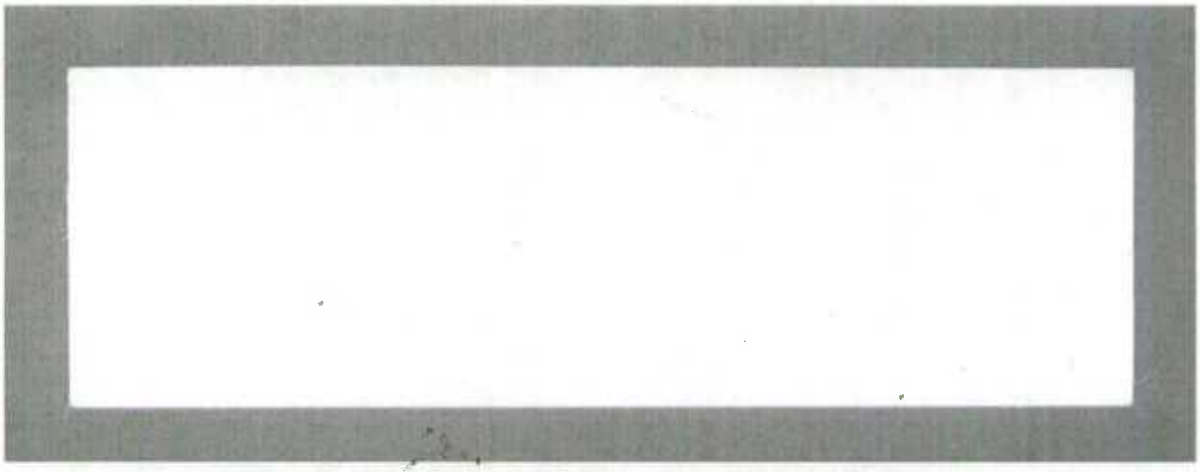
Statistics Canada Statistique Canada



Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes auprès des entreprises

Canada

SAMPLING DESIGNS FOR THE UNIFIED ENTERPRISE SURVEYS: THE EARLY YEARS

By

Michelle Simard, Claude Girard, Marie-Noëlle Parent, Jocelyn Smith

BSMD-2001-003E

SAMPLING DESIGNS FOR THE UNIFIED ENTERPRISE SURVEYS: THE EARLY YEARS

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

ABSTRACT

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.

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.

RÉSUMÉ

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.

Table of Contents

List of tables and figures

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

Survey name	NAICS covered	Activity years
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

Survey name	NAICS covered	Survey years[1]	Tax years
Construction	23	1997, 1999	1998, 2000
Taxi & Limousines	4853	1997, 1999	1998, 2000

[1] 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).

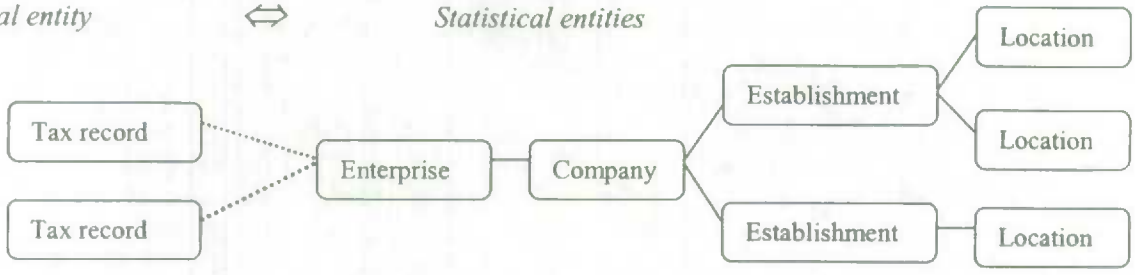
Figure 1: Statistical entities available on the Business Register

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

Legal entity ⇔ *Statistical entities*

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[2]	**Non-Employers**		
		With GST Sales above $ 30 000.	**With GST Sales below $ 30 000.**	**No GST Account**
Incorporated	APRIL 1987	October 1997 May 1998 October 1998 October 1999	May 1998 October 1998 October 1999	October 1998 October 1999
Unincorporated	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.

[2] 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

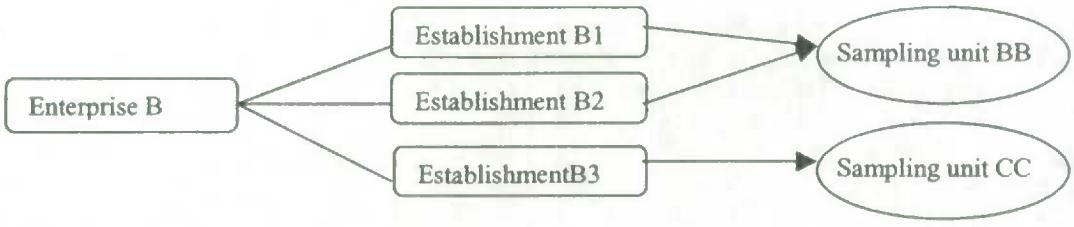
NAICS / Prov.	Prov. 1	Prov.2	Prov. j	Prov.13
NAICS 2311	$Cell_{11}$	$Cell_{12}$	$Cell_{1j}$	$Cell_{113}$
NAICS 23221	$Cell_{21}$	$Cell_{22}$	$Cell_{2j}$	$Cell_{213}$
...	$Cell_{i1}$	$Cell_{i2}$	$Cell_{ij}$	$Cell_{i13}$
NAICS 7224	$Cell_{n1}$	$Cell_{n2}$	$Cell_{nj}$	$Cell_{n13}$

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

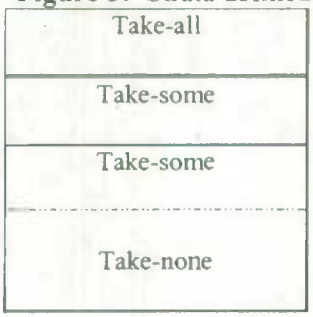
UES 1998 - 1999	UES 2000
30 000	30 000
45 844	45 000
70 057	70 000
107 056	105 000
163 597	160 000
250 000	250 000

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 stratifacation 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 references 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

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

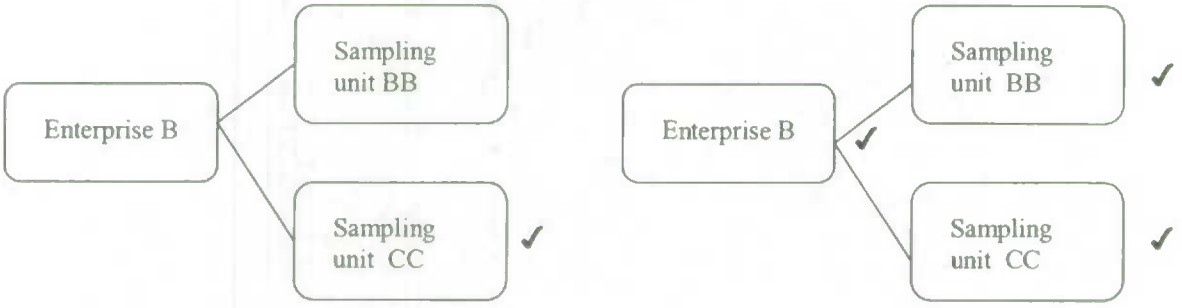
8 NETWORK SAMPLING

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 intital 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 Hidiroglou (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 π_i and π_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 $i_1,...,i_n$ and S^* is the final sample then

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

where *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, CVs are calculated to be inversely proportional to the revenue total (GBI_s) as shown in (3).

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

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

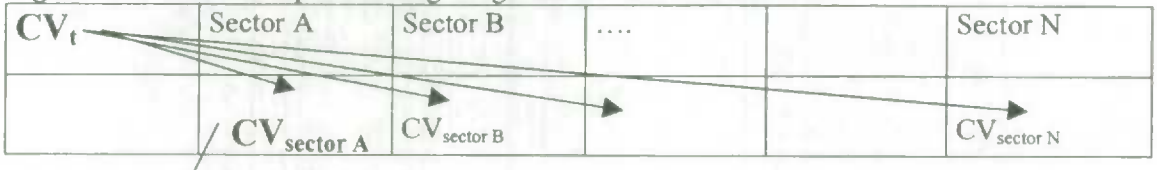


Figure 7 B: Second step: Sectorial raking

$CV_{sector\ A}$	Prov. 1 $CV_{prov\ 1}$	Prov. 2 $CV_{prov\ 2}$					Prov. 13 $CV_{prov\ 13}$
NAICS4 A1 $CV_{stratum\ A1}$							
NAICS4 A2 $CV_{stratum\ A1}$				$cell_{s*p}$ 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):

$$CV_p = \frac{CV_t GBI_t}{\sqrt{\sum_p GBI_p^2}} \tag{4}$$

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, $CV^{(0)}_{s*p}$, are computed using the following equation:

$$CV^{(0)}_{s*p} = 1/2 * \frac{CV_s GBI_s}{\sqrt{\sum_p GBI^2_{s*p}}} + 1/2 * \frac{CV_p GBI_p}{GBI^{s\exp}_{s*p} \sqrt{\sum_s GBI^{2-2s\exp}_{s*p}}} \tag{5}$$

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

$$CV^{(i)}_{s*p} = \frac{CV^{(i-1)}_{s*p} CV_s CV_p GBI_s GBI_p}{\sqrt{\sum_s \left(CV^{(i-1)}_{s*p}\right)^2 GBI^2_{s*p}} * \sqrt{\sum_p \left(CV^{(i-1)}_{s*p}\right)^2 GBI^2_{s*p}}} \tag{6}$$

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):

$$CV^*_p = \frac{CV_s GBI_s}{\sqrt{\sum_p (GBI^*)^2_p}} \tag{7}$$

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

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

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.

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

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

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
Total	114	1047	10.9

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
Total Canada	114	1047	10.9

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 form 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 aplication 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)

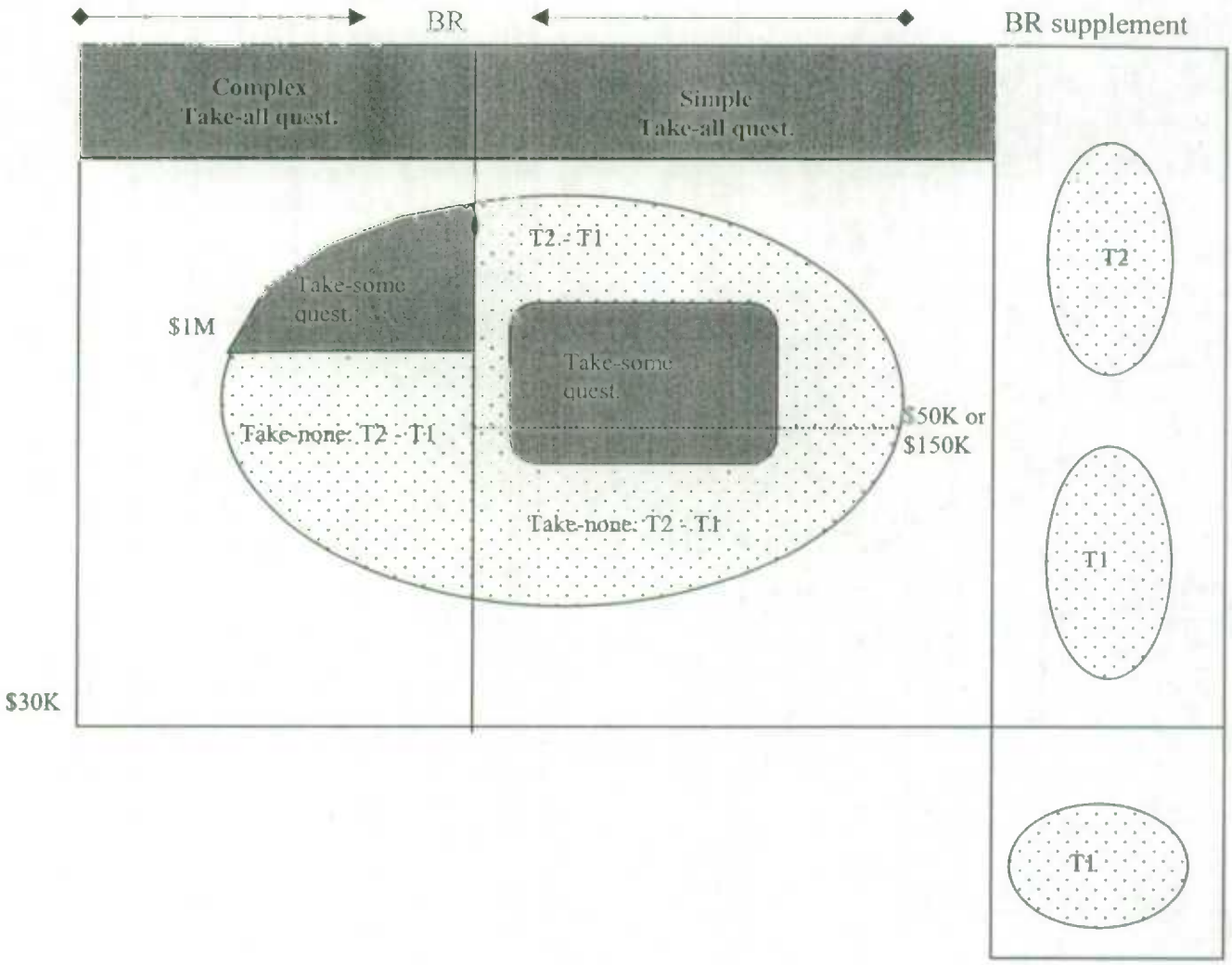
The SCF is the exact same picture of the population as the SUF, with the additions of all flags created in the sampling programs as described above plus the indication whether the units were selected to receive a questionnaire or not. All sampling information is recorded on the SCF. If there were rotation 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.

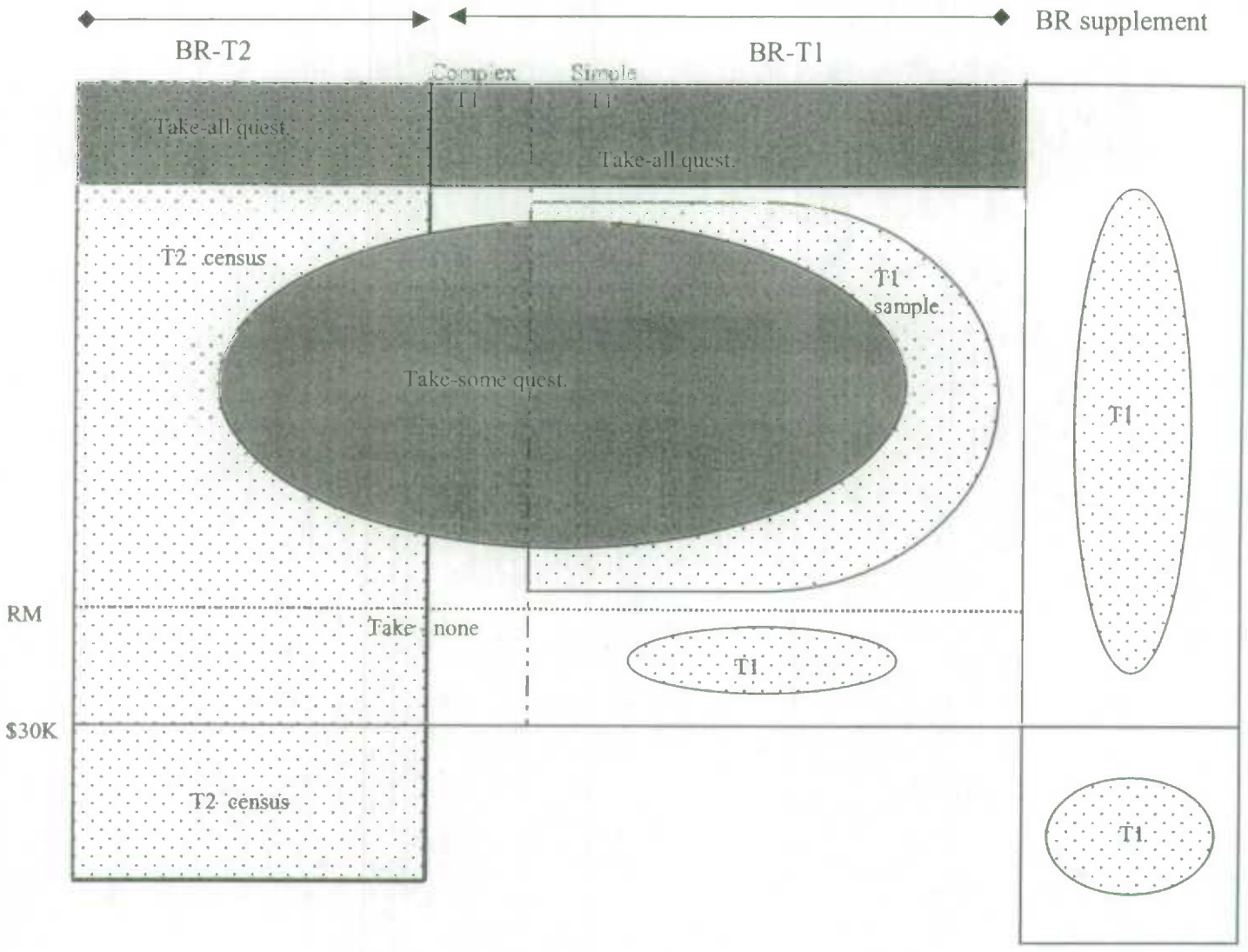
Appendix 2: Target Population, Samples and Exclusion thresholds.

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
TOTAL	n/a	n/a	n/a	n/a	**186276**	**15439**

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
SPECPUB	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
TOTAL	**38957**	**36606**	**489143**	**43972**	**528100**	**80578**

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
SPECPUB	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
TOTAL	15465	13538	483678	40902	**499143**	**54440**

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
TOTAL	**38368**	**35388**	**388543**	**41050**	**426911**	**76438**

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
TOTAL	**14221**	**11792**	**382788**	**37726**	**397009**	**49518**