Proceedings of Statistics Canada Symposium 2001
Achieving Data Quality in a Statistical Agency: A Methodological Perspective

COORDINATING SAMPLES USING THE MICROSTRATA METHODOLOGY

Pascal Rivière[1]

ABSTRACT

The issue of sample co-ordination is crucial for business surveys as it is a way of smoothing out the survey burden. In many co-ordination methodologies, the random numbers characterising the units are permanent and the way of selecting the sample varies. In the microstrata methodology, it is the selection function that is permanent. On the other hand, random numbers are systematically permuted between units for different co-ordination purposes : smoothing out the burden, minimising the overlap between two surveys, or updating panels. Permutations are made in intersections of strata named microstrata. This method has good mathematical properties and gives a general approach for sample co-ordination in which births, deaths and strata changes are automatically handled. There is no particular contraint on stratifications and rotation rates of panels. Two softwares have been written to implement the method and its evolutions: SALOMON in 1998, and MICROSTRAT in 2001.

KEY WORDS: Stratification; Survey sampling ; Co-ordination of samples ; Burden ; Panels ; Vectors of Random Numbers ; Permutations.

1. INTRODUCTION

Sample co-ordination in business surveys has many different goals, as mentioned by Royce (2000). Two main aims can be isolated. The first one is to control the survey burden on businesses, by minimizing the recovery rate between two samples (negative co-ordination), or by smoothing out the cumulative burden. The second one is to update panels, which means controlling the overlap between two consecutive samples of the panel, controlling the time of inclusion of a unit in the panel, and handling for units arriving in the scope or exiting from the scope. It raises a lot of issues (e.g. co-ordination between enterprises samples and establishments samples), and it is impossible to take into account all kinds of constraints simultaneously.

There is a long history of sample co-ordination methodologies in many different countries, for example the Jales method (Atmer, Thulin & Backlund 1975), the EDS methodology (De Ree 1983), the Ocean system (Cotton, Hesse 1992), synchronised sampling (McKenzie, Gross 2000), or Poisson mixture sampling (Teikari 2001). Hesse (1999) gives a detailed presentation of the main methods. The microstrata method, which is presented here, has been developed in 1998 in the framework of a Eurostat's SUPCOM project on sample co-ordination (Rivière 1998). The aim was to have a general software to co-ordinate samples, usable by a non-expert for different kinds of co-ordinations.

We will start by explaining what does « sample co-ordination » mean, and by describing briefly the existing methods. After that we will describe the principle of sequential selection of co-ordinated samples, and the idea of burden-based permutations of random numbers, which will lead to the method of appropriate burden-based permutations in « microstrata ». We will then present the main features of the software that

[1] Pascal Rivière, University of Southampton, Department of Social Statistics, Highfield, Southampton S017 1BJ, UNITED KINGDOM

implemented the method, basically from the user's point of view. We will conclude by drawing up a list of characteristics, advantages and disadvantages of the method.

2. SEQUENTIAL SELECTION OF COORDINATED STRATIFIED SAMPLES

In the following, all samplings will be Simple Stratified Random Samplings Without Replacement (SSRSWOR), which correspond to the majority of actual sampling designs in business surveys. Other sampling designs can be used, which requires other ways of co-ordinating. For example, Ohlsson (2000) analyses the co-ordination of pps samples.

In actual business surveys, it is often difficult to know in advance all the information on future sampling designs. This means that it is impossible to perform a global co-ordination of all samples (for example all the samples of the same year). Therefore the co-ordination can only be done survey by survey. For a given survey, our aim will be to select a sample that is co-ordinated with all the previous samples, and such that the inclusion probabilities are controlled. The problem is now to define how to build co-ordinate samples.

In a sequence of samples $(s^1,...,s^K)$, every sample s^k will be entirely characterised by : a sampling design (which includes a stratification), a vector of random numbers, and a selection function.

Sampling design

A <u>stratification</u> is defined as a partition of the universe: $U = \bigcup_{h=0}^{H_k} U_h^k$, where U_h^k = stratum h of survey k

By way of convention, U_0^k is the subset of units outside the scope, with a sampling rate of 0.

The sampling frame for survey k is then $\bigcup_{h=1}^{H_k} U_h^k$. H_k = number of sampling strata in survey k.

As samplings are SSRSWOR, sampling designs can always be defined as the list of sampling sizes for all strata. For every survey k and for every sampling stratum h, the total number of units is denoted N_h^k , and the sampling design gives the number of units n_h^k to take in the sample in this stratum.

We have : $|U| = N = \sum_{h=0}^{H_k} N_h^k$, $|s^k| = n^k = \sum_{h=0}^{H_k} n_h^k$, where $n_0^k = 0 \quad \forall k$

By definition the inclusion probabilities are: $\forall j \in U_h^k, \ \ \pi_j^k = \left(N_h^k\right)^{-1} n_h^k$

In SSRSWOR sampling design can also be described by sampling rates, but then it is necessary to come up with numbers of units, using rounding techniques (Cox 1987). We will restrict ourselves for purposes of simplification, to cases where the sampling design is defined using numbers of units per stratum.

The k-th sampling design is then characterised by: $Q^k = \{(U_0^k, 0), (U_1^k, n_1^k), ..., (U_{H_k}^k, n_{H_k}^k)\}$

Vectors of random numbers

In our context, a « random number » is a random variable drawn in a uniform distribution on [0 , 1), and associated to a given unit. Each unit has its own random number. A Vector of Random Numbers (VRN) is a vector whose N components are the random numbers of all the units of U. Such a vector is a basic instrument to select samples. The fact that random numbers are i.i.d. is essential if we want to select samples easily and with a perfect control of inclusion probabilities.

For that reason the very first VRN ω^1 is such that all the random numbers are i.i.d. Then, contrary to many other methods, the random numbers will not be « permanent » : before each survey *k*, a new VRN ω^k will be built, deduced from previous VRNs in a way that will be explained in the next chapter. For each new unit (birth, for example), a new random number is drawn (and associated to the unit), independently from previous random numbers.

Therefore, before the *k*-th sample selection, each unit *j* has a random number ω_j^k.

Sample selection

Inclusion of a given unit *j* in the sample s^k can be indicated by a membership indicator I_j^k, whose value is 1 if $j \in s^k$, and 0 otherwise. Given the sampling design, and the VRN, the selection of sample *k* using SSRSWOR technique, requires to build a deterministic function that provides the vector of membership indicators, such that these indicators satisfy the constraints of the sampling design. It also has to be left invariant by any permutation of components within the sampling strata. A simple way to satisfy these constraints consists in selecting, in every stratum *h*, the units with the n_h^k lowest random numbers.

Hence a sequence of sample selections will be defined as a sequence $\left((Q^1,\omega^1),\ldots,(Q^K,\omega^K)\right)$, such that every VRN ω^k is either a function of previous VRNs or a vector independent from all previous VRNs, and such that all sample selections are done by taking the lowest random numbers in each stratum.

3. BURDEN-BASED PERMUTATIONS

Filling-in a questionnaire of a business survey is a long and tedious task. Many enterprises would complain about this unwanted quantity of work, called « survey burden » or « survey load ». For a given set of sample designs, the co-ordination of samples can not reduce the total load (over all units and all surveys), but it can allow to smooth it out. De Ree (1983) proposed to sort the units by increasing cumulated burden just before sample selection. Therefore the less burdened units will be more likely to be selected. We suggest to use such a sorting to permute random numbers between units, which means that the random numbers will not be permanent. As sortings are made by increasing cumulated burden, we can ensure that the less a unit is burdened, the more it is likely to be selected.

Cumulative burden

For each survey, a burden coefficient is defined. It is supposed to represent the time or cost required to fill in the questionnaire. Burden coefficients might vary from a survey to another, as these costs can vary a lot. There can be several burden coefficients for a given survey, but only one for each sampling stratum. These coefficients are not a real estimation of the burden : such an estimation is very difficult, and the burden includes a subjective part that cannot be measured. In practice, burden coefficients are integers.

Let c_j^k be the burden coefficient for unit *j* in survey *k*. Then, by definition, the cumulative burden from survey *m* to survey *K*, named (*m,K*)-cumulated burden is, for unit *j*:

$$CB_j^{m,K} = \sum_{k=m}^{K} c_j^k I_j^k$$

Cost-based permutations and burden-based permutations

A permutation of components of a vector X is said to be cost-based if there exists a vector $C = (c_1,...,c_N)$ such that the components of the new vector X' are ordered according to the cost vector:

$$x'_i < x'_j \Leftrightarrow c_i < c_j \text{ or } (c_i = c_j \text{ and } x_i < x_j)$$

A permutation is said to be burden-based if it is a cost-based permutation in which the cost vector is a vector of cumulative burdens, which is a random vector. Such a permutation ensures that within each partition subset, the random numbers are sorted according to the total burden. Thus the most burdened units have the largest random numbers in each subset.

4. APPROPRIATE BURDEN-BASED PERMUTATIONS IN MICROSTRATA

Principle of microstrata

Now the main question is : where do the sortings take place ? If sorting (and permutating) was done in sampling strata, for example, it would raise an issue. Suppose for example that we selected 2 samples A and B, both stratified using numbers of employees. In A, all businesses with over 20 staff are in the sample, whereas below, the survey rate is 50%. In B, we have a single stratum with businesses between 10 and 49 employees. Suppose now that a permutation based on increasing cumulated burden is made in B sampling strata. Then obviously the units between 20 and 49 staff will be placed at the end of [0 , 1] and we will thus have an overrepresentation of companies with 10 to 19 staff in the next sample. In other words, the probabilities of inclusion will be lower in the 20-49 range. Intuitively, a permutation in 10-49 is inadequate because units in 10-19 range and units in 20-49 range do not have the same « past » in terms of sampling strata.

The first way to solve this problem is to oblige the user to have sampling strata based on an elementary partition: that is what (Van Huis & al 1994) call " building blocks ". The user of this co-ordination system is not free to choose his sampling strata : they have to be combinations of building blocks. This technique is efficient though it raises an issue if we aim to build a general co-ordination system : in such a system, the user should be able to define the stratification he wants, without a priori constraints. Moreover, even if sampling strata are defined by unions of building blocks, it is not enough to take into account of births and strata changes in a general way.

The original idea about "microstrata" comes from this idea to let the user free to choose his stratification, and to have something general enough for births and strata changes. The overall principle is very simple : instead of imposing a priori building blocks that are kinds of intersections of sampling strata, we will try to deduce it from the stratifications of previous samples, whatever those stratifications are. The microstrata are simply intersections of previous sampling strata. If we want to smooth out the cumulative burden for the surveys m to K, the microstrata will be defined as the intersection between the sampling strata of all these surveys. Every microstratum A will then be written as:

$$A = U^m_{h_{m1}} \cap ... \cap U^K_{h_K}, \text{ where } 0 \le h_m \le H_m, ..., 0 \le h_K \le H_K$$

Appropriate burden-based sequences of co-ordinated samples

Given a sequence of samples, a (m,K)-permutation is simply a permutation of ω^K within (m,K)-microstrata, in which sortings are made by increasing (m,K)-cumulated burden. Such a permutation is said to be « appropriate » in the sense that samples used to calculate the cumulated burden are the same as samples used to build the microstratification : cumulated burden and microstratification have the same starting sample *m* and the same ending sample *K*.

Appropriate permutations are the basis of the microstrata methodology. The idea is to perform appropriate burden-based permutations after each sample selection, which enables one to spread the cumulated burden. For every permutation, all burden coefficients of all units could be reallocated if needed. The only constraint is that for a given survey *k*, burden coefficients have to be constant over sampling strata of *k*.

A burden-based sequence of co-ordinated samples is then a sequence of samples $\left((Q^1, \omega^1), ..., (Q^K, \omega^K)\right)$ such that :

- ω^1 is an i.i.d. random vector
- $\forall k, 1 < k \leq K, \exists m_k, q_k, 1 \leq m_k \leq q_k < k$ / ω^k is a (m_k, q_k) -permutation of ω^{q_k}

Main property of burden-based sequences of co-ordinated samples

(Rivière 2001) shows that in a burden-based sequence of samples, if the starting sample is the same in all burden-based permutations (which means that $m_k = m \ \forall k$), then all the random vectors ω^k are i.i.d. Consequently, if we always use appropriate permutations starting from a given « first » survey, then we are sure that all inclusion probabilities of all orders are controlled.

It is important to notice that with such an approach of burden smoothing, co-ordination is made sample by sample (each sample being co-ordinated with the previous ones to minimise the total load), and not on several samples simultaneously. This is a deliberate choice, because in many NSIs, sampling designs of business surveys are not known a long time in advance.

Independence between microstrata and sample selection

A main feature is the fact that sortings are not made at the sample selection step. Before sample selection, and before even knowing the sampling design of the next sample to be drawn, random numbers are permuted in each microstratum to make sure that the bigger the random number, the bigger the cumulative burden. Microstrata are then intersections of sampling strata of previous surveys, but obviously they do not take account of the sampling strata of the sample to be drawn.

Therefore the sample selection itself has absolutely nothing to do with microstrata. Given the vector of random numbers, given the sampling designs, the procedure is totally deterministic : it consists in selecting the lowest random numbers in each stratum. It means that the selection function is always the same, as said previously. This is an essential difference with methodologies based on permanent random numbers (like JALES): in these methods, random numbers are fixed and the selection function varies (for example : sample selection defined by a starting point in [0 , 1) and a direction). In the microstrata methodology, the random numbers always change, but the selection function is fixed.

5. METHODOLOGY USED IN PRACTICE: SEVERAL PERMUTATIONS

The methodology based on microstratifications can have a major drawback : the more the burden is accumulated over a time, the smaller the microstrata are. In the case of tiny microstrata, permutations are not very useful, and it is clear that they are totally inefficient in microstrata that contain only one unit. That is why the permutation technique has to be improved to be used in practice. How can we build a method based on the microstrata principle that avoids the issue of microstrata size ? Actually, the idea is to make several sortings, at several levels. In the following, we will distinguish between global negative co-ordination (aimed at smoothing out the cumulative burden) and specific co-ordination with another sample (either positive or negative).

Global negative co-ordination

In the cumulative burden technique mentioned above, the starting sample of the cumulation plays a major role. To avoid the obstacle presented by the small size of the microstrata, we will carry out several appropriate burden-based permutations, with different starting samples, and therefore different sizes of microstratifications.

Suppose that K samples were selected. The problem is now to build a new VRN ω^{K+1} that will be used to select the K+1-th survey. To do that, the idea is to take the last VRN ω^{K}, and then to perform three (m,K)-permutations, with different values of m :

- the first one in the K-th sampling strata, with sorting by increasing membership indicator to the K-th survey
- the second one in the microstrata starting from the first survey of the current year y, with sorting by increasing cumulated burden since the beginning of the year
- the third one in the microstrata starting from the first survey of the year y-2, with sorting by increasing cumulated burden since the beginning of the year y-2

We can see the advantage of this method – the first sort will allow regular rotation of units over wide strata (as wide as possible in fact since these are the sampling strata). However, this first rotation does not take into account cumulated burdens and is somewhat in the spirit of the renumbering technique in the Ocean system (Cotton, Hesse 1992). The second sort makes it possible to rotate on the basis of cumulated burdens on microstrata which are not yet too small. The third sort is the major one for burden smoothing. It makes it possible to manage a cumulative burden over quite a long period, bearing in mind that, as we have emphasised, there would be little effect in the case of minuscule microstrata.

Having performed the sort, we get a new VRN ω^{K+1}. At this stage we do not necessarily know what will be the sampling design of the $K+1$-th survey. When the $K+1$-th sampling design is known, the sample selection is always the same : in each stratum h, we select the n_h units with the lowest random numbers (n_h given by the sampling design).

Example of microstrata

Let us consider a population with 33 units. Suppose we have 3 surveys: the first 2 have two sampling strata named 1 and 2 (plus the outside scope), the third has 3 (plus the outside scope). Each survey defines a partition of the population. For each survey and each unit, the stratum number is given.

Table 1 : decomposition of the population in microstrata

Survey 1	00	1	11 22	11 222	0	0 11111 2222	1 2	111 22222
Survey 2	00	0	11 11	22 222	0	1 11111 1111	1 1	222 22222
Survey 3	00	1	11 11	11 111	2	2 22222 2222	3 3	333 33333
Microstrata	**1**	**2**	**3 4**	**5 6**	**7**	**8 9 10**	**11 12**	**13 14**

There are 14 (1,3)-microstrata, where we intersect strata of the three surveys (1 to 14). There are 8 (2,3)-microstrata, in which we intersect 2nd and third survey (columns of table 1 : {1}, {2}, {3,4}, {5,6}, {7}, {8,9,10}, {11,12}, {13,14}). And there are 4 (3,3)-microstrata, which are the sampling strata of the third survey : {1}, {2-6}, {7-10}, and {11-14}.

Specific co-ordination

Every sample drawing in our method has to have as an input a stratification, a sampling design and co-ordination characteristics. Global negative co-ordination is one possible way of co-ordinating samples (= one possible co-ordination characteristic), but there are others.

The user of a sample co-ordination system may also want to co-ordinate his survey with another one in a positive or negative manner. This co-ordination can still lead to the performance of sortings in some partition or other. At the end of the sortings, whatever they are, a random number will still be associated to each unit. Then, using these random numbers the sample is always selected in the same manner. For that reason, describing the co-ordination means describing the permutations prior to it.

In the chosen approach the starting point is to admit that it is impossible to meet all the objectives simultaneously : for a given sample, we can not have in the same time a good specific co-ordination and a sample that allows to smooth out the burden. As the sample selection is always the same (given the VRN), the problem is to find an i.i.d. VRN ω^{K+1} that will ensure the specific co-ordination we want.

As we do not want to mix global co-ordination and specific co-ordination, we will not have to perform the first three permutations (see 5.1). Then the idea is simple : to co-ordinate the *K*+1-th sample A with sample B, we will first take the VRN ω^{K_B} *that was used for the sample selection of B.* Then :

- if we want to co-ordinate A and B in a positive way (100% recovery), $\omega_j^{K+1} = \omega_j^{K+1}\ \forall j$
- if we want to co-ordinate A and B negatively, $\omega_j^{K+1} = 1 - \omega_j^{K+1}\ \forall j$

It means that there is no permutation at all. It is easy to see that this approach of specific co-ordination is strictly equivalent to the technique used in JALES. In this method we would have a perfect control of the recovery rate if the definition of the stratification did not change and if units stayed in their initial strata.

Positive co-ordination with rotation rate and panel updating

In such a co-ordination, the aim is to make sure that from a sample to another, a proportion *r* of units will quit the sample, as a proportion 1-*r* will stay. *r* is called the <u>rotation rate</u>, 1-*r* being the <u>recovery rate</u>. We can use a similar approach as in the previous case, where *r* was either 0 or 1.

Hence we start with random numbers used for the selection of sample B. Then if we except births, units that were in sample B have the lowest random numbers in B sampling strata. To make a co-ordination with a recovery rate *r*, the technique will be similar (but not identical) to the previous one, except that the permutation will be done in B sampling strata and not in microstrata.

Let us consider a sampling stratum *h* with N_h units, n_h having been retained in sample B. Then the idea is to take the $r.n_h$ <u>first</u> units and to put them at the end of the interval [0 , 1), which amounts to sliding the random numbers of the $N_h - r.n_h$ remaining units downwards. Because of births and deaths, it is not certain that the first n_h units were all in sample B. On the other hand, n_h is fixed (cf. B sampling design).

Example : 1 stratum, 12 units, 50% sampling rate for B, rotation rate 1/3.

Before permutation :

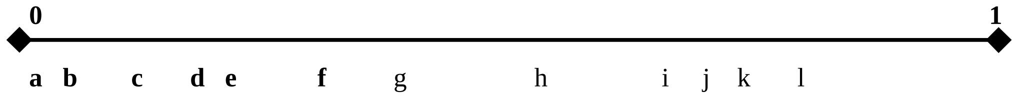

Fig. 1 : positions of the initial random numbers of the units

We can see that in survey B, units a,b,c,d,e,f were in the sample. After permutation, we get :

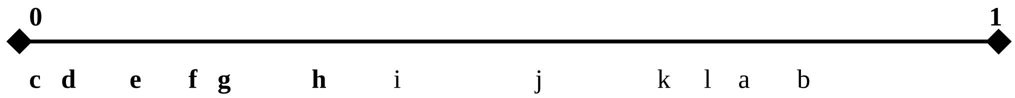

Fig. 2 : positions after permutation

The units that stay in the sample are now c,d,e,f, the new units are still g and h, and a and b leave the sample. If [0 , 1) is seen as a circle and not as a line transect, it is clear that this technique consists in sliding down the units along the circle, and is perfectly adapted to panels. In practice the necessity of rounding makes things a little bit more complicated.

Panels

A panel is a sequence of samples, in which each sample is co-ordinated with the previous one, with a given recovery rate. In panel handling it is also important to control the time of inclusion of units in the panel.

Panels can be seen as a special example of positive co-ordination with given recovery rate. Therefore there is nothing special with panels in the proposed methodology. It implies that the stratification can vary from a sample to another, and that the recovery rate is free. On the other hand, nothing is done explicitly to control the time of inclusion. However, it is controlled implicitly. For example, if the sampling rate is fixed, if the rotation rate is $1/p$, p being an integer, if there are no births, deaths, nor strata changes, and if we neglect the issue of rounding, then it can be shown that the maximum time of inclusion is p.

Handling for births, deaths and strata changes

In many co-ordination methods, births, deaths and strata changes are very difficult to manage. With the microstrata methodology, the generality of the principle ensures that the question of births is automatically handled without doing anything : by convention, in previous surveys, a birth always belonged to a very particular stratum, the « out-of-scope » stratum (in which the sampling rate is 0).

Therefore, suppose that a microstratum contains a birth : as a microstratum is defined as the intersection of previous sampling strata, all the units of this microstratum were in the « out-of-scope » stratum beforehand, which means that all these units are births. The microstrata principle leads to automatic creations of « births microstrata ». The user does not have to care about them: implicitly, a birth microstratum is built. The same goes for strata changes : corresponding microstrata get automatically created. And handling for deaths does not raise any particular issue.

Obviously, births, deaths and strata changes have many negative impacts, in terms of control of burden smoothing, or in terms of control of the overlap (panels or negative co-ordination). But the point here is that they do not require any add-on in the methodology : there is nothing particular to be developed in order to handle those cases. Moreover, the user of the method does not have to care about that.

6. IMPLEMENTATION OF THE METHOD

In 1998, in the framework of the Eurostat SUPCOM project, the SALOMON software was written using the microstrata methodology (Rivière 1998,1999). It had several features : definition of the structure of the sampling frame, sampling frame updating, definition of scope and stratification of a survey, choice of the type of co-ordination (global negative, negative, positive, positive with recovery rate, none), sample selection. This software is available on PC and has been given by Eurostat in 1999 to several National Statistical Institutes that used it, particularly in Eastern Europe. SALOMON, which had been written in less than one year, had some drawbacks and it could be slow because of the huge number of permutations.

Instead of improving this software, INSEE decided to build a brand new one, MICROSTRAT, with nothing in common with SALOMON in terms of computer programs. The idea was to address many issues like specific coordination (in SALOMON, another method of specific coordination was used, that proved less efficient), and to renew completely the way of storing data in order to optimise the permutations. In the following, we will describe the main common features between both tools.

In the system, stratification variables are modified at least once a year : there is a « main annual updating », generally in the beginning of January or at the end of December. After this updating, microstratifications which are the basis of the first two sorts (of global negative co-ordination) are automatically renewed. That is because the " cumulative burden since beginning of year *y* " (or *y*-2) is not the same and does not correspond to the same stratifications, for " year *y* " (or *y*-2) has changed. Other updatings can be done during the year but they have no influence on the definition of microstrata.

The software provides an interface to define in an interactive way the sampling strata and the sampling design. After the definition of scope and stratification, strata identifiers (in practice, integers) are automatically calculated and then allocated to all the units. These strata numbers will be successive positive integers, number 0 by convention representing the stratum " outside the scope ". Definition of the scope by the user should implicitly lead to the allocation of the number 0. The calculation for each unit of its stratum number is made for each sample. The software has then in its database the strata numbers of all units for all surveys since the beginning of year *y*-2. When a new unit is introduced into the population (e.g. a birth), the strata numbers of this unit for the previous surveys are automatically instantiated at 0 (this is automatic and invisible to the user).

Then by default, the coordination mode is "global coordination". Otherwise the user describes which survey he wants to make a positive (or negative) co-ordination with. Then except for the definition of specific (panel or not) co-ordinations, the user does not need to worry about anything as regards co-ordination, and does not even need to know about the existence of microstrata. The construction of microstrata, the sorts carried out, are part of the internal mechanics of the software, its engine, and are of no particular interest for the user. All the necessary information is stored for 3 years. This means, for each unit and each survey: stratum number, presence in the sample or not, random number.

7. CONCLUSION

The proposed system is general and based on four principles :

1) To perform sample co-ordination, instead of acting on the selection function, we can change the vectors of random numbers and always keep the same selection function.
2) To smooth out the total burden, the idea is to permute the components of the vectors of random numbers, using sorting by cumulated burdens in adequate sub-populations named microstrata.
3) Specific co-ordination is done by reusing the random numbers of the survey that we want to co-ordinate with. Specific co-ordination with a recovery rate r ($0<r<1$) requires an additional deterministic permutation.

4) Panels are a particular case of specific positive co-ordination with receovery rate. Time of inclusion is not handled explicitly, but only implicitly.

Then, if we take the user's standpoint, the first main characteristic of a tool implementing the method (SALOMON or MICROSTRAT) is that the user is free to choose the stratifications, the recovery rates (in the case of panels), and the kinds of co-ordination. The second one is that the method does not need any particular treatment for births, changes of stratification variables, units entering or leaving the scope : many microstrata are built but the user does not have to care. But it is important to mention that co-ordination requirements are numerous (burden smoothing, panels, specific co-ordination), and that whatever the tool, it is impossible to meet all the requirements simultaneously : the methodologist has to decide what are the most important ones.

ACKNOWLEDGEMENTS

I would like to thank Chris Skinner, Robert Clark and Patrick Hernandez for many fruitful discussions.

REFERENCES

Atmer, J.G., Thulin G., and Backlund S. (1975). Coordination of samples with the JALES technique, *Statistik Tidskrift*, 13, pp. 443-450.

Cotton F., Hesse C. (1992). Co-ordinated selection of stratified samples. *Proceedings of Statistics Canada Symposium November 92*.

Cox L. H. (1987). A constructive procedure for unbiased controlled rounding. *Journal of the American Statistical Association*, 82, 520-524.

De Ree J. (1983). A system of co-ordinated sampling to spread response burden of enterprises. Prepared by the Netherlands Central Bureau of Statistics (CES/AC.48/43) for the Conference of European Statisticians (United Nations).

Hesse C. (1999) : Sampling co-ordination : a review by country, *INSEE, Working paper E9908*

McKenzie R., Gross B. (2000). Synchronised sampling. *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association.

Ohlsson E. (2000). Co-ordination of pps samples over time, *Proceedings of the Second International Conference on Establishment Surveys*, American Statistical Association.

Rivière, P. (1998) : Description of the chosen method, *deliverable 2 of SUPCOM 1996 project (part « co-ordination of samples »), pp11-33, Eurostat*

Rivière, P. (1999): Co-ordination of samples: the microstrata methodology, *Proceedings of the 13th Roundtable on Business Survey Frames, INSEE, Paris*

Rivière, P. (2001): Random permutations of random vectors as a way of co-ordinating samples, Working paper of the university of Southampton, June 2001

Royce D. (2000). Issues in coordinated sampling at Statistics Canada, *Proceedings of the Second International Conference on Establishment Surveys, American Statistical Association*.

Teikari I. (2001) Poisson mixture sampling in controlling the distribution of response burden in longitudinal and cross-section business surveys. *Tutkimuksia Forskningsrapporter Research Reports 232, Statistics Finland*

Van Huis M., Koeijers E. and De Ree J. (1994). Response burden and co-ordinated sampling for economic surveys. *Netherlands Official Statistics*, volume 9.