Catalogue No. 93-18

INSERTING JOINERS INTO A LONGITUDINAL PANEL OF HOUSEHOLDS AND INDIVIDUALS: SIMULATIONS

February 1993

Bernard Gailly, CEPS/INSTEAD

Pierre Lavallée, Social Survey Methods Division, Statistics Canada

ABSTRACT

Through simulation studies, alternate weighting procedures for panel surveys are compared. Three essential points emerge from the simulations conducted using the PSELL sample (*****Panel Socio-économique *****Liewen zu Lëtzeburg**+**) in 1990.

- The different models tested produce comparable results: the distributions obtained using alternative models are similar to the distributions obtained using the model currently in use.

- The Ernst model has three advantages.
  - the estimator is unbiased;
  - it systematically produces results that conform most closely to the distributions obtained using the traditional model;
  - it rests on a postulate more limited in impact than the postulate underlying the Rao model; joiners increase intra-household variability and add to host household information; and they have specific traits in common with host household members.

  Joiners do not necessarily resemble the entire sample into which they enter and do not represent the target population as a whole at a given point in time. They give a better picture of the development of host households by introducing new information. They have no individual existence independent of the host household.

  Weight regulation occurs within each household. It is not implemented for the sample as a whole (Rao).

- These conclusions are in accordance with the results obtained by Lavallée and Hunter (1992), making them more generally applicable.

Appendix I contains the results of a second comparison based on the 1987 PSELL sample. They support the results obtained using the 1990 sample.

TABLE OF CONTENTS

1. INTRODUCTION

The "Panel Socio-économique *****Liewen zu Lëtzeburg**+**" (PSELL) was established in 1985. The Panel follows the same persons and households from year to year. Like all longitudinal studies, it has one primary objective: collecting information that reflects year by year developments in household standard of living, lifestyle and living conditions.

Annual observation provides successive pictures of the situation in the target population. Each annual sample may be used in the same way as a cross-sectional survey. However, the longitudinality of the survey makes cross-sectional analysis difficult.

The initial sample of PSELL has never been replenished. The sample has therefore undergone progressive distortion: its evolution reflects some aspects of demographic change, but other specific effects as well. More specifically:

- The general demographic changes it reflects include birth, marriage, separation, emigration and death.

- It displays specific effects related to evolution of the sample. These specific effects have two sources: the sample has its own specific demography and its composition is influenced by the repetition of the survey in a number of waves.

These specific effects may be managed in such a way that the sample retains its basic statistical properties, generally via the weighting procedures.

A number of weighting methods exist. The objective of this paper is to evaluate various weighting procedures and propose a solution that improves sample sensitivity to general demographic phenomena.

1.1. General Demographic Change

The sample reflects the demographic evolution in the population because it "represents" the reference population. Events such as birth, death, marriage, divorce, separation and emigration occur in the sample and in the population at large.

However not all these events are taken into account in the same way by the PSELL.

They are taken into account in all descriptions of household characteristics. The arrival of an additional person increases household size. The contribution of a new income has an impact on total household income. The marriage of a child results in establishment of a new household.

But these demographic events are not always reflected in the sample of individuals: for example, only the descendants of the initial sample members are considered legitimate members of the sample. All other joiners are considered illegitimate and excluded from cross-sectional analysis because of the difficulty of assigning them a probability of individual selection without skewing the sample. How does one calculate the probability of initial selection of a young husband who entered the sample in 1989? Although he was a member of the target population in 1985, he was not selected for the initial sample.

1.2. Effects of Sampling

1.2.1. Internal Demographic Effects

The PSELL sample of individuals does not currently reflect all aspects demographic change. "Joiners" reflect changes but are not taken into account in cross-sectional analysis.

This position leads to a considerable loss of information.

In fact, the persons and households forming the sample constantly interact with others in the target population. They sometimes, in conjunction with persons who do not belong to the Panel, form new family units. Or such persons join them.

All these interactions between the sample and the population are observed in the context of the Panel. Most joiners are willing to cooperate with the observers and thus enter the cross-sectional sample.

Their characteristics are taken into account in analysis of household living conditions, but not in analysis of individuals.

This loss of information is detrimental to sample representativeness and longevity.

Some demographic phenomena cannot be taken into account in cross-sectional studies, as long as Panel joiners - although included in the observation - are excluded from cross-sectional analysis.

1.2.2. Effects of Repeated Observation

Annual waves of the survey also has an impact on sample evolution.

Some members of the initial sample refuse to respond after a number of years of participation. Refusal may be temporary and some members agree to participate again following a more or less protracted absence.

An annual cross-sectional weighting procedure can offset the effects of refusals by members of the initial sample, thus preserving the statistical qualities of the sample (unbiased estimators, accuracy of estimates).

However the current procedure ignores the entry of joiners in the sample. It manages internal Panel development without solving the most serious threat to its longevity: inexorably, sample size shrinks as the result of deaths only partially offset by births, emigration not offset by immigration, and refusals not offset by the entry of joiners.

Sample weighting, although it has become a common procedure, entails particular problems with regard to samples covered by a longitudinal study. This paper suggests and compares a number of solutions to this problem.

1.3. Simulations

The weighting procedure implemented to date is summarized in the first part of the paper under the heading "traditional weighting".

A number of weighting models can solve the problem of allocating an effective weight to joiners. The models presented respectively by Ernst (1989) and

J.N.K. Rao (unpublished paper) have been tested and discussed by P. Lavallée and L. Hunter (1992).

The second part of this paper is devoted to replication of these works.

The method used to determine the probabilities of selection for joiners is presented for each model. Variants are also proposed to evaluate each model.

Each weighting model is applied to the sample observed in 1990. The results are compared to the estimates obtained using the "traditional" weighting model, knowing that the latter does not take joiners into account.

Finally, the model we consider the most appropriate is applied to the sample observed in 1987 and the results compared to the estimates obtained using the "traditional" weighting method. This simulation confirms the stability of the effects of the model selected and facilitates observation of some of the effects linked to longitudinal aspects of the study.

2. "TRADITIONAL" SAMPLING AND WEIGHTING

2.1. Initial Sample

The initial PSELL sample is a simple random sample. It represents the persons resident in Luxembourg and registered with the social security system or another social protection system. Each of these persons is a documented wage-earner. The file covers neither all households nor all residents of Luxembourg.

Each person selected leads to an address, which corresponds to a household.

The selection procedure allows for a number of wage-earners to lead to the same address. This means that some households have a greater probability of selection than others: the larger the number of wage-earners in a household, the greater the probability of that household being selected.

To correct this bias, the number of wage-earners present in each household must be identified. As all household members are interviewed, this may be done *a posteriori*.

The initial weight of each household is the reciprocal of the number of wage-earners in that household.

As all household members take part in the survey, the initial weight of household members is equal to the weight of the household.

- The weight of the household is necessarily equal to the mean of the individual weights.

- Each member of the household has the same weight, which gives these weights consistency. However, for reasons to be explained later, this rule will not be respected later on.

It does not appear useful to introduce other corrections to the initial sample, as refusal to respond does not significantly alter sample structure with respect to target population.

2.2. Individual Status within Sample

Beginning with the second wave of the survey, the weights were first calculated at the individual, not the household, level.

The procedure takes into account the fact that individual trajectories may be more or less independent of the initial household.

However, with time it becomes increasingly difficult to identify corresponding households because of the changes produced by entry, withdrawal, marriage and separation.

Before calculating individual weights, the sample must be broken down into 6 sub-groups as all members are not taken into account in the same way.

- Panel members who refuse to respond: these are persons selected in the initial sample (t_0) who refuse to respond in subsequent years (t_{0+n}).

- Panel members present in the cross-sectional sample: these are persons selected in the initial sample (t_0) who are present in the observed sample in subsequent years (t_{0+n}). These persons may be absent at the time of, or refuse to respond to one or more waves. Their individual weight may be recalculated each time they enter the sample.

- New members: these are children born concurrently into the population and the sample (between (t_{0+n}) and ($t_{0+(n-1)}$)). Of these children, only direct descendants of Panel members are considered "legitimate" members. It is sufficient that one of the two ascendants be a Panel member present in the sample.

These new members are assigned a weight equal to the mean weight of their direct ascendants present in the sample at the time they make their entry (different cases are possible).

- Joiners present in the sample: these persons enter the sample at time (t_{0+n}). Most of them could have belonged to the target population at the time the initial sample was selected (t_0). They could have been selected, but were not. As they enter the households in the sample by chance, it is difficult to allocate a probability of selection to them. They are assigned a weight of "0", which means they are never taken into account in cross-sectional analysis of the sample of individuals.

- Joiners who leave the sample: these persons contribute to defining household characteristics when they are present in the sample. When they leave, they move into this category because they are no longer taken into account in the annual calculation of individual response rate and definition of household characteristics.

- Emigrants and the deceased make up the sixth category: these persons reflect general demographic movement. They must not be taken into account in the calculation of response rate.

Only the first two categories are used in calculating individual weights.

2.3. Calculating Weights

Each year, the initial sample observed at (t_0) is broken down into subcategories corresponding to sharply contrasting response rates. These response rates are calculated with respect to the initial sample, not the previous year's sample.

The weight of each member present in the sample is the reciprocal of the response rate of the population subcategory to which that member belongs. The weight obtained at time t_{0+n} is corrected by the initial individual weight at time t_0.

Joiners are assigned a weight of "0".

New members are assigned a weight equal to the mean weight of their direct ascendants (weights of "0" are taken into account).

Each household is assigned a weight equal to the mean weight of its members.

2.4. Implications

The process has a number of implications.

1. Household members may have different weights.

Each member's trajectory that is independent from that followed by other members of the household to which he belongs at a given point in time. It is thus fairly logical to assign different weights to different household members at a given point in time.

There is one disadvantage: not all members of a household have the same weight for estimating the values of one characteristic. Some members "represent" the household to a greater extent than others. Uniform weights would undoubtedly give greater consistency to individual and household estimates taken separately and between both.

2. The weights of household members may be different from the household weight.

Each member follows a trajectory that does not exactly match that of the household. In an individual capacity, he may refuse to respond to the wave, belong to a sample sub-category that is particularly well represented, leave the household and found his own household or, in conjunction with joiners, form a new household.

Consider a Panel member who belongs to a group of persons particularly disinclined to take part in surveys.

He is assigned a very high individual weight because he represents a large number of persons who resemble him.

This person moves to a household whose members do not belong to the initial sample. The weight of this new household is inversely proportional to the number of illegitimate members that it contains. Because these illegitimate members were assigned an individual weight of "0", only the weight of the legitimate Panel member is divided by the total number of persons in the new household.

The greater the number of illegitimate members in a household, the smaller the weight of the household. However this weight has no effect on the individual weight of an legitimate Panel member living in the household, who retains his high individual weight.

The number of illegitimate members in a household influences household weight, but does not change the individual weight of a person who is an legitimate Panel member.

If household weight is reallocated to all household members, the legitimate Panel member receives an individual weight below his initial weight. This operation was done in year one because household members belonged to the target population at the time the sample was drawn and because all participated in the survey.

If the operation is to be repeated in subsequent years, a number of assumptions, the importance of which is explained in the following paragraph, must be made.

3. The departure of an illegitimate member has no effect on the weight of the other members of the household.

When a joiner leaves a host household, the Panel continues to track him. He retains his initial individual weight of "0" and forms a new household, the weight of which remains "0".

The departure of this member usually increases the weight of the host household (all other things being equal) as individual weights are now divided by a figure representing fewer members with weights of "0".

Conversely, the departure has no effect on the individual weights of other household members. This means that the weight of "0" assigned to joiners eliminates the need to make any assumption on whether or not these persons "resemble" other members of the host household or not.

If the departure of an illegitimate member increased the weight of legitimate members belonging to the same household, the other members would "offset" the departure of the illegitimate member. Everything would proceed "as if" the legitimate members of the household were under-represented following the departure of a person who resembles them, at least in some respects.

4. The ratio between household weight and weight of household members remains constant: by definition, household weight is always equal to the mean weight of household members.

The household is a composite entity with variable content: the probability of a household belonging to the sample depends on the probability of its members remaining in the sample and in the household. A household has no independent existence, but evolves with its members. Household weight is simply the logical consequence of the weights of household members.

Sample evolution does not alter this property.

2.5. Disadvantages

This weighting method has one major disadvantage.

It systematically excludes joiners from individual cross-sectional analysis. It hinders adjustment of the initial sample to reflect demographic phenomena displayed by joiners.

A large volume of information remains unexploited.

The weighting procedure is designed to solve only cross-sectional problems of sample representativeness. It doesn't take into account the consequences of longitudinal observation of the sample and makes no provision for moderating the effects of erosion of the sample of individuals.

The following simulations illustrate two possible procedures that may be used to solve this problem.

3. THE "SHARE APPROACH" (ERNST)

3.1. Principles

The approach described by Ernst (1989) and tested by Lavallée and Hunter (1992) entails a sharing of weights among legitimate and illegitimate members ("a share approach" (Lavallée and Hunter)). It allocates a weight other than "0" to joiners so that they may be taken into account in cross-sectional analysis of individual characteristics.

Initially,

- the weight of each legitimate Panel member is the inverse of his selection probability, and
- the weight of joiners is set at "0".

The weight of the household is equal to the mean of the weights of its members.

Next, the household weight is reallocated to each household member.

The latter step, the only thing that distinguishes Ernst's model from the traditional model, has significant consequences.

3.2. Implications

The first two features of the procedure were reported by Lavallée and Hunter.

1. It may be demonstrated that the estimator is unbiased (because the initial weight of joiners is and remains "0") (cf. Appendix 2).

The procedure adopted in year one is repeated annually. The 1985 wage-earners are used as "reference individuals" for calculating household weight, even though the selection probability of other members interviewed is unknown. The household weight is then allocated to all household members. All members are thus selected in each household.

In the following years, legitimate household members are the "reference individuals". As their probability of selection is known, household weight may be calculated. The household weight may then be allocated to all household members, even though the initial selection probability of illegitimate members is unknown. All members are thus observed in each household.

2. All the individuals who belong to a household have the same individual weight.

All the members represent the household in the same way. This uniformity gives consistency between the estimates of individual values and household values.

3. There is no basic contradiction between this feature and the principle of the independence of the individual trajectory with respect to other household members.

Each year the initial weight of each household member reflects his individual trajectory within the Panel: it takes into account status (Panel member, illegitimate member, direct descendant of legitimate Panel member, response rate of sub-sample to which he belongs as an individual).

This weight may be different from the weight of other household members (cf. section 2.4).

When a person leaves a household (establishes a home alone or joins another household), he is following a different trajectory from that of the other members of that household. The weight he carries away is his initial weight, not the weight of the household he leaves. His initial weight and final weight are again independent of the weights of the members of the household he leaves.

If this person establishes a home alone, his initial weight and final weight are equal.

If the leaver founds a household or joins another household, his initial weight remains unchanged. Only his final weight takes the composition of the new household into account.

4. The independence of the individual trajectory with respect to the household is also maintained.

In the traditional model, the individual weight of legitimate members is not influenced by the presence of illegitimate members in the household.

In the Ernst model, the weight of a household is smaller if the household contains a large number of illegitimate members: the weight of the legitimate members must be divided by a larger number of members whose weight is "0". It follows logically that the final weight of legitimate members is reduced: they "inherit" the household weight.

However the final weight of individuals is strictly linked to the establishment (more or less provisional) of the household to which they belong.

This consequence of the model presented by Ernst rests on a postulate stated in the following paragraphs.

The movement from initial individual weight through household weight to final individual weight occurs as if the joiners resembled the legitimate member(s) who make up the household. In some respects, these joiners demonstrate specific traits common to the legitimate members of the host household; in other respects, they differ.

Of course, the illegitimate members carry new information, and weight sharing allows for input of this information. However the weight-share operation also produces effects that are inadmissible without concurrent assumption of partial resemblance between legitimate members and the joiners.

The postulate may be demonstrated as follows.

The persons who represent a population category that is well represented or over-represented within a sample are usually assigned a relatively small weight. This is a basic principle of all weighting procedures.

Weight sharing among legitimate members and joiners reduces the final individual weight of legitimate members. It must be assumed that the presence of these illegitimate members in the household enables the legitimate members to represent a smaller number of persons with whom they share specific common traits.

5. An illegitimate member enters the Panel because he has joined a given household. His existence in the Panel is and continues to be totally dependent on the household in question.

Because it is impossible to calculate initial selection probability for illegitimate members, they are assigned an initial weight of zero.

Their final weight is the weight of the host household. If they leave this household and establish a home alone or join a household made up exclusively of illegitimate members, their initial weight is zero, the household weight is zero and their final weight is zero.

6. The model suggested by Ernst and the traditional model necessarily preclude cross-sectional analysis, and exclude individuals and households made up solely of joiners (whose weight is zero).

To summarize:

1. The originality of the model developed by Ernst rests essentially on movement from initial individual weight through household weight to final individual weight.

2. This characteristic of the model allows for consideration of new information introduced by joiners. At the same time, it assumes that the illegitimate members have some traits in common with the legitimate members of the host household.

3. However the impact of this assumption is weaker than that of the postulate underlying the Rao model.

3.3. Simulation

The traditional model and the Ernst model calculate household weight in the same way. For this reason, the simulation may be limited to comparison of estimated individual values.

Twenty variables were selected. The first eight describe individual characteristics; the last twelve describe characteristics of the household and the head of the household to which the individual belongs. These characteristics of the household and the head of the household are allocated to the individual. They describe the family context in which the individual lives.

Characteristics individual	Characteristics household	Characteristics head of household
sex	number of persons	sex
age	number of children	age
age by sex	number of adults	marital status
marital status	number of jobs	job
nationality	type of dwelling	marital status by sex
adult/child	canton of residence	job by sex
relationship to head of household		
labour force status		

Table 1 compares the estimates of these variables obtained using each weighting method.

Column 1 shows the distributions obtained using the weighting model currently used by PSELL.

It would have been preferable to compare the estimates obtained using each procedure with "real" data observable in the population. However these data are not available. Census data can be used to conduct tests on the sample observed in 1991.

The estimates obtained using traditional weighting are the only reference data available. We acknowledge their reliability, except for the fact that they do no take sufficient account of demographic changes in the reference population. An alternative procedure that takes these changes into account may be preferable, if it can provide unbiased estimates, equivalent (or identical) to the estimates produced by the traditional method.

Column 2 shows the estimates weighted using the Ernst model.

Taking illegitimate members into account (after weighting) increases sample size by 11.6%.

In 1990, 92 persons belonged to households made up exclusively of illegitimate members. Their final weight remains "0". They are automatically excluded from estimation of individual values. The real size of the sample is thus 4641 persons rather than 4735 persons.

Comparison of results essentially demonstrates two things.

1. The disparity between the figures is generally less than 1 percentage point for individual characteristics: taking into account illegitimate members has a slight tendency to lower the weight of the head of the household's children relative to the weight of persons otherwise linked to the head of the household. In other words, illegitimate members are more likely to be adults than children.

2. The characteristics of the household and the head of the household, defined as characteristics of the individual environment, display rigorously identical distributions.

TABLE 1

Comparison of Estimates Using 1990 Weighting Method

(Traditional and Household Weight Sharing)

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Individual Characteristics		
Sex		
1. Male	48.1	48.4
2. Female	51.9	51.6
Age		
1. < 18 years	22.4	21.8
2. 18 - 24 years	10.7	10.9
3. 25 - 34 years	14.8	14.9
4. 35 - 44 years	14.5	14.7
5. 45 - 54 years	13.1	13.2
6. 55 - 64 years	10.8	10.9
7. > 64 years	13.7	13.6
Age by sex		
1. Male < 18 years	11.1	11.1
2. Male 18 - 64 years	32.0	32.4
3. Male > 64 years	5.0	4.9
1. Female < 18 years	11.2	10.8
2. Female 18 - 64 years	32.0	32.1
3. Female > 64 years	8.7	8.7

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Marital Status		
1. Single	40.3	39.7
2. Married	48.3	49.0
3. Widow(er)	8.3	8.2
4. Divorced	2.1	2.1
5. Separated	1.0	1.0
Nationality		
1. Luxemburger	79.7	79.5
2. EEC	18.5	18.7
3. Non-EEC	1.8	1.8
Adult / Child		
1. Adult	74.9	75.4
2. Child	25.1	24.6
Relationship to Head of Household		
1. Head of Household	37.4	37.4
2. Spouse	23.9	24.1
3. Common law partner	0.9	1.1
4. Friend	0.1	0.3
5. Son/Daughter	33.8	32.7
6. Other	3.9	4.4

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Labour Market Status		
1. Child	25.1	24.6
2. Pension, disability	2.9	2.9
3. Pension, retirement	9.2	9.2
4. Pension, survivor	5.7	5.6
5. Work or illness	39.4	39.5
6. Job seeker	0.6	0.6
7. Housewife, Other	17.1	17.6
Household Characteristics		
Number of Persons in Household		
1 person	8.9	8.9
2 persons	20.1	20.1
3 persons	23.6	23.6
4 persons	27.7	27.7
5 persons	12.7	12.7
6 persons	5.0	5.0
7 persons	1.3	1.3
8 persons	0.7	0.7

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Number of Children in Household		
No children	43.8	43.8
1 child	22.3	22.3
2 children	22.4	22.4
3 children	8.5	8.5
4 children	2.7	2.7
5 children	0.3	0.3
Number of Adults in Household		
1 adult	11.4	11.4
2 adults	59.6	59.6
3 adults	17.4	17.4
4 adults	8.2	8.2
5 adults	2.2	2.2
6 adults	0.7	0.7
7 adults	0.5	0.5
8 adults	0.0	0.0

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Number of Persons in Household who have Jobs		
No job	18.2	18.2
1 job	43.1	43.1
2 jobs	28.5	28.5
3 jobs	7.1	7.1
4 jobs	2.5	2.5
5 jobs	0.3	0.3
6 jobs	0.3	0.3
Housing Type		
1. Rural house	7.7	7.7
2. Single family house	34.0	34.0
3. Semi-detached house	10.3	10.3
4. Terraced house	25.3	25.3
5. 2-4 dwellings	13.4	13.3
6. 5-19 dwellings	6.3	6.3
7. More than 20 dwellings	0.7	0.7
8. Makeshift housing	0.1	0.1
9. Retirement home	0.2	0.2
10. n.i.e.	2.0	2.1

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Canton		
1. Luxembourg - urban	15.4	15.4
2. Capellen	7.7	7.7
3. Esch sur Alzette	33.7	33.7
4. Luxembourg - rural	9.1	9.1
5. Mersch	4.2	4.2
6. Clervaux	2.7	2.7
7. Diekirch	5.8	5.8
8. Redange	3.3	3.3
9. Vianden	0.7	0.7
10. Wiltz	3.3	3.3
11. Echternach	3.8	3.8
12. Grevenmacher	6.2	6.2
13. Remich	4.1	4.1
Characteristics of Head of Household		
Sex of Head of Household		
1. Male	85.9	85.9
2. Female	14.1	14.1

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Age of Head of Household		
1. < 24 years	1.7	1.7
2. 25 - 34 years	16.2	16.2
3. 35 - 44 years	26.5	26.5
4. 45 - 54 years	23.8	23.8
5. 55 - 64 years	17.1	17.1
6. > 64 years	14.7	14.7
Marital Status of Head of Household		
1. Single	6.4	6.4
2. Married	79.1	79.1
3. Separated	1.3	1.3
4. Divorced	3.1	3.1
5. Widow(er)	10.1	10.1
Employment of Head of Household		
1. Unemployed	29.4	29.4
2. Employed	70.6	70.6
Marital Status of Head of Household by Sex		
1. Unmarried man	7.3	7.3
2. Married man	78.7	78.7
3. Unmarried woman	13.6	13.6
4. Married woman	0.4	0.4

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4159)	(n=4641)
Employment Status of Head of Household by Sex		
1. Unemployed man	20.3	20.3
2. Employed man	65.6	65.6
3. Unemployed woman	9.1	9.1
4. Employed woman	5.0	5.0

4. "COMPOSITE ESTIMATOR" (RAO)

4.1. Principles

Lavallée and Hunter describe the model proposed by J.N.K. Rao as follows.

- The approach focuses more on the estimator formulation than on the weighting of each of the selected individuals (on the estimator rather than the estimates). It uses a model-based approach to construct an estimator for illegitimate members.

- The model starts with the idea that the legitimate members (selected in 1985) and illegitimate members (who entered the samples in subsequent years) form two separate sub-samples drawn from the same population.

- Each sub-sample gives a population estimate. This implies that the estimates for the illegitimate member sub-group rest on a postulate: the illegitimate

members are treated as if they had been selected using stratified simple random sampling.

- The method allows for:
  - calculation of the selection probabilities of members in both sub-groups,
  - construction of a composite estimator of weights that combines calculation of the weights in each of the two sub-groups,
  - taking into account the relative importance of each sub-group within the total sample.
- The weight of each legitimate Panel member is equal to the inverse selection probability (traditional model, Ernst model).
- The illegitimate members are classified according to a reference variable (e.g. canton of residence). These members are over-represented in some cantons and under-represented in others. The representation rates define the selection probabilities of illegitimate members.

(In 1990, in the absence of data on the target population, the sample of legitimate members was used as reference population).

The illegitimate members are weighted according to the inverse of their selection probability.

- The final weight estimator, applicable to the entire sample, is made up of two weighting factors that correspond respectively to the legitimate member sub-sample and the illegitimate member sub-sample. The weights of both sub-samples are taken into account in proportion to their relative size in the total sample.

- The mean of the weights of household members defines the household weight. This phase is common to all three models.

- The household weight may then be reallocated to each member (Ernst model), although this step is not obligatory. Individual weights may be used as they stand because illegitimate members are assigned an initial individual weight not equal to "0".

Allocation of the household weight to household members has the advantage of giving consistency between the estimates for individuals and for households. All the members of a household have the same weight and represent the household in the same way.

This operation is fundamental to the Ernst model, as it establishes the vector of the weights of illegitimate members. In the Rao model, the operation may be omitted, as illegitimate members are assigned an initial weight independent of household weight.

This difference between the two models stems from the fact that they are based on different assumptions, which have very different implications.

4.2. Implications

The first two properties of this model have been described by Lavallée and Hunter.

1. The composite estimator is biased. The bias comes from the estimator of the weights of illegitimate members. These members do not form a probability sample as they are not drawn at random and the stratification variable is selected arbitrarily. The significance of the bias is directly related to the size of the sub-sample relative to the total sample.

2. All the members of a household may have the same individual weight, as long as the household weight is reallocated to all members. This uniformity of weight restores consistency between estimates of individual values and household values.

3. There is no basic contradiction between this property and the concept of the independence of individual trajectories.

   ! If household weight is not reallocated to the members, each individual (legitimate or not) keeps the initial cross-sectional weight.

   Each year, the initial weight of each household member reflects that person's individual trajectory within the Panel, taking status into account (Panel member, illegitimate member, direct descendant of a legitimate Panel member, annual response rate in the population sub-category to which he belongs as legitimate Panel member, weight of stratum to which he belongs as illegitimate Panel member).

Heterogeneity of the weights of household members increases with diversity of origin and trajectory.

! If household weight is reallocated to each member, the initial cross-sectional weight of a member remains independent of that of other household members.

When a person leaves a given household, establishes a home alone or joins another household, he follows a trajectory different from the trajectories followed by the members of that household. His initial weight and final weight are once again independent of the weights of the members of that household.

If he establishes a home alone, his initial weight and final weight are identical.

If he founds or joins another household, his initial weight is unchanged; his final weight, in his capacity as member of a new household, takes into account the composition of that new household.

4. The independence of the individual trajectory in relation to the household is also respected. When an individual leaves a household, he carries away only his initial individual weight.

Calculation of household weight is done by the same method used in the Ernst model and the traditional model. There is however a significant difference between the Rao and Ernst models.

In the Ernst model, the final individual weight of household members is smaller if the household contains a larger number of illegitimate members, as the weight of the legitimate members is divided by a number representing more members (cf. Section 3.2). Allocation of weights within each household is based on the specific configuration of that household.

This is not true of the Rao model. Legitimate members no longer necessarily represent a smaller number of individuals who resemble them.

Regulating the weights of household members no longer depends on internal household configuration. It is subject to external circumstances, i.e. governed by the variable used to stratify and weight the sub-sample of illegitimate members.

The Ernst postulate no longer applies: although joiners provide a clearer picture of changes that have occurred in host households, there is no assumption that they resemble in any way the members of the household they join.

The impact of the Rao postulate is broader.

The joiners are weighted relative to a target population (any characteristic thereof). It is therefore assumed that they represent this population. At the same time, this target population is the population that the legitimate members are supposed to represent. This is the equivalent of saying that both sub-samples, drawn from a common parent population, resemble one another (given an error factor).

It may be demonstrated that the assumption is too general. The resemblance of the two sub-samples depends on the characteristics selected.

If the two sub-samples resemble one another, they must display parallel distribution of the variables used for comparison.

If they do not, the sample of illegitimate members must over-represent or under-represent some population classes. Two solutions are possible.

- It is assumed that the two sub-samples do not resemble one another, the initial postulate is rejected and illegitimate members are not added to the total sample; or

- the effects of over-representation are corrected by weighting the sub-sample of illegitimate members; its particular properties are nullified and the contribution of the new information specifically aimed at regenerating the initial sample is rejected. (The significance of this type of weighting of a non-probability sample is uncertain).

The two sub-samples do in fact resemble one another in a number of ways: sex, sex of head of household, and canton of residence are the three criteria that give the closest resemblance between the two sub-samples (cf. Table 3).

The three variables with the greatest disparities between the two sub-samples are individual age, age of head of household and relationship to head of household (cf. Table 3). Joiners are over-represented among:

- persons in the 25 - 34 age group,

- common law partners, friends, and other relationships to the head of the household, to the disadvantage of the head of the household's children, and

- members who belong to households in which the head of the household is less than 34 years of age.

The two sub-samples do not resemble one another systematically. The other variables display more or less strong disparities between the two sub-samples. Inclusion of illegitimate members assumes that their sample will be adjusted to offset the effects of areas of over (under)-representation.

Three further points can shed additional light on the effects of this procedure for weighting the sample of illegitimate members.

5. The traditional model and the Ernst model have one point in common: joiners are members of the Panel only because of their membership in a household. Their initial individual weight is zero. When they leave the household, establish a home alone or join a household made up exclusively of illegitimate members, their individual weight is zero, the household weight is zero and their final weight is zero.

The Rao model is different from the other two in three ways.

- The initial weight of illegitimate members is no longer "0". An initial weight is allocated on the basis of status in relation to the stratification variable.

  This implies that illegitimate members have an existence within the Panel independent of the household to which they belong. When they leave the host household and establish a home alone, they still have an initial individual weight linked to their position in relation to the stratification variable for the sample of illegitimate members.

- The joiners receive *a priori* a higher initial weight in the Rao model than in the Ernst model.

  Joiners are already assigned a weight greater than "0" when they are included in the calculation of household weight: households that receive joiners receive *a priori* a larger weight with this model.

- The first two models automatically exclude any illegitimate member belonging to a household made up exclusively of illegitimate members. These members are not included in cross-sectional analysis of individual characteristics.

  The Rao model does not automatically exclude these members, whose initial weight is not "0". This means that they may be taken into account in cross-sectional analysis of individual characteristics, even if they belong to a household made up exclusively of illegitimate members.

4.3. Simulation

Table 2 takes the twenty variables used in the first simulation and compares them to the estimates obtained using the traditional weighting model.

Column 1 gives the distributions obtained with the traditional weighting model currently used by PSELL.

Column 2 gives the distributions obtained with a preliminary version of the Rao composite estimation model. In accordance with the author's suggestions, the stratification variable for the sample of illegitimate members (canton of residence) is by nature extremely general.

The two sub-samples are distributed in similar fashion in the various cantons (cf. Table 3). As a result, the effect of weighting is very limited and the points of disparity between the two sub-samples remain evident.

This variable has the advantage of allowing characteristics of joiners to be taken into account through allocation of an initial weight to illegitimate members. But it is difficult to see the point of a weighting method that has no effect on sample profile (apart from the fact that the advisability of weighting a non-probability sample is unclear).

Column 3 suggests an alternative. The sample of illegitimate members is weighted according to the two variables that show the greatest disparity from one sample to the other: age of head of household and relationship to head of household.

The object of the exercise is clear: adjust the sample of illegitimate members to make it resemble the sample of legitimate members as closely as possible. The resemblance between the two sub-samples is better ensured ... to the disadvantage of the new information contributed by the joiners.

Column 4 also shows distribution after stratification of the sample of illegitimate members according to age of head of household and relationship to head of household. A further precaution ensures a better comparison.

The illegitimate members who have left the host household and are now in a household made up exclusively of illegitimate members are excluded voluntarily, even though the model allocates a weight to them.

Comparison of results essentially demonstrates four points.

1. The different methods produce very similar results.

2. When the sample of illegitimate members is stratified by canton, 23 cases show disparities equal to or greater than 1 point (18 between 1 and 2 points; 5 greater than 2 points). It is probable that these disparities are statistically negligible, although they are more numerous with this model than with the others.

3. When the sample of illegitimate members is stratified by relationship to head of household and age of head of household, 11 cases show disparities equal to or greater than 1 point.

4. When the sample of illegitimate members is stratified by relationship to head of household and age of head of household, excluding voluntarily the members who have no relationship to an legitimate member, 9 cases show disparities equal to or greater than 1 point.

TABLE 2

Comparison of Estimates According to Weighting Method, in 1990

- Traditional
- Composite weight A (stratification by canton)
- Composite weight B (stratification by relationship to head of household and age)
- Composite weight B (excluded: households with no legitimate member)

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Individual Characteristics				
Sex				
1. Male	48.1	48.6	48.6	48.8
2. Female	51.9	51.4	51.4	51.2
Age				
1. < 18 years	22.4	21.7	21.9	21.8
2. 18 - 24 years	10.7	11.9	11.3	11.2
3. 25 - 34 years	14.8	16.9	15.4	15.5
4. 35 - 44 years	14.5	14.3	14.9	14.7
5. 45 - 54 years	13.1	12.4	12.9	13.2
6. 55 - 64 years	10.8	10.2	10.6	10.5
7. > 64 years	13.7	12.6	13.0	13.1
Age by Sex				
1. Male < 18 years	11.1	11.0	11.0	11.1
2. Male 18 - 64 years	32.0	33.1	32.8	32.8
3. Male > 64 years	5.0	4.6	4.8	4.9
1. Female < 18 years	11.2	10.8	10.9	10.7
2. Female 18 - 64 years	32.0	32.5	32.3	32.2
3. Female > 64 years	8.7	8.0	8.2	8.3

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Nationality				
1. Luxemburger	79.7	79.2	79.5	79.7
2. EEC	18.5	19.1	18.9	18.7
3. Non-EEC	1.8	1.7	1.6	1.6
Marital Status				
1. Single	40.3	40.1	40.0	39.8
2. Married	48.3	48.6	48.2	48.8
3. Widow(er)	8.3	7.8	8.1	7.9
4. Divorced	2.1	2.3	2.4	2.3
5. Separated	1.0	1.2	1.3	1.2
Adult / Child				
1. Adult	74.9	75.8	75.5	75.6
2. Child	25.1	24.2	24.5	24.4
Relationship to Head of Household				
1. Head of household	37.4	37.0	36.8	36.6
2. Spouse	23.9	23.8	23.7	23.9
3. Common law partner	0.9	1.7	1.4	1.4
4. Friend	0.1	0.5	0.4	0.3
5. Son/ Daughter	33.8	31.5	32.1	32.2
6. Other	3.9	5.5	5.6	5.6

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Labour Market Status				
1. Child	25.1	24.2	24.5	24.4
2. Pension, disability	2.9	2.8	2.9	2.8
3. Pension, retirement	9.2	8.6	8.8	9.0
4. Pension, survivor	5.7	5.3	5.6	5.5
5. Work or illness	39.4	41.6	40.6	40.4
6. Job seeker	0.6	0.7	0.7	0.7
7. Housewife, Other	17.1	16.9	16.9	17.2
Household Characteristics				
Number of Persons in Household				
1 person	8.9	8.1	8.3	8.0
2 persons	20.1	21.0	19.7	20.0
3 persons	23.6	24.2	24.2	23.9
4 persons	27.7	26.8	27.2	27.3
5 persons	12.7	12.1	12.2	12.2
6 persons	5.0	5.4	5.7	5.8
7 persons	1.3	1.7	2.0	2.1
8 persons	0.7	0.7	0.7	0.7

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Number of Children in Household				
No children	43.8	43.9	43.4	43.5
1 child	22.3	23.6	23.4	23.4
2 children	22.4	21.2	21.6	21.4
3 children	8.5	8.2	8.4	8.4
4 children	2.7	2.6	2.5	2.6
5 children	0.3	0.5	0.7	0.7
Number of Adults in Household				
1 adult	11.4	10.4	10.7	10.3
2 adults	59.6	60.2	58.7	58.8
3 adults	17.4	17.2	18.2	18.2
4 adults	8.2	8.3	8.1	8.3
5 adults	2.2	2.5	2.8	2.9
6 adults	0.7	0.9	1.0	1.0
7 adults	0.5	0.4	0.5	0.4
8 adults	0.0	0.1	0.0	0.1

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Number of Persons in Household who have Jobs				
No job	18.2	16.5	17.0	16.9
1 job	43.1	42.2	42.6	43.0
2 jobs	28.5	30.8	29.6	29.2
3 jobs	7.1	7.1	7.2	7.1
4 jobs	2.5	2.8	3.1	3.2
5 jobs	0.3	0.3	0.3	0.3
6 jobs	0.3	0.3	0.2	0.3
Housing Type				
1. Rural house	7.7	7.4	7.3	7.5
2. Single family house	34.0	33.5	33.5	33.6
3. Semi-detached house	10.3	10.2	10.3	10.5
4. Terraced house	25.3	25.4	25.8	25.6
5. 2 - 4 dwellings	13.4	14.1	13.9	13.8
6. 5 - 19 dwellings	6.3	6.5	6.2	6.1
7. more than 20 dwellings	0.7	0.7	0.7	0.7
8. Makeshift housing	0.1	0.1	0.1	0.1
9. Retirement home	0.2	0.1	0.1	0.1
10. n.i.e.	2.0	2.0	2.0	2.0

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Canton				
1. Luxembourg - urban	15.4	15.4	15.5	15.4
2. Capellen	7.7	7.7	7.6	7.4
3. Esch sur Alzette	33.7	33.8	34.7	34.4
4. Luxembourg - rural	9.1	9.0	8.7	8.8
5. Mersch	4.2	4.2	4.3	4.4
6. Clervaux	2.7	2.8	2.9	3.0
7. Diekirch	5.8	5.8	5.6	5.8
8. Redange	3.3	3.3	3.2	3.0
9. Vianden	0.7	0.7	0.7	0.7
10. Wiltz	3.3	3.3	3.2	3.3
11. Echternach	3.8	3.8	3.9	3.8
12. Grevenmacher	6.2	6.1	5.8	6.0
13. Remich	4.1	4.1	3.8	4.0
Characteristics of Head of Household				
Sex of Head of Household				
1. Male	85.9	85.9	85.9	86.4
2. Female	14.1	14.1	14.1	13.6

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Age of Head of Household				
1. < 24 years	1.7	2.8	2.0	2.0
2. 25 - 34 years	16.2	19.2	16.7	16.6
3. 35 - 44 years	26.5	25.9	27.5	26.9
4. 45 - 54 years	23.8	22.5	23.4	24.0
5. 55 - 64 years	17.1	16.2	16.6	16.6
6. > 64 years	14.7	13.4	13.8	13.9
Marital Status of Head of Household				
1. Single	6.4	7.5	6.9	6.7
2. Married	79.1	77.7	77.9	78.8
3. Separated	1.3	1.6	1.5	1.4
4. Divorced	3.1	3.5	3.7	3.3
5. Widow(er)	10.1	9.8	10.0	9.8
Employment of Head of Household				
1. Unemployed	29.4	27.8	28.5	28.6
2. Employed	70.6	72.2	71.5	71.4

	Traditional Weighting Members	Composite Weight A (strati-fication by canton)	Composite weight B (str.: relat to HH age of HH)	Composite weight B (excluded: illegitimate households)
	Frequency (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)
Marital Status of Head of Household by Sex				
1. Unmarried man	7.3	8.7	8.4	8.1
2. Married man	78.7	77.2	77.5	78.3
3. Unmarried woman	13.6	13.7	13.7	13.2
4. Married woman	0.4	0.4	0.4	0.4
Employment of Head of Household by Sex				
1. Unemployed man	20.3	19.1	19.7	19.8
2. Employed man	65.6	66.8	66.2	66.6
3. Unemployed woman	9.1	8.7	8.9	8.8
4. Employed woman	5.0	5.4	5.3	4.8

TABLE 3

COMPARISON OF SUB-SAMPLES USING SIX VARIABLES

Comparison of Samples by Canton of Residence

Canton	Samples	
	Legitimate members	Illegitimate members
Luxembourg - urban	15.4%	13.4%
Capellen	7.7%	5.2%
Esch/Alzette	33.7%	42.5%
Luxembourg. rural	9.1%	6.4%
Mersch	4.2%	4.3%
Clervaux	2.7%	4.0%
Diekirch	5.8%	4.9%
Redange	3.3%	2.3%
Vianden	0.7%	1.0%
Wiltz	3.3%	4.0%
Echternach	3.8%	3.5%
Grevenmacher	6.1%	4.3%
Remich	4.1%	4.2%
Total	100.0%	100.0%
N=	4159	576

Comparison of Samples by Sex of Members (frequency in %)

Sex	Sample	
	Legitimate members	Illegitimate members
Male	48.1%	50.9%
Female	51.9%	49.1%
Total	100.0%	100.0%
N=	4159	576

Comparison of Samples by Sex of Head of Household (frequency in %)

Sex of Head of Household	Sample	
	Legitimate members	Illegitimate members
Male	85.9%	85.6%
Female	14.1%	14.4%
Total	100.0%	100.0%
N=	4159	576

Comparison of Samples by Age Category
(frequency in %)

Age Category	Sample	
	Legitimate members	Illegitimate members
< 18 years	22.4%	16.8%
18 - 24 years	10.7%	21.9%
25 - 34 years	14.8%	35.1%
35 - 44 years	14.5%	10.2%
45 - 54 years	13.1%	5.9%
55 - 64 years	10.8%	4.2%
> 64 years	13.7%	5.9%
Total	100.0%	100.0%
N=	4159	576

Comparison of Samples by Relationship to Head of Household
(frequency in %)

Relationship to Head of Household	Sample	
	Legitimate members	Illegitimate members
Head of household	37.4%	32.1%
Spouse	23.9%	21.7%
Common law partners	0.9%	7.6%
Friend	0.1%	3.6%
Son/Daughter	33.8%	16.1%
Other	3.9%	18.9%
Total	100.0%	100.0%
N=	4159	576

Comparison of Samples by Age of Head of Household
(frequency in %)

Age of Head of Household	Sample	
	Legitimate members	Illegitimate members
< 24 years	1.6%	11.1%
25 - 34 years	16.2%	40.3%
35 - 44 years	26.5%	21.7%
45 - 54 years	23.9%	12.7%
55 - 64 years	17.1%	9.5%
> 64 years	14.7%	4.7%
Total	100.0%	100.0%
N=	4159	576

5. CONCLUSION

Comparison of results demonstrate three essential points.

1. The different models give very comparable results: disparities in the distributions obtained with each model and the model in current use are very small.

2. The Ernst model has three advantages:

   - the weight estimator is unbiased;

   - it systematically produces results most similar to the distributions obtained using the traditional model;

- ! the postulate on which it rests has a lesser impact than the postulate underlying the Rao model: although joiners give a clearer picture of changes that occur in host households, in the Ernst model, this addition presupposes the further assumption that these joiners share some traits in common with the members of the host household.

They do not necessarily resemble the entire sample into which they enter and they are not representative of the target population as a whole at a given point in time; but they do express changes that occur in host households.

They have no individual existence independent of the host households.

Regulation of weights is conducted within each household. It is not conducted for the sample as a whole (Rao).

3. The fact that these conclusions are in strict compliance with the results obtained by Lavallée and Hunter (1992) makes them more generally applicable.

Appendix 1 gives the results of a comparison made using the sample observed in 1987 (PSELL). The results support these conclusions and the results obtained using the sample observed in 1990.

BIBLIOGRAPHY

Ernst, L. (1989), "Weighting Issues for Longitudinal Household and Family Estimates" in Panel Surveys, John Wiley and Sons, New York, 1989.

Gailly B. (1992), PSELL : "Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1989", Document PSELL No. 48, CEPS/Instead, Walferdange, 1992.

Lavallée P., Hunter L. (1992) "Weighting for the Survey of Labour and Income Dynamics", contribution to Statistics Canada Symposium 92 "Design and Analysis of Longitudinal Surveys", Ottawa, 1992 (forthcoming).

Rao J.N.K., unpublished paper.

APPENDIX 1

Comparison of Estimates Using 1987 Weighting Method

(Traditional and Share Approach)

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Individual Characteristics		
Sex		
1. Male	48.2	48.2
2. Female	51.8	51.8
Age		
1. < 18 years	23.6	23.3
2. 18 - 24 years	11.1	11.3
3. 25 - 34 years	15.2	15.6
4. 35 - 44 years	14.6	14.6
5. 45 - 54 years	12.6	12.4
6. 55 - 64 years	10.5	10.4
7. > 64 years	12.4	12.4

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Age by Sex		
1. Male < 18 years	12.0	11.8
2. Male 18 - 64 years	31.5	31.6
3. Male > 64 years	4.7	4.7
1. Female < 18 years	11.6	11.5
2. Female 18 - 64 years	32.5	32.7
3. Female > 64 years	7.7	7.7
Marital Status		
1. Single	40.0	40.0
2. Married	48.8	49.0
3. Widow(er)	8.1	8.0
4. Divorced	2.3	2.2
5. Separated	0.8	0.8
Nationality		
1. Luxemburger	77.8	77.7
2. EEC	20.3	20.4
3. Non-EEC	1.9	1.9
Adult / Child		
1. Adult	74.7	75.0
2. Child	25.3	25.0

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Relationship to Head of Household		
1. Head of household	36.6	36.4
2. Spouse	23.9	24.1
3. Common law partner	0.9	0.9
4. Friend	0.2	0.2
5. Son/Daughter	34.2	33.6
6. Other	4.2	4.8
Labour Market Status		
1. Child	25.3	25.0
2. Pension, disability	3.0	3.0
3. Pension, retirement	7.8	7.7
4. Pension, survivor	5.5	5.4
5. Work or illness	39.7	39.9
6. Job seeker	0.8	1.0
7. Housewife, Other	17.9	18.0

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Household Characteristics		
Number of persons in Household		
1 person	8.0	8.0
2 persons	19.6	19.6
3 persons	22.8	22.8
4 persons	28.7	28.7
5 persons	12.9	12.9
6 persons	4.9	4.9
7 persons	2.1	2.1
8 persons	0.7	0.7
9 persons and over	0.3	0.3
Number of Children in Household		
No children	41.7	41.7
1 child	23.6	23.6
2 children	22.8	22.8
3 children	9.1	9.1
4 children	2.7	2.7
5 children	0.1	0.1

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Number of Adults in Household		
1 adult	10.4	10.4
2 adults	58.4	58.4
3 adults	18.9	18.9
4 adults	7.8	7.8
5 adults	2.8	2.8
6 adults	1.4	1.4
7 adults	0.2	0.2
9 adults	0.1	0.1
Number of Persons in Household who have Jobs		
No job	17.4	17.4
1 job	42.2	42.2
2 jobs	30.0	30.0
3 jobs	6.9	6.9
4 jobs	2.5	2.5
5 jobs	0.9	0.9
6 jobs	0.1	0.1

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Housing Type		
1. Rural house	9.3	9.3
2. Single family house	34.6	34.6
3. Semi-detached house	9.1	9.1
4. Terraced house	25.3	25.3
5. 2 - 4 dwellings	13.1	13.1
6. 5 - 19 dwellings	5.7	5.7
7. more than 20 dwellings	1.1	1.1
8. Makeshift housing	0.1	0.1
9. n.i.e.	1.7	1.7

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Canton		
1. Luxembourg - urban	17.7	17.7
2. Capellen	6.9	6.9
3. Esch sur Alzette	34.3	34.3
4. Luxembourg - rural	8.8	8.8
5. Mersch	3.7	3.7
6. Clervaux	2.5	2.5
7. Diekirch	6.7	6.7
8. Redange	2.8	2.8
9. Vianden	0.8	0.8
10. Wiltz	3.0	3.0
11. Echternach	3.5	3.5
12. Grevenmacher	5.7	5.7
13. Remich	3.6	3.6
Characteristics of Head of Household		
Sex of Head of Household		
1. Male	86.3	86.3
2. Female	13.7	13.7

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Age of Head of Household		
1. < 24 years	1.8	1.8
2. 25 - 34 years	18.7	18.7
3. 35 - 44 years	27.3	27.3
4. 45 - 54 years	23.0	23.0
5. 55 - 64 years	15.8	15.8
6. > 64 years	13.4	13.4
Marital Status of Head of Household		
1. Single	6.5	6.5
2. Married	79.6	79.6
3. Separated	1.2	1.2
4. Divorced	2.7	2.7
5. Widow(er)	10.0	10.0
Employment of Head of Household		
1. Unemployed	26.7	26.7
2. Employed	73.3	73.3
Marital Status of Head of Household by Sex		
1. Unmarried man	7.0	7.0
2. Married man	79.3	79.3
3. Unmarried woman	13.4	13.4
4. Married woman	0.3	0.3

	Traditional Weighting Members	Mean Household Weight M./non-M.
	Frequency (%)	
	(n=4648)	(n=4914)
Employment of Head of Household by Sex		
1. Unemployed man	17.6	17.6
2. Employed man	68.7	68.7
3. Unemployed woman	9.1	9.1
4. Employed woman	4.6	4.6

APPENDIX 2

The Ernst Procedure is Unbiased

The validity of this statement has been demonstrated by Ernst (1992, pp. 143 & c.). Another version may be found in Lavallée and Hunter (1992).

Generally speaking, any random variable X that characterizes a population of size

$$X' \sum_{i=1}^{N} x_i$$

N may be expressed as
where x_i is the value (x) of this variable for any i^{th} unit of the population.

A traditional survey estimates variable X by selecting a probability sample such that any individual i in population N has a known positive probability (p_i) of being selected in the sample (probability of selection).

There are two possible cases.

1. All individuals have the same probability of being selected. In this case, the estimate of X may be written

$$X) ' \frac{1}{p_i} \sum_{i=1}^{n} x_i$$

where X' is the estimate of X based on values (x) observed in the set of individuals (i) who belong to a sample made up of n individuals. Each individual is assigned a weight that is the reciprocal of its selection probability ($1/p_i$).

A strict equivalent may be written

$$X' \sum_{i=1}^{N} w_i x_i$$

with $w_i = 1/p_i$ for any individual who belongs to the sample and $w_i = 0$ for any individual who does not belong to the sample, w_i being the weight allocated to each individual. This weight (w_i) is equal to $1/p_i$ if all the individuals have the same probability of selection. The weight is equal to zero for any individual who does not belong to the sample.

2. All individuals do not have the same probability of selection (p_i). These different probabilities are known. The estimate of X may be written in the same manner, i.e.

$$X' \sum_{i=1}^{N} w_i x_i$$

However, w_i (the weight of the individuals) must be calculated: $w_i = 1/p_i$, the weight of individual i is the reciprocal of his selection probability if he belongs to the sample and 0 if he does not belong to the sample.

The following problem arises: is it possible to insert into the Panel sample individuals who were not selected at the time of initial selection of the Panel sample? These individuals entered a household at some time during the Panel and the probability of their entry is unknown.

Is there not a risk of skewing the estimate of X by allocating a positive weight to these individuals?

In fact, the estimate of X remains unbiased, if the mathematical expectation of the weight estimator is equal to 1:

$$E(X^{)}) = X \qquad (1)$$

$$\text{if } E(w_i) = 1 \qquad (2)$$

the mathematical expectation of X' is an unbiased estimator of X (1) if the mathematical expectation of w_i is an unbiased estimator of 1, a condition that is met if the mathematical expectation of weight estimator (w_i) is equal to 1 (2).

Defining an unbiased household weighting procedure is thus the equivalent of defining a random variable W_i capable of satisfying the condition in equation (2) (Ernst, 1992, p. 144).

The mathematical expectation of any random variable Y is given by

$$E(Y) = \sum_{y \in S} y \times P(Y = y)$$

The mathematical expectation of variable Y is equal to the sum of the products obtained by multiplying each value of y by the probability of that value being realized (i.e. probability P that Y is equal to y, taking into account that y may have any value).

As a result, the mathematical expectation of the weight estimator is given by the same equation, i.e.

$$E(W_{mi}) = \sum_{w \in S} w \times P(W_{mi} = w) \qquad (3)$$

The mathematical expectation of the estimator of weights (W) for individuals (i) belonging to households ($_{mi}$), i.e. $E(w_{mi})$, corresponds to a sum of products: the products of the possible values of the weights of individuals who belong to households (w) by the probabilities of obtaining these values $P(W_{mi}=w)$.

In this case, W_{mi} (the weights of individuals who belong to the households) may display two initial values as we wish to take into account both legitimate and illegitimate members:

1. let $W_{mi} = 1/p_{mi}$ for the legitimate members whose probability of selection (p_{mi}) is known, because they belong to a household (m) selected for the initial sample,

2. let $W_{mi} = 0$ for the illegitimate members whose probability of selection cannot be calculated, because they did not belong to a household selected for the initial sample, although they probably did belong to the target population at the time that sample was drawn.

By applying equation (3) :

$$E(W_{mi}) = \frac{1}{p_{mi}} \times P(W_{mi} = 1/p_{mi}) + 0 \times P(W_{mi} = 0)$$

$E(W_{mi}) = (1/p_{mi} \text{ X } P(W_{mi}=1/p_{mi})) + (0 \text{ X } P(W_{mi}=0))$,

where $0 \text{ X } P(W_{mi}=0)$ is necessarily equal to 0 and $P(W_{mi}=1/p_{mi})$ is the probability of the initial weight having the value $1/p_{mi}$; this corresponds exactly to the probability of the individual being selected for the sample. Thus:

$E(W_{mi}) = (1/p_{mi} \text{ X } p_{mi}))+0 = 1$. *Quod erat demonstrandum*.

Note that it is not necessary to know the value of the probabilities for illegitimate members: it is simply nullified.