer comme un tableau à deux dimensions où les lignes correspondent à l'agrégation de l'industrie et les colonnes, aux 10 provinces et trois territoires, comme il est montré au tableau 4. L'agrégation est soit un code à quatre chiffres ou un code à cinq chiffres du SCIAN.

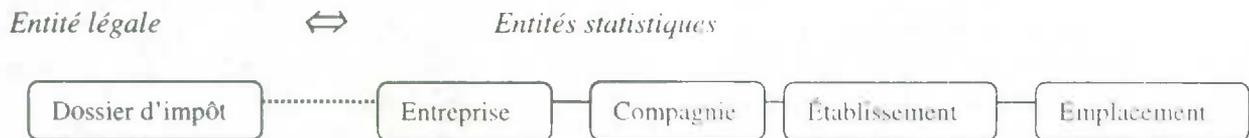
**Tableau 4 :** Illustration d'une cellule

SCIAN/Province	Province 1	Province 2	Province j	Province 13
SCIAN 2311	Cellule <sub>11</sub>	Cellule <sub>12</sub>	Cellule <sub>1j</sub>	Cellule <sub>113</sub>
SCIAN 23221	Cellule <sub>21</sub>	Cellule <sub>22</sub>	Cellule <sub>2j</sub>	Cellule <sub>213</sub>
...	Cellule <sub>i1</sub>	Cellule <sub>i2</sub>	Cellule <sub>ij</sub>	Cellule <sub>i13</sub>
SCIAN 7224	Cellule <sub>n1</sub>	Cellule <sub>n2</sub>	Cellule <sub>nj</sub>	Cellule <sub>n13</sub>

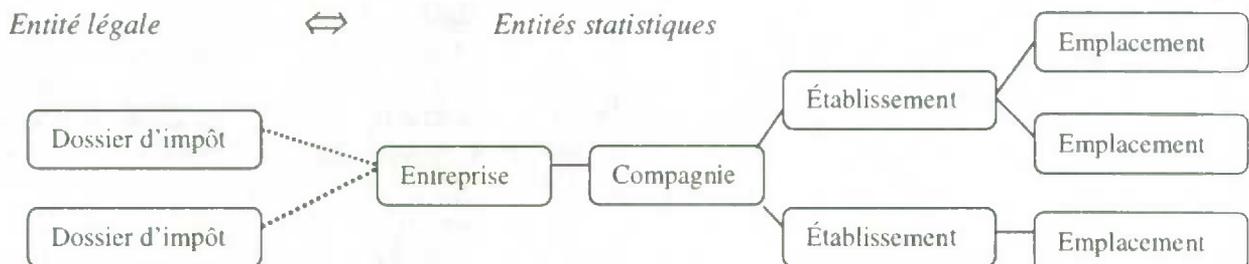
<sup>2</sup> Les entreprises avec salariés et les entreprises sans salariés sont définies selon les renseignements sur la paie disponible auprès de l'ADRC.

Un niveau important dans l'EUE est le niveau de l'entité légale. L'Agence des douanes et du revenu Canada (ADRC) définit le niveau en collaboration avec Statistique Canada en fonction de la structure juridique qui définit l'entreprise et par laquelle elle opère. L'entité légale est déterminante dans le contexte de l'EUE pour ce qui est de l'acquisition des données et de la stratégie en matière d'estimation parce que c'est à ce niveau qu'on dispose de l'information auxiliaire, c'est-à-dire les dossiers administratifs. Les deux niveaux statistiques utilisés dans le cadre de l'EUE sont l'entreprise et l'établissement. On définit l'établissement comme une entité de production physique exploitée dans une province et dans une branche d'activité données. On lui attribue un code industriel à six chiffres selon le Système de classification des industries de l'Amérique du Nord (SCIAN). L'entreprise est l'entité administrative qui gère le ou les établissements. C'est à ce niveau que les états financiers sont produits. Aucun code de province n'est attribué à l'entreprise. On lui attribue cependant le code de classification de l'industrie dominante de ses établissements correspondants. Les règles de dominance se fondent sur la variable du revenu. Les figures 2 et 3 présentent les deux types d'entreprises, de même que les relations entre les entités.

**Figure 2 : Entreprise simple**



**Figure 3 : Entreprise complexe**



### 3.1 Évolution de la couverture de la base pour l'EUE

Depuis ses débuts, la couverture de l'EUE a augmenté. Chaque année, on a ajouté à la base, différents types d'unités, ce qui a fait que la population cible a augmenté en nombre. Les types figurent au tableau 3. De juillet 1997 à octobre 1999, on a ajouté dans la base presque 1,2 million d'enregistrements d'entreprises sans salariés. Il s'agit là d'un changement d'envergure qui a des répercussions importantes sur l'analyste et les utilisateurs puisque chaque année la population cible varie. À l'avenir, on ne prévoit pas introduire de nouvelles unités dans le RE, sauf pour ce qui est des naissances de nouvelles entreprises, comme d'habitude.

Laboratoires d'essai	54138	1998, 99, 2000
Services de traduction	54193	1998, 99, 2000
Commerce de gros	41	1998, 99, 2000

Tableau 2 : Liste des enquêtes couvertes par l'EUE — Données administratives seulement

Nom de l'enquête	SCIAN	Années d'enquête <sup>1</sup>	Exercices
Construction	23	1997, 1999	1998, 2000
Taxis et limousines	4853	1997, 1999	1998, 2000

### 3 LA BASE DE SONDAGE

Tous les processus d'échantillonnage commencent par la création d'une base, qui dans le cadre de l'EUE, consiste en une liste de toutes les unités de la population cible. La base est produite par le Registre des entreprises (RE), le système de base de données de Statistique Canada qui permet de gérer, de mettre à jour et de produire tous les renseignements d'échantillonnage nécessaires aux enquêtes auprès des entreprises. Le RE a été remanié au milieu des années 1980. Il s'agit d'une base de données complexes fondée sur les renseignements tirés des dossiers administratifs. Le RE crée et gère les fichiers de population, de l'échantillon et de la collecte à l'égard de toutes les enquêtes qui l'utilisent. Les fichiers de départ de l'EUE sont abordés à l'annexe 1. La base de l'EUE couvre l'univers des entreprises canadiennes : tant les entreprises avec salariés que celles sans salariés. Elle renferme des renseignements sur chacune des entreprises, de même que sur les liens à d'autres entités statistiques ou administratives. Parmi les renseignements que l'on trouve dans le RE, mentionnons les codes des provinces et des industries, les variables se rapportant à la taille (le revenu et le nombre d'employés), l'adresse, la dénomination sociale, l'identificateur statistique unique, le numéro d'entreprise (NE) qui permet de coupler l'entité commerciale aux dossiers administratifs, les dates de démarrage et de fermeture des entreprises, les indicateurs de la complexité de la structure, des renseignements sur la couverture, etc.

On peut se servir comme unité d'échantillonnage de n'importe lequel des quatre niveaux d'entités statistiques disponibles, c'est-à-dire l'entreprise, la compagnie, l'établissement et l'emplacement qui se trouvent sur le RE. Il en est question à la figure 1. La création de ces entités statistiques se fonde sur un ensemble de règles et de critères types. On utilise souvent les deux premiers niveaux quand l'objectif de l'enquête est de produire des renseignements financiers, c'est-à-dire des revenus et des dépenses consolidés, de même que l'état financier de l'entreprise, entre autres choses. Quant aux deux derniers niveaux, on les utilise le plus souvent quand l'enquête vise à produire des renseignements opérationnels ou des données sur la production, c'est-à-dire le type d'activité, le salaire, le nombre d'unités produites, etc. Vous trouverez de plus amples détails à ce sujet dans Cuthill (1990).

Figure 1 : Entités statistiques qui figurent sur le Registre des entreprises



<sup>1</sup> Il s'agit de ou des années à l'égard desquelles un questionnaire a été envoyé ou sera envoyé dans le cadre de l'EUE.

l'information ou des données auxiliaires, en provenance des dossiers administratifs qui, dans le cadre de la stratégie d'estimation permettront de produire des estimations plus efficaces. À la section 12, on décrit certains détails sur les données administratives sélectionnés dans l'échantillon de première phase. L'échantillon de deuxième phase est l'échantillon principal qui permet d'obtenir des données d'enquête au moyen de questionnaires. On produit des renseignements financiers et non financiers afin de répondre aux exigences en matière d'estimation à l'égard de l'échantillon. La première phase chevauche complètement la deuxième phase et ce sont les mêmes unités d'échantillonnage qui sont utilisés dans les deux phases.

En 1997, on a enquêté dans le cadre de l'EUE, pour la première année de référence, sept industries pilotes qui n'avaient pas fait l'objet d'une enquête depuis longtemps. En 1998, on a intégré 16 industries supplémentaires, y compris le commerce de gros. Les années suivantes, deux autres enquêtes importantes ont été intégrées à l'EUE : le commerce de détail en 1999 et la fabrication en 2000. Cependant, pour l'EUE 1998 et de 2000, deux industries n'ont pas été enquêtées au cours de la deuxième phase, c'est-à-dire que seul l'échantillon de la première phase a été prélevé. Pour ces deux industries, les estimations financières et non financières seront établies uniquement à partir des dossiers administratifs. Les tableaux 1 et 2 présentent l'ensemble des industries enquêtées dans le cadre de l'EUE et leurs années d'activité dans la plate-forme de l'EUE.

**Tableau 1 : Liste des enquêtes couvertes par l'EUE —Données d'enquête et données administratives**

<u>Nom de l'enquête</u>	<u>SCIAN</u>	<u>Années d'activité</u>
Services de comptabilité et de tenue de livres	5412	1998, 99, 00
Aquaculture	1125	1997,98, 99, 2000
Éditeurs de livres	51113 (en fonction de l'activité)	2000
Messageries	492	1997, 98, 99, 2000
Éditeurs de bases de données et de répertoires	51114	1998, 99, 2000
Services d'emploi	5613	1998, 2000
Services de restauration et débits de boissons	722	1997, 98, 99, 2000
Services de prospection et de levées géophysiques	54136, 54137	1998, 99, 2000
Bailleurs de biens immobiliers	5311, 531310	1997, 98, 99
Gestion	5416	1998, 99, 2000
Fabrication	31, 32, 33 (y compris 113,1153)	2000
Extraction de minerais	2122,2123	2000
Éditeurs de journaux	51111	1998, 99, 2000
Détaillants hors-magasins	454, mais non 454310	1999, 2000
Détaillants hors-magasins (autres)	454310	2000
Agents immobiliers	5312, 531320, 531390	1997, 98, 99, 2000
Réparation et entretien de véhicules automobiles	8111	1999
Réparation et entretien (autres)	8112, 8113, 8114	1999, 2000
Commerce de détail	44, 45, mais non 454	1999, 2000
Services spécialisés de design	5414, 54132	1998, 99, 2000
Éditeurs spécialisés	51119	1998, 99

## **1 INTRODUCTION**

Le Projet d'amélioration des statistiques économiques provinciales (PASEP) entame sa quatrième année de référence pour l'année 2000. Ce projet a été lancé en 1996, au moment où trois provinces ont signé une entente avec le gouvernement canadien visant à harmoniser leur taxe de vente provinciale avec la taxe de vente nationale (la taxe sur les produits et services [TPS]). Dans le cadre de cette entente, le gouvernement canadien perçoit la taxe de vente harmonisée (TVH) durant l'année et à la fin de chaque exercice, il attribue la part qui revient à chaque province. Statistique Canada a le mandat de produire des estimations provinciales fiables qui serviront à établir la part provinciale. Cependant, l'infrastructure et les programmes d'enquête ne pouvaient ni appuyer ni satisfaire les nouvelles exigences.

Pour s'acquitter de ses nouvelles responsabilités, Statistique Canada a procédé à une refonte majeure de ses programmes annuels auprès des entreprises, de certains de ses programmes s'adressant aux ménages, de même qu'à une restructuration du Système de comptabilité nationale et des tableaux provinciaux d'entrées-sorties. L'une de ses principales améliorations est la mise en œuvre d'une nouvelle enquête annuelle auprès des entreprises qui deviendra le véhicule de production des estimations annuelles pour l'ensemble des industries au niveau de détails nécessaires par province et par branche d'activité : l'Enquête unifiée auprès des entreprises (EUE). L'EUE a été créée et élaborée en fonction de méthodes fiables, cohérentes et de qualité.

L'objectif du présent document est de décrire les caractéristiques d'échantillonnage de l'EUE au fil des ans. Le plan de sondage a été largement modifié depuis ses débuts en 1997. Même si l'objectif de l'enquête a toujours été de produire des estimations fiables au niveau provincial tous les ans, l'arrivée de nouvelles industries a influé sur les exigences en matière de niveaux industriels à produire. En outre, l'intégration des programmes continus dans l'échantillon a imposé des exigences historiques, qui ne correspondaient pas nécessairement à l'objectif du PASEP de contrôler le fardeau de réponse. L'importance relative de certaines industries comparativement à d'autres, de même que la dimension provinciale ont fait du processus de répartition de l'échantillon une question déterminante pour tous les gestionnaires. Pour toutes ces raisons, les échantillons ont fait l'objet d'une évaluation et d'un remaniement tous les ans, dont il est question dans le présent document.

Le présent document est structuré de la façon suivante. Sauf pour la section 12, toutes les sections traitent des caractéristiques et des paramètres de l'échantillon de deuxième phase de l'EUE pour différentes années de référence. Les premières sections présentent les industries et les caractéristiques de base de l'échantillon, les bases de sondage, de même que les concepts utilisés tout au long de l'EUE. Les paramètres d'échantillonnage sont décrits dans les autres sections. Il s'agit notamment de la stratification, de la répartition et de la sélection de l'échantillon. On traite de l'échantillonnage par réseau à la section 8. Les sections 9 et 10 décrivent les nouvelles méthodes utilisées pour calculer les coefficients de variation cibles (c.v.) d'une enquête et d'une province à l'autre. À la section 11, il est question d'une nouvelle approche visant à cerner et à traiter les cellules négligeables de façon à alléger le fardeau de réponse. La section 12 présente brièvement l'échantillon de première phase pour lequel on recueille de l'information auxiliaire en vue de la stratégie d'estimation. Enfin, aux dernières sections, on présente les orientations futures et les remerciements.

## **2 PLANS D'ENQUÊTE**

La principale caractéristique du plan est de s'appuyer sur une approche à deux phases. Deux échantillons sont prélevés tous les ans. L'échantillon de la première phase permet de recueillir de

11	Définition des cellules négligeables	19
11.1	Définition des critères	20
11.2	Résultats préliminaires	20
12	Échantillon de première phase : Information auxiliaire	22
13	Orientation future	22
14	Remerciements	23
15	Références	23

## Annexes

1	Fichiers de départ	24
2	Population cible, échantillons et seuils d'exclusion	26
3	Taille de l'échantillon pour l'EUE 1997	28
4	Taille de l'échantillon pour l'EUE 1998	29
5	Taille de l'échantillon pour l'EUE 1999	30
6	Taille de l'échantillon pour l'EUE 2000	31

## Liste des tableaux et figures

		Page
Table 1	Liste des enquêtes couvertes par l'EUE – Données d'enquête et données administratives	4
Table 2	Liste des enquêtes couvertes par l'EUE – Données administratives seulement	5
Table 3	Évolution de la base de sondage de l'EUE	7
Table 4	Illustration d'une cellule	7
Table 5	Valeurs utilisées pour les exclusions	9
Table 6	Taux de non-réponse pour l'EUE en pourcentage	11
Table 7	Taux des imperfections de la base pour l'EUE en pourcentage	12
Table 8	Définition des secteurs	15
Table 9	Résultats de la mise à l'essai de différents centiles	21
Table 10	Cellules négligeables par enquête	21
Table 11	Cellules négligeables par province	22
Figure 1	Entités statistiques qui figurent sur le Registre des entreprises	5
Figure 2	Entreprise simple	6
Figure 3	Entreprise complexe	6
Figure 4	Unités d'échantillonnage pour l'EUE	8
Figure 5	Strates définies dans une cellule	10
Figure 6	Avant et après l'échantillonnage par réseau	13
Figure 7A	Première étape : Répartition du c.v. global en c.v. sectoriels	16
Figure 7B	Deuxième étape : Ratissage sectoriel	16

# Les plans d'échantillonnage des enquêtes unifiées sur les entreprises :

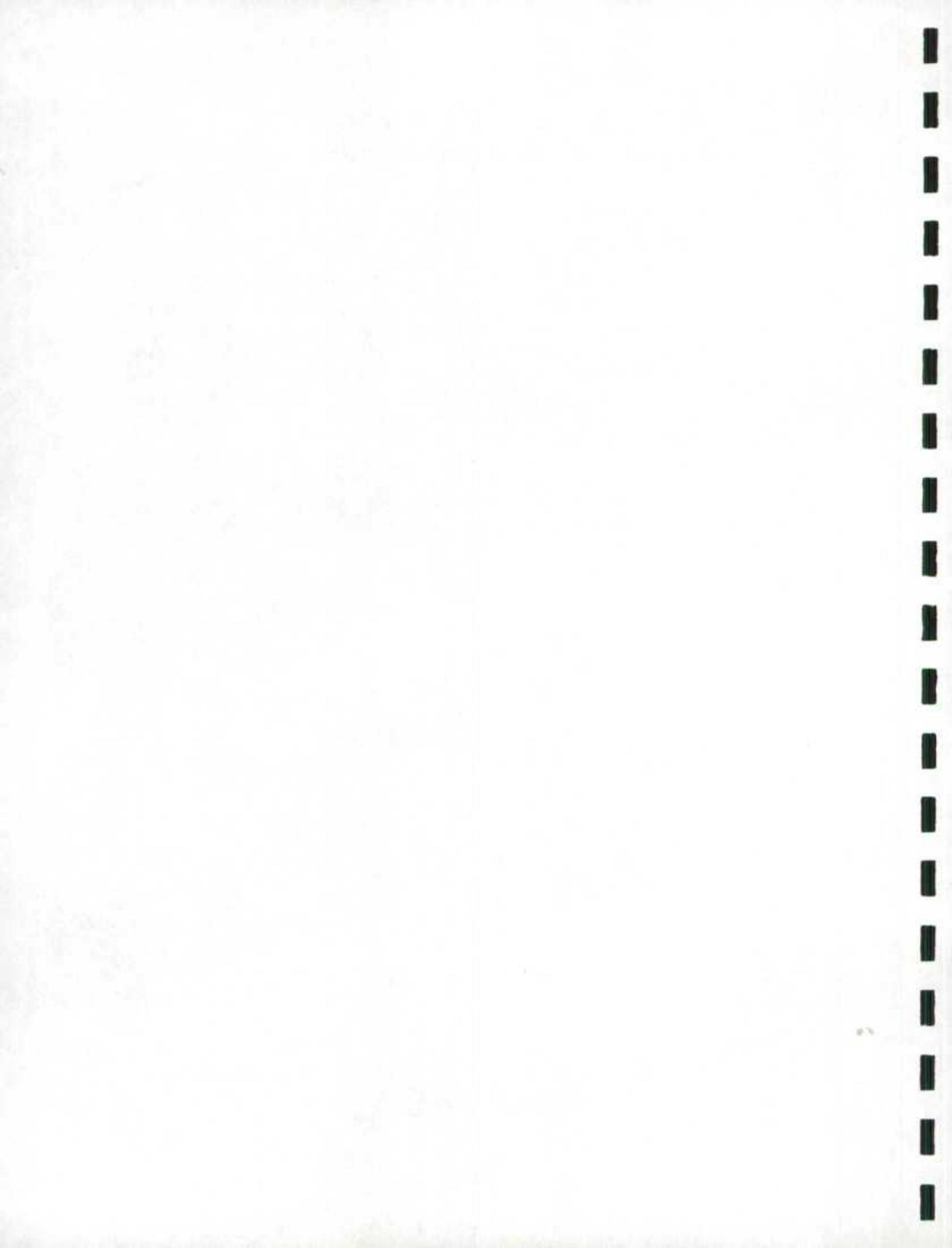
## Les premières années

Michelle Simard, Claude Girard, Marie-Noëlle Parent, Jocelyn Smith  
Division des Méthodes d'enquêtes entreprises

Mars 2001

### Table des matières

Sections	Page	
1	Introduction	3
2	Plans d'enquête	3
3	La base de sondage	5
3.1	Évolution de la couverture de la base pour l'EUE	6
4	Concepts s'appliquant à l'EUE	7
4.1	Concept de cellules	7
4.2	Unité d'échantillonnage	8
4.3	Entreprises simples et complexes	8
5	Stratification	9
5.1	Seuils (Royce-Maranda) et stratification	9
5.2	Stratification	10
5.3	Unités à tirage obligatoire	10
6	Répartition de l'échantillon	10
6.1	Taux d'inflation de l'échantillon	11
7	Sélection de l'échantillon	12
8	Échantillonnage par réseau	12
8.1	Probabilités d'inclusion et poids de sondage	13
9	Détermination du CV des cellules	14
9.1	EUE 1997 et EUE 1998	14
9.2	EUE 1999 : Ratissage par le quotient	15
9.2.1	Première application	15
9.2.2	Deuxième application	15
10	EUE 2000 : Nouvelle proposition	17
10.1	Facteur de multiplication	17
10.2	Taille d'échantillon	18
10.3	Répartition en puissance	18
10.3.1	Introduction d'un exposant provincial	18
10.4	EUE 2000 : Répartition du c.v.	19



# Les plans d'échantillonnage des enquêtes unifiées sur les entreprises :

## Les premières années

La nouvelle Enquête Unifiée sur les Entreprises (EUE) intègre différentes enquêtes annuelles de Statistique Canada. Cette enquête a été modifiée annuellement depuis son implantation en 1997. Certains concepts employés pour l'échantillonnage sont demeurés les mêmes depuis le début du projet. Ceux-ci incluent l'utilisation d'une seule base de sondage, une approche à deux phases, l'algorithme de stratification, l'instauration des numéros aléatoires pour la sélection et la répartition de l'échantillon, mais d'autres ont été modifiées. Entre autres, une nouvelle méthode de détermination du coefficient de variation pour enquête à buts multiples a été développée pour l'année de référence 1999. Également, pour l'année de référence 2000, de nouvelles stratégies ont été développées pour rencontrer les spécifications divergentes des industries en terme de précision ainsi qu'une nouvelle approche pour contrôler le fardeau de réponse.

### Note de reconnaissance

*Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.*

Pour plus de renseignements sur ce document, veuillez communiquer avec Bonnie Bercik (613) 951-6790 ou Diane Proulx (613) 951-7192  
Télécopieur : (613) 951-0411

STATISTICS CANADA LIBRARY  
B BLOTHÉQUE STATISTIQUE CANADA



1010322036

C. 2

CA 005



Projet d'amélioration des statistiques  
économiques provinciales

**Les plans d'échantillonnage  
des enquêtes unifiées sur  
les entreprises :**

**Les premières années**

Série technique

Numéro 69

Project to Improve Provincial  
Economic Statistics

**Sampling Designs for  
The Unified Enterprise  
Surveys:**

**The Early Years**

Technical Series

Number 69



Internet : [www.statcan.ca](http://www.statcan.ca)

Intranet : <http://pasep>



Statistique  
Canada

Statistics  
Canada

Canada