

PANEL SOCIO-ECONOMIQUE

"LIEWEN ZU LËTZEBUERG"

DOCUMENT PSELL N° 117

DECEMBRE 1998

**REPRESENTATIVITE ET PONDERATION
DES ECHANTILLONS DU PSELL 2
1994-1997**

par

Dr Bernard GAILLY

**CEPS/Instead
Differdange
Grand-Duché de Luxembourg**

1998

Présentation du programme P S E L L . 2

Les informations présentées dans ce cahier proviennent du programme PSELL.2 développé par la Division "Population et Ménages" du C.E.P.S./Instead. Le Grand-Duché de Luxembourg dispose d'un instrument exceptionnel permettant de connaître les conditions d'existence des personnes et des ménages qui y vivent depuis 1985 : le panel socio-économique "Liewen zu Lëtzebuerg" (PSELL).

Dans le cadre de ce programme, de nombreuses informations sont récoltées chaque année sur les principaux aspects de la vie de la population du pays :

- conditions de logement, équipement et composition des ménages
- principales dépenses
- précarité
- endettement
- position scolaire des enfants
- position socioprofessionnelle des adultes
- revenus, ...

En 1994, cette étude a fêté son dixième anniversaire. Sur le plan scientifique, cet événement représentait certainement un succès parce qu'il est très rare qu'un même programme de recherche puisse être développé sur une période aussi longue. Une large part de ce succès revient toutefois aux milliers de personnes qui, au fil des années, ont accepté de recevoir chez elles nos enquêteurs et de participer à ce vaste programme ; par leur contribution, elles ont permis de réunir un capital de connaissances inestimable, couvrant dix ans de la vie de la population de notre pays.

Les données récoltées ont déjà fait l'objet de nombreuses études publiées pour la plupart au CEPS/Instead dans les séries suivantes :

- ☞ Documents PSELL (voir liste en annexe)
- ☞ Notes de Recherche
- ☞ PSELL INFO
- ☞ ECOCEPS.
- ☞ Population & Emploi - Série "Conditions de vie"

A partir de 1994 l'échantillon de l'étude a été rénové. Il compte désormais 8232 personnes réparties dans 2978 ménages (avant pondération). Cet échantillon évolue comme la population du pays. Il prend en compte les naissances, l'immigration, les mariages, les décès et l'émigration.

Pour plus d'informations

(I. BOUVY)

Tel: (00 352) 58 58 55- 513

Fax: (00 352) 58 55 60

Document produit par le

CEPS/Instead

*Centre d'études de populations, de pauvreté et de politiques socio-économiques
B.P. 48 - L 4501 Differdange*

Président : Gaston Schaber

SOMMAIRE

| | | |
|---------------------|--|----|
| ATTENTION | | 4 |
| CHAPITRE I | QU'EST-CE QU'UNE POPULATION ET QU'EST-CE QU'UN ECHANTILLON ? | 5 |
| CHAPITRE II | QU'EST-CE QU'UN ECHANTILLON REPRESENTATIF ? | 11 |
| CHAPITRE III | QU'EST-CE QU'UN ECHANTILLON LONGITUDINAL ET QU'EST-CE QU'UN ECHANTILLON TRANSVERSAL ? | 15 |
| 1. | Un échantillon longitudinal | 17 |
| 2. | Des échantillons transversaux | 17 |
| 3. | Unités longitudinales et unités transversales | 18 |
| CHAPITRE IV | QU'EST-CE QU'UN TAUX DE REPONSE CORRECTIF ? | 21 |
| 1. | Taux de réponse correctif transversal | 23 |
| 2. | Taux de réponse correctif longitudinal | 24 |
| CHAPITRE V | A QUOI SERT-IL DE "PONDERER" LES UNITES D'OBSERVATION D'UN ECHANTILLON | 25 |
| 1. | Représentativité et pondération | 27 |
| 2. | Pondérer l'échantillon initial | 27 |
| CHAPITRE VI | FAUT-IL REPONDERER L'ECHANTILLON LONGITUDINAL ET COMMENT | 31 |
| 1. | Indication de repondération | 33 |
| 2. | Le taux de réponse correctif longitudinal | 34 |
| 3. | La non-réponse est-elle ignorable ? | 35 |
| 4. | Les poids longitudinaux ajustés | 37 |
| CHAPITRE VII | SAUVEGARDER LA REPRESENTATIVITE ANNUELLE DES ECHANTILLONS TRANSVERSAUX | 39 |
| 1. | Concernant la représentativité | 41 |
| 2. | Concernant la repondération des échantillons transversaux | 42 |
| | 2.1. Repondération de l'ensemble des personnes qui forment les ménages | 42 |
| | 2.2. Pondération des naissances : les nouveau-nés | 44 |
| | 2.3. Pondération des naissances : les nouveaux immigrants | 45 |
| 3. | Profils des échantillons transversaux pondérés | 47 |

ATTENTION !

Les utilisateurs du PSELL.2 sont invités à lire attentivement les indications qui suivent surtout s'ils ont déjà utilisé les données du PSELL.1 en recourant au système de pondération. Un certain nombre de modifications ont été apportées à la technique de pondération et les résultats de ces modifications peuvent paraître troublants si l'on saisi mal leurs origines : certains concepts de base ont été soit clarifiés soit modifiés et certains dispositifs purement techniques ont été mieux adaptés à la nature de l'évolution des échantillons.

1. Qu'est-ce qu'une population et qu'est ce qu'un échantillon ?
2. Qu'est-ce qu'un échantillon représentatif ?
3. Qu'est ce qu'un échantillon longitudinal et qu'est-ce qu'un échantillon transversal ?
4. Qu'est-ce qu'un taux de réponse correctif ?
5. A quoi sert-il de « pondérer » les unités d'observation d'un échantillon ?
6. Faut-il repondérer l'échantillon longitudinal et comment ?
7. Comment sauvegarder la représentativité annuelle des échantillons transversaux d'un panel ?

Nous n'envisagerons pas les aspects trop techniques de ces différentes procédures. Nous essayerons simplement de montrer qu'il n'est pas vain d'apporter le plus de soin possible à ces différents aspects de la gestion des échantillons générés par un panel.

CHAPITRE I

**QU'EST-CE QU'UNE POPULATION
ET QU'EST-CE QU'UN ECHANTILLON ?**

I. QU'EST-CE QU'UNE POPULATION ET QU'EST-CE QU'UN ECHANTILLON ?

La population est un univers auquel on souhaite étendre l'ensemble des conclusions d'une étude. Ces conclusions ne pourront être étendues au-delà de cet univers.

La population visée par le PSELL se limite aux personnes liées directement ou indirectement au système de sécurité et de protection sociale luxembourgeois.

Par exemple, les conclusions des études relatives aux 'revenus' ou à la 'pauvreté' ne concernent en aucun cas 'l'ensemble des personnes qui résident au Luxembourg'. Les fonctionnaires étrangers, les agents de sociétés étrangères installés provisoirement au Luxembourg et liés au système de sécurité sociale de leur pays ainsi que les Luxembourgeois (frontaliers) travaillant en dehors du pays ne sont pas visés par cette étude.

L'échantillon (S) extrait de cette population (U) est formé par l'ensemble de toutes les unités sélectionnées avant toute observation et quel que soit le statut de ces unités après l'observation.

S est composé de m personnes appartenant à U composé de M personnes. Il se présente généralement sous la forme suivante après l'observation. Pour m adresses distribuées, s comprend :

- **m1** 'hors champ'

dont :

- décès
- émigrés
- en ménages collectifs
- autres erreurs liées à la nature de la population

- **m2** 'enquêtes réalisées'

- **m3** 'échecs'

dont :

- refus
- adresses inexactes ou inconnues
- impossible à réaliser (maladie, handicap, (...))
- impossible à contacter (absences, (...))
- autres impossibilités concernant des personnes dans le champ.

L'échantillon est composé de toutes les unités (m) sélectionnées. Soit, les enquêtes réalisées, les hors champ et les échecs.

Les répondants (soit $m_2 + m_1$) comprennent non seulement les enquêtes réalisées (m_2) mais aussi les 'hors champ' (m_1).

Les 'hors champ', soit m_1 , étaient bien recensés dans la population au moment où l'échantillon a été tiré¹. Ils ont donc influencé le tirage de l'échantillon. Ce sont des répondants pour lesquels toutes les variables mesurées prendront une valeur '0' ou 'non-réponse'. Nous verrons par la suite qu'il n'y a pas lieu de les écarter ni de la définition, ni de la composition de l'échantillon.

Le premier échantillon du PSELL.2 est sélectionné au sein d'une population exhaustive de '**titulaires principaux de revenus**' correspondant au fichier de l'Inspection Générale de la Sécurité sociale soit 154 534 titulaires principaux. Ces titulaires principaux sont aussi bien des personnes qui travaillent que des pensionnés ou des enfants titulaires d'une pension d'orphelin.

Cet échantillon compte 5 713 titulaires de revenus sélectionnés de manière aléatoire :

- chaque unité a *la même probabilité d'être choisie* (pas de sur-représentation, pas de stratification)
- et cette *probabilité est différente de 0* (sans biais).

Aucun sous-groupe n'est sur-représenté ni sous-représenté de manière volontaire. La sous-représentation de certains groupes pourrait être regrettable si leur intérêt particulier pour l'étude ne se révélait que plus tard.

Les titulaires de revenus ne sont que des **unités de sélection** conduisant à des **unités d'observation** : les ménages et les membres de ces ménages.

Chaque titulaire de revenu conduit à une adresse. A chaque adresse correspond un ménage. Les 5 713 titulaires de revenus conduisent donc à 5 713 ménages.

¹ Il est important de noter qu'aucune correction ne peut être introduite dans la base de sondage suite à l'observation sous peine de fausser les probabilités de sondage.

Ces 5 713 ménages forment un échantillon qui se répartit de la manière suivante :

Tableau 1 :

| <i>STATUT</i> | <i>FREQUENCES</i> | <i>POURCENT</i> |
|----------------------|-------------------|-----------------|
| Enquêtes réalisées | 2978 | 52.1 |
| « Hors Champ » | 269 | 4.7 |
| - décédés | 16 | |
| - émigrés | 61 | |
| - ménages collectifs | 192 | |
| « Echecs » | 2466 | 43.2 |
| - refus | 1922 | |
| - toujours absents | 410 | |
| - mauvaises adresses | 134 | |
| TOTAL | 5713 | 100.0 |

Le taux de sondage des ménages est de 3.69% ou, en d'autres termes, chaque titulaire principal en représente 27 (154 534 / 5 713).

Le taux de réponse est de 56.8 %. Chaque titulaire principal ayant répondu en représente 50 dans la population (154 534 / 3 247) et 1.76 dans l'échantillon (5 713 / 3 247).

Si un titulaire principal conduit à une et à une seule adresse, donc à un et à un seul ménage, l'inverse n'est pas nécessairement vrai. Un ménage peut être composé de plusieurs titulaires principaux parce que plusieurs titulaires peuvent conduire au même ménage.

L'échantillon des titulaires principaux était un échantillon aléatoire simple. La sélection de chaque unité était équiprobable. L'échantillon des ménages, par contre, ne répond plus à cette condition. La probabilité de sélection des ménages est inégale. Plus le nombre de titulaires principaux conduisant à un même ménage est élevé, plus le ménage a une probabilité de sélection élevée.

Afin de rétablir l'égalité des probabilités d'inclusion des ménages (équiprobabilité), ces derniers reçoivent un poids inverse au nombre de titulaires principaux susceptibles de conduire à leur adresse. Ce nombre n'est connu qu'au terme de l'observation des ménages.

Au sein de chaque ménage, tous les membres sont observés. Lorsque l'équiprobabilité de sélection des ménages est rétablie, tous les individus appartenant à ces ménages retrouvent eux aussi une probabilité de sélection égale. Chaque membre reçoit le poids du ménage et l'équiprobabilité de ces unités d'observation est rétablie.

L'ensemble des membres sélectionnés au cours du tirage du premier échantillon sont des **membres longitudinaux** : nous reviendrons sur ce point de manière plus détaillée par la suite.

Ces 10 967 individus appartenant aux 5 713 ménages forment un échantillon qui se répartit de la manière suivante :

Tableau 2

| <i>STATUT</i> | <i>FREQUENCES</i> | <i>POURCENT</i> |
|--------------------|-------------------|-----------------|
| Enquêtes réalisées | 8 232 | 75.1 |
| « Hors Champ » | 269 | 2.4 |
| « Echechs » | 2 466 | 22.5 |
| TOTAL | 10 967 | 100.0 |

Les enquêtes réalisées et les hors champ forment l'ensemble des répondants.

- Les membres observés ont reçu un poids égal au poids de leur ménage.
- Les 'titulaires hors champ' conduisent à des 'ménages hors champ'. Par définition, ces ménages 'hors champ' ne peuvent être observés. Nous estimons qu'un titulaire 'hors champ' correspond à un ménage 'hors champ' et à un individu 'hors champ'. Il reçoit donc un poids unitaire.

Ces individus 'hors champ' entrent également dans la composition de l'échantillon des individus. Ils y apportent donc leur poids.

- Les ménages ayant refusé de répondre ne peuvent être observés. Nous estimons qu'un refus correspond à un titulaire, à un ménage et à un seul individu. Il reçoit un poids unitaire et entre à ce titre dans le calcul de la taille de l'échantillon des individus.

CHAPITRE II

QU'EST-CE QU'UN ECHANTILLON REPRESENTATIF ?

II. QU'EST-CE QU'UN ECHANTILLON REPRESENTATIF ?

Un échantillon est représentatif de la population dont il est extrait **si et seulement si toutes les unités de la base de sondage ont une probabilité strictement supérieure à zéro d'être sélectionnées.**

Si ce critère est respecté, l'échantillon peut être utilisé **pour produire des estimations sans biais de quantités reliées à la population** qui lui correspond.

Si le fichier contient des erreurs ou des unités qui ne devraient plus s'y trouver (les 'hors champ'), il est exclu d'écarter volontairement ces unités de l'échantillon. Ces unités appartiennent à la base de sondage. Ecarter ces personnes de l'échantillon après les avoir sélectionnées reviendrait à faire 'comme si' leur probabilité de sondage était nulle alors que ce n'est manifestement pas le cas.

CHAPITRE III

**QU'EST-CE QU'UN ECHANTILLON LONGITUDINAL
ET QU'EST-CE QU'UN ECHANTILLON TRANSVERSAL ?**

III. QU'EST-CE QU'UN ECHANTILLON LONGITUDINAL ET QU'EST-CE QU'UN ECHANTILLON TRANSVERSAL ?

L'échantillon sélectionné en 1995 est destiné à la réalisation d'un panel. Un panel est une étude qui a la particularité de poursuivre deux objectifs simultanément. D'une part, il permet de suivre la trajectoire d'un certain nombre d'individus au cours d'un certain nombre d'années. D'autre part, il doit permettre de produire chaque année des estimations sans biais de quantités relatives à la population au sein de laquelle il évolue.

1. UN ECHANTILLON LONGITUDINAL

'Suivre la trajectoire d'un certain nombre d'individus au cours d'un certain nombre d'années' est une opération strictement longitudinale qui s'effectue sur la base d'un **échantillon d'unités longitudinales** : chaque année les mêmes individus sont observés et la quantité d'informations disponibles pour chacun d'eux augmente ainsi progressivement.

L'échantillon longitudinal est composé et n'est composé que des individus sélectionnés la première année. Cet échantillon **est représentatif et n'est représentatif que de la population en l'année où il a été tiré.**

L'évolution de la population de référence au fil des années n'a strictement aucun effet sur l'échantillon longitudinal ni sur son évolution.

2 DES ECHANTILLONS TRANSVERSAUX

'Produire chaque année des estimations sans biais de quantités relatives à la population au sein de laquelle l'échantillon évolue' est une opération strictement transversale ou annuelle (si les observations ont lieu chaque année). Cette opération s'effectue sur la base **d'échantillons transversaux**. Un échantillon transversal est un échantillon ponctuel dont la fonction est de **représenter l'état de la population au moment où il est observé**. Cette représentativité transversale suppose notamment que chaque année de nouveaux individus entrent dans l'échantillon longitudinal.

Les échantillons transversaux sont donc composés

- des individus longitudinaux
- d'individus présents dans le pays au moment du tirage du premier échantillon, qui n'ont pas été sélectionnés et qui rejoignent les individus longitudinaux dans leurs ménages
- et d'individus absents du pays au moment du tirage du premier échantillon et qui sont 'nés' dans le pays depuis lors, soit dans les ménages formés par les individus longitudinaux, soit par le biais d'une sélection annuelle volontaire d'un sous-échantillon représentatif de ces nouvelles 'naissances' ('nouveau-nés' et immigrants).

3. *UNITES LONGITUDINALES ET UNITES TRANSVERSALES*

Seuls les individus sont des **unités longitudinales**. Ils ne peuvent ni se diviser ni se fusionner. Ils restent identifiables sans ambiguïté d'année en année.

Ces unités doivent appartenir à un échantillon représentatif d'une population à un moment donné. Ce ne sont, par exemple, que les individus sélectionnés la première année du panel. Ce pourrait être aussi la combinaison d'un sous-ensemble aléatoire d'individus appartenant au premier échantillon sélectionné pour le panel et d'un sous-ensemble aléatoire d'individus sélectionnés dans la population de référence quatre années plus tard (renouvellement de l'échantillon par rotation).

Les **unités transversales** sont

- les ménages (unités qui peuvent se diviser ou se fusionner),
- les individus présents dans le pays au moment du tirage du premier échantillon qui rejoignent les individus longitudinaux dans leurs ménages
- et les individus 'nés' dans le pays depuis le tirage du premier échantillon.

Leur existence dans les échantillons annuels est ponctuelle. Elle maintient la représentativité transversale des échantillons successifs.

Bien que les ménages ne soient que des unités transversales, leurs caractéristiques peuvent toujours être attribuées à leurs membres (individus longitudinaux). Les caractéristiques des ménages peuvent évoluer, les ménages peuvent changer de structure, se diviser, se fusionner mais l'unité individuelle d'observation longitudinale reste identique à elle-même. Elle sera simplement caractérisée par les modifications du contexte familial dans lequel elle évolue.

Il est également possible de suivre des ménages à travers le temps. Dans ce cas, l'identité d'un ménage s'établira, par exemple, en fonction de l'identité de la « personne de référence ». Il convient de ne pas perdre de vue que cette référence a été choisie et définie d'une manière plus ou moins arbitraire. En aucun cas cet échantillon de ménage ne sera à l'image de l'évolution des ménages dans une population de référence.

CHAPITRE IV

QU'EST-CE QU'UN TAUX DE REPONSE CORRECTIF ?

IV. QU'EST-CE QU'UN TAUX DE REPONSE CORRECTIF ?

Si l'on se réfère au tableau de présentation de la distribution des individus dans le premier échantillon, on peut considérer que 75.1% des individus ont participé à l'observation. Dans ce cas, on parlera d'un taux de réponse descriptif.

On peut aussi affirmer que le taux de réponse est de 77.5% des individus appartenant à l'échantillon. Dans ce second cas, il s'agit d'un **taux de réponse correctif**.

Un taux de réponse descriptif fournit des informations concernant le déroulement du processus de sélection : '269 individus hors champ dans un échantillon de 10 967 individus représentent un taux de 2.4% de hors champ'. Un tel taux est aussi un taux descriptif.

Un taux de réponse correctif prend en compte toute l'information nécessaire pour corriger l'estimateur des poids des individus en fonction de la non-réponse. Il correspond au rapport entre les répondants et le nombre d'unités dans l'échantillon.

Dans ce calcul, les hors champ ne peuvent pas être considérés comme des 'non-répondants'.

1. TAUX DE REPONSE CORRECTIF TRANSVERSAL

La première année d'observation d'un panel fournit un et seul échantillon purement transversal et identique à tout échantillon utilisé pour n'importe quel type d'enquête. Le taux de réponse correctif transversal des individus de la première vague du panel est donc

$$R_{tr}(t) = \frac{\text{nombre de répondants (au temps } t)}{\text{nombre d'unités dans l'échantillon (au temps } t)}$$

soit :

$$\frac{8232 + 269}{10\,967} = 77.5\%$$

2. *TAUX DE REPONSE CORRECTIF LONGITUDINAL*

Le taux de réponse correctif longitudinal correspond au rapport entre *le nombre d'unités longitudinales* ayant répondu à une vague *t* et *le nombre d'unités dans la première vague*.

Soit :

$$Rlg(t) = \frac{\text{nombre d'unités longitudinales qui répondent (au temps } t)}{\text{nombre d'unités dans l'échantillon de la vague 1}}$$

Ce taux de réponse peut se calculer pour l'ensemble de l'échantillon ou au niveau de différentes strates si l'on soupçonne que certaines catégories de la population ont des raisons de répondre moins souvent que d'autres.

CHAPITRE V

**A QUOI SERT-IL DE "PONDERER" LES UNITES D'OBSERVATION
D'UN ECHANTILLON ?**

V. A QUOI SERT-IL DE "PONDERER" LES UNITES D'OBSERVATION D'UN ECHANTILLON ?

1. REPRESENTATIVITE ET PONDERATION

Il est impératif de comprendre que la pondération des unités d'un échantillon ne peut rien changer au caractère représentatif ou non-représentatif de l'échantillon. La condition sous laquelle la représentativité d'un échantillon est assurée, a été clairement énoncée.

Si l'échantillon est représentatif, il permet de produire des estimations sans biais de quantités relatives à la population dont il est extrait. **La pondération sert à améliorer la qualité des estimations** en remédiant aux perturbations introduites éventuellement par d'autres sources de biais telles que, par exemple, la non-réponse.

Si un échantillon n'est pas représentatif (choix volontaire des unités observées, échantillon de 'prototypes', échantillon longitudinal 'cylindré'), la pondération des unités de cet échantillon ne le rendra jamais représentatif de quoi que ce soit (si ce n'est d'un univers mental qui peut, par ailleurs, avoir sa propre cohérence interne).

2. PONDERER L'ECHANTILLON INITIAL

L'échantillon initial est un échantillon transversal.

Un poids a déjà été attribué aux membres de l'échantillon initial (inversement proportionnel au nombre de titulaires principaux de revenus). Ce poids a permis de rendre aux unités d'observation un caractère équiprobable que les unités de sélection de l'échantillon ne permettaient pas d'assurer.

Cette technique suppose que l'on connaisse très exactement la clé de passage des unités de sélection aux unités d'observation. D'une manière générale, elle suppose que l'on connaisse la structure des liens qui unissent les unités appartenant à un univers quelconque 'U1' et les unités appartenant à un autre univers 'U2'.

Un retour en arrière s'impose, ici, car le rétablissement de l'équiprobabilité entre les unités d'observation supposait une opération préalable dont il n'a pas encore été fait état.

première étape

Tous les titulaires principaux de revenus n'avaient pas répondu aux sollicitations des enquêteurs (2 466 refus). La non-réponse de ces titulaires principaux entraînait la non-réponse d'un certain nombre de ménages et donc d'un certain nombre d'individus. Ce phénomène risquait donc d'introduire certains biais dans les estimations des quantités mesurées par l'étude dans la mesure où la non-réponse est non-ignorable et l'on dit de la non-réponse qu'elle est non-ignorable lorsque la probabilité de répondre ou de ne pas répondre est liée à des variables mesurées par l'enquête.

En l'occurrence, la non-réponse était liée à la distribution géographique des titulaires principaux (par canton).

Le parallélisme entre la distribution des titulaires principaux dans le fichier de référence et dans l'échantillon qui nous a été fourni est tel qu'une sur-représentation ou sous-représentation géographique des ménages au sein de l'échantillon des adresses qui ont été mises à notre disposition est tout à fait exclue.

Par contre, le taux de réponse obtenu suite à l'enquête présente des disparités selon les cantons. Un taux de correction peut donc être calculé pour chaque titulaire résidant dans chaque canton et un premier poids peut être calculé pour les titulaires de revenus afin de compenser ces taux de réponse inégaux selon les cantons de résidence.

Ce poids de sélection correspond, pour chaque titulaire, à **l'inverse du taux de réponse correctif transversal** dans son canton :

$$Rtr(t) = \frac{\text{nombre de répondants dans le canton}}{\text{nombre d'unités dans l'échantillon du canton}}$$

Les titulaires répondants ainsi pondérés se distribuent dans les cantons de la même manière que les titulaires de l'ensemble de l'échantillon. L'inégalité des probabilités de sélection selon les cantons est compensée.

Les titulaires ayant refusé de répondre sont remplacés par les poids accordés aux titulaires ayant répondu et, ceci, de manière inversement proportionnelle au taux de réponse dans chaque canton. La taille initiale de l'échantillon n'est donc pas modifiée. Les refus sont compensés par un accroissement du poids des répondants.

Cette pondération a un double résultat. Elle maintient le degré de précision des estimations qui seront produites : n (taille de l'échantillon) reste constante. Elle améliore la qualité des résultats en corrigeant la non-réponse non-ignorable.

deuxième étape

Les ménages héritent de ces poids de sélection attribués aux titulaires de revenus puisque chaque titulaire conduit à un et un seul ménage.

troisième étape

Ce n'est qu'ensuite que les ménages sont pondérés en raison inverse du nombre de titulaires principaux susceptibles de conduire à chaque ménage.

Le poids d'un ménage est donc égal à l'inverse du nombre des **titulaires principaux pondérés** susceptibles de conduire à ce ménage

C'est ce poids des ménages qui est attribué à leurs membres. La physionomie de l'échantillon initial s'en trouve bien entendu modifiée.

Tableau 3 :

Structure de l'échantillon des ménages avant et après pondération

| <i>STATUT</i> | <i>FREQ. 1994 avant pondération</i> | <i>FREQ. 1994 après pondération</i> |
|--------------------|---|---|
| Enquêtes réalisées | 8 232 | 10 497 |
| « Hors Champ » | 269 | 470 |
| « Echechs » | 2 466 | - |
| TOTAL | 10 967 | 10 967 |

Tableau 4 :

Distribution des taux de réponse des titulaires selon les cantons

| <i>Canton</i> | <i>Taux de réponse</i> |
|----------------|------------------------|
| LUXEMBOURG | .599 |
| CAPELLEN | .574 |
| ESCH / ALZETTE | .583 |
| LUX. CAMPAGNE | .617 |
| MERSCH | .523 |
| CLERVAUX | .549 |
| DIEKIRCH | .421 |
| REDANGE | .588 |
| VIANDEN | .500 |
| WILTZ | .583 |
| ECHTERNACH | .504 |
| GREVENMACHER | .528 |
| REMICH | .534 |
| TOTAL | .568 |

Tableau 5 :

Comparaison des distributions des titulaires principaux selon les cantons dans la population et dans l'échantillon total

| <i>CANTON</i> | <i>POPULATION FREQUENCE</i> | <i>ECHANTILLON FREQUENCE</i> |
|------------------|---------------------------------|----------------------------------|
| LUXEMBOURG | 19.6 | 20.2 |
| CAPELLEN | 7.7 | 7.4 |
| ESCH / ALZETTE | 32.2 | 31.7 |
| LUX. CAMPAGNE | 9.3 | 9.3 |
| MERSCH | 5.1 | 4.9 |
| CLERVAUX | 2.6 | 2.5 |
| DIEKIRCH | 5.9 | 5.7 |
| REDANGE | 2.8 | 3.1 |
| VIANDEN | 0.6 | 0.6 |
| WILTZ | 2.7 | 2.6 |
| ECHTERNACH | 3.4 | 3.6 |
| GREVENMACHER | 4.6 | 5.0 |
| REMICH | 3.5 | 3.3 |
| Donnée manquante | 0.0 | 0.0 |
| TOTAL | 154 534 | 5713 |

CHAPITRE VI

**FAUT-IL REPONDERER L'ECHANTILLON LONGITUDINAL ET
COMMENT ?**

VI. FAUT-IL REPONDERER L'ECHANTILLON LONGITUDINAL ET COMMENT ?

1. INDICATION DE REPONDERATION

Le PSELL.2, comme tous les panels, est soumis à un phénomène inévitable : l'érosion de l'échantillon suite à la répétition des observations effectuées dans les mêmes ménages. Cette érosion dite 'attrition' de l'échantillon provient de la lassitude des personnes interrogées à de multiples reprises qui refusent de poursuivre leur collaboration à une étude de longue durée¹.

Cette érosion liée à la non-réponse annuelle des membres **longitudinaux réduit le nombre des répondants** et **peut être une source de biais si la non-réponse est non-ignorable**. La repondération annuelle des unités de l'échantillon longitudinal est destinée à réduire ces biais et à améliorer la qualité des estimations.

Les taux de réponse des individus longitudinaux font donc l'objet d'une analyse très intensive. Chaque année, on examine la répartition des taux de réponse selon un grand nombre de variables mesurées en 1994 (première vague d'observation de l'échantillon). L'objectif de cette procédure est de repérer les catégories d'individus importantes pour l'étude qui manifesteraient une plus faible propension à répondre que les autres catégories d'individus. Si aucune concentration de la non-réponse n'apparaît dans aucune catégorie de population stratégique pour l'étude, on est en droit de considérer que la non-réponse est un phénomène 'général' ou aléatoire, sans effet sur la qualité des résultats de l'étude. Dans le cas contraire, il y a lieu de repondérer les membres longitudinaux en s'efforçant de compenser cette distribution inégale des taux de réponse.

Il s'agit d'expliquer le mieux possible la non-réponse. La part de la non-réponse qui reste 'inexpliquée' par ces variables est 'ignorable' parce qu'elle n'est pas liée à des variables mesurées par l'enquête. Les personnes qui ne répondent pas ne se distinguent des autres personnes que par le seul fait qu'elles se lassent de répondre.

¹ On distingue ce type de « non-réponse totale de vague » ou « chronique » de la « non-réponse partielle » où certaines observations visées par l'enquête sont manquantes pour certaines unités d'observation (dans ce dernier cas, on aura généralement recours à différentes procédures « d'imputation »).

2. LE TAUX DE REPONSE CORRECTIF LONGITUDINAL

Le taux de réponse correctif longitudinal se calcule selon la formule

$$Rlg(t) = \frac{\text{nombre d'unités longitudinales qui répondent (au temps } t)}{\text{nombre d'unités dans l'échantillon de la vague 1}}$$

La variable dépendante à prendre en compte est 'le taux de réponse correctif'. Ce taux de réponse inclut les 'hors champ' dans les répondants. Ces 'hors champ' ne sont pas seulement les 'hors champ' enregistrés en 1994 mais aussi les nouveaux 'hors champ' apparus comme tels au cours des observations successives (les nouvelles personnes décédées, émigrées et en ménage collectif).

Ce taux de réponse est de 75.8% en 1997 (enquête 1998), soit, en valeurs corrigées par le poids de sélection calculé en 1994, 7 279 enquêtes réalisées et 1 029 'hors champ' (hors champ de 1994, augmentés des décès, des émigrés et des personnes passées en ménages collectifs depuis 1994).

Tableau 6 :

Distributions de l'échantillon longitudinal en 1994, 1995 et 1996 pondérées par le poids de sélection (wt94)

| <i>STATUT</i> | <i>FREQ. 1994 Pondérées</i> | <i>FREQ. 1995 Sous wt94</i> | <i>FREQ. 1996 Sous wt94</i> | <i>FREQ. 1997 sous wt94</i> |
|--------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Enquêtes réalisées | 10 497 | 8 477 | 7 888 | 7 279 |
| « Hors Champ » | 470 | 700 | 872 | 1 029 |
| « Echecs » | - | 1 790 | 2 206 | 2 659 |
| TOTAL | 10 967 | 10 967 | 10 967 | 10 967 |

Les refus de 1994 ont été compensés par les poids des répondants de 1994. Seuls les nouveaux refus enregistrés en 1995, en 1996 et en 1997 apparaissent. Du numérateur, ils sont passés au dénominateur de la formule de calcul du taux de réponse. Ils représentent un total de 2 659 individus en 1997.

La taille de l'échantillon longitudinal ne se modifie pas d'une année à l'autre. Seules les enquêtes réalisées tendent à diminuer. L'échantillon compte toujours 10 967 individus pondérés par le poids de sélection initial. Ces 10 967 individus circulent entre les trois groupes emportant avec eux, le poids qui leur a été attribué en 1994.

Le taux de réponse correctif longitudinal se rapporte chaque année à l'échantillon initial pondéré (1994). Le nombre (pondéré) des 'refus' et des 'hors champ' est donc cumulatif.

Des 'refus' des années antérieures pourront être récupérés par la suite; des émigrés 'hors champ' pourront éventuellement revenir dans leur ménage d'origine mais, inexorablement, la proportion des enquêtes réalisées et des 'hors champ' (les 'répondants') aura tendance à diminuer et la proportion des échecs aura tendance à augmenter. Les poids des répondants tendront à augmenter pour compenser ces échecs mais la taille de l'échantillon longitudinal restera constante pendant toute la durée de vie de cet échantillon.

3. *LA NON-REPONSE EST-ELLE IGNORABLE ?*

La non-réponse est-elle ignorable ou risque-t-elle d'introduire certains biais parce qu'elle est liée au moins partiellement à des variables mesurées par l'enquête ? Pour tenter d'expliquer cette non-réponse, il est fait appel à des variables qui couvrent différents volets de l'étude :

- Le niveau de formation le plus élevé atteint par chaque personne.
- Le nombre d'emplois dans le ménage.
- La diversité des revenus existants dans le ménage.
- Le montant de ces revenus.
- La taille du ménage.
- Le nombre d'adultes.
- Le nombre d'enfants.
- Le revenu disponible pour chaque unité de consommation.
- Le fait d'être propriétaire ou locataire de son logement.
- Le type d'habitation (individuel, individuel en série, appartement dans un immeuble plus ou moins grand).

- Des caractéristiques individuelles : âge, sexe, nationalité, ancienneté dans le pays, localisation géographique (canton).
- Des caractéristiques du chef de ménage : âge, sexe ; a-t-il des revenus propres ? a-t-il un emploi ?
- L'effet 'enquêteur' : la distribution des adresses aux enquêteurs s'effectue selon des zones géographiques plus détaillées que les cantons.

Toutes ces variables sont transformées en 'dummies' et entrées dans une analyse discriminante¹ 'pas à pas' (stepwise).

Un 'arbre' des taux de réponse est constitué à partir des variables les plus discriminantes (division hiérarchique descendante). Les variables sont introduites aussi longtemps qu'elles accroissent significativement le pourcentage de variance expliquée. L'analyse est arrêtée dès que les variables ajoutées distinguent des sous-catégories de trop petite taille.

Les variables les plus discriminantes sont surtout : la distribution géographique et certaines catégories de revenus (sans effet linéaire). Le nombre d'enfants, la taille du ménage et l'âge du chef de ménage influencent également les taux de réponse.

La combinaison de ces variables ne permet d'expliquer que 6% du taux de réponse en 1997. Nous émettons l'hypothèse selon laquelle la part non-expliquée de la non-réponse est 'négligeable' en ce sens qu'elle n'est pas liée aux variables importantes pour l'objet de l'étude.

¹ Pour mémoire, la variable dépendante est une variable dichotomique. Les probabilités de réponse peuvent être modélisées à l'aide d'une régression logistique. L'analyse discriminante, bien qu'elle soit fondée sur un autre modèle, donne des résultats équivalents selon le point de vue qui nous occupe.

4. *LES POIDS LONGITUDINAUX AJUSTES*

Les poids des individus longitudinaux doivent être ajustés en 1997 en raison de la répartition non aléatoire des taux de réponses.

Wil 97 est égal au produit

- des poids des individus longitudinaux en 1994 (**wt94**)
- et de l'inverse des taux de réponse en 1997 (**1 / Rlg (1997)**)

soit :

$$wil\ 97 = wt\ (94) * \frac{1}{Rlg\ (97)}$$

soit :

$$\frac{1}{\text{probabilité de sélection en 94}} * \frac{\text{n unités en 94}}{\text{n rép. longitud. en 1997}}$$

sachant que

- le rapport entre le nombre d'individus longitudinaux ayant répondu en 1997
- et le nombre d'unités dans l'échantillon en 1994
- peut varier selon les strates définies par les variables discriminantes.

Wil 97 est un estimateur sans biais si **Rlg (97)** est égal à la vraie probabilité de réponse des unités à la vague 1997.

Tableau 7 :

**Distribution de l'échantillon longitudinal pondéré
en 1994 (wt94), 1995 (wil95), 1996 (wil96) et en 1997 (wil97)**

| <i>STATUT</i> | <i>FREQ. 1994 pondérées</i> | <i>FREQ. 1995 pondérées</i> | <i>FREQ. 1996 pondérées</i> | <i>FREQ. 1997 pondérées</i> |
|--------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Enquêtes réalisées | 10 497 | 10 177 | 9 992 | 9 738 |
| « Hors Champ » | 470 | 790 | 975 | 1 229 |
| TOTAL | 10 967 | 10 967 | 10 967 | 10 967 |

CHAPITRE VII

**SAUVEGARDER LA REPRESENTATIVITE ANNUELLE DES
ECHANTILLONS TRANSVERSAUX**

VII. SAUVEGARDER LA REPRESENTATIVITE ANNUELLE DES ECHANTILLONS TRANSVERSAUX

Chaque nouvelle vague d'enquêtes correspond non seulement à une nouvelle étape dans l'élaboration de l'échantillon longitudinal mais aussi à la création d'un nouvel échantillon transversal. Ce nouvel échantillon transversal pose deux problèmes.

D'une part, chaque échantillon transversal doit rester **représentatif de la population au moment de la nouvelle observation**. D'autre part, il doit prendre en compte l'ajustement des **poids des individus longitudinaux** en fonction du taux de réponse longitudinal.

1. CONCERNANT LA REPRESENTATIVITE

Le nouvel échantillon transversal reste représentatif de la population s'il reproduit intégralement tous les phénomènes démographiques qui se sont produits dans la population. Or, spontanément, il ne les reproduit que partiellement.

- Il génère des 'naissances' : de nouveaux ménages se forment (par mariage), des nouveau-nés apparaissent dans les ménages et des cohabitants rejoignent les ménages formés par les individus longitudinaux.
- Il génère des 'morts' : des ménages se défont (par divorce qui génèrent à leur tour de nouveaux ménages), des personnes décèdent, d'autres quittent le pays ou entrent dans des ménages collectifs.
- Il ne reproduit pratiquement pas les 'naissances' par immigration.

Pour prendre en compte les naissances, un nouvel échantillon supplémentaire devrait être introduit chaque année comprenant des nouveau-nés et des immigrants.

Le panel suit des ménages. Il produit donc par lui-même les nouveau-nés. Il n'est donc pas nécessaire de sélectionner chaque année un échantillon supplémentaire de nouveau-nés.

Les immigrants, par contre, ne se conforment pas au même modèle. Les ménages composés d'individus longitudinaux intègrent bien quelques nouveaux immigrants, chaque année, dans le cadre de la recomposition de familles ou suite à des mariages. Ce phénomène est tout à fait marginal. Dans la plupart des cas, les nouveaux immigrants forment eux-mêmes des ménages. Ils échappent entièrement à l'observation du panel. Ils doivent donc être introduits volontairement dans l'échantillon. Pour des raisons pratiques, cette opération n'est effectuée qu'un an sur deux.

L'observation effectuée en 1998, concernant l'année 1997, étant la quatrième vague, l'échantillon de nouveaux immigrants, introduit dans le panel un an plutôt, reste sous observation. Un nouvel échantillon d'immigrants sera introduit en 1999.

Tableau 8 :

**Composition des échantillons transversaux
en 1994, 1995, 1996 et 1997 (sans pondération)**

| <i>Statut dans l'échantillon</i> | <i>1994</i> | <i>1995</i> | <i>1996*</i> | <i>1997*</i> |
|---|--------------|--------------|--------------|--------------|
| Indiv. Longitudinaux : « réalisés » | 8 232 | 6 597 | 6 119 | 5 627 |
| Nouveaux individus transversaux (Cohabitants, bébés, immigrants) | | 210 | 658 | 1 027 |
| Total | 8 232 | 6 807 | 6 777 | 6 654 |

* y compris les nouveaux immigrants entrés au pays en 1995 et 1996

**2. CONCERNANT LA REPONDERATION DES ECHANTILLONS
TRANSVERSAUX**

Pour la clarté nous désignerons dorénavant par '**poids initial**', le poids annuel des individus longitudinaux (présents dans chaque fichier transversal).

Nous considérons des **ménages entiers** et non pas seulement les personnes. Cette procédure est efficace lorsqu'on souhaite produire des estimations au niveau des ménages.

Par contre, cette approche pose le **problème de l'entrée de nouveaux membres** dans les ménages. Ces nouveaux membres pouvaient être présents dans la population au moment de la sélection de l'échantillon longitudinal mais ils n'ont pas été sélectionnés. Dans ces conditions, comment faut-il les pondérer ?

2.1 Repondération de l'ensemble des personnes qui forment les ménages

Admettons dans un premier temps que la population de référence ne soit pas modifiée. Elle n'a connu aucune 'naissances' et aucun 'décès'

L'observation de ménages entiers montre que des nouvelles personnes sont venues s'ajouter aux membres du ménage observé l'année précédente.

Ces 'membres ajoutés' étaient présents dans la population de référence (U) au moment de la sélection de l'échantillon (S) en 1994. Ils n'ont pas été sélectionnés. Ils sont observés au cours d'une vague ultérieure (en 1995 ou en 1997) parce qu'ils ont rejoint un ménage comprenant, au moins, un membre (m) de l'échantillon (S), soit un membre longitudinal.

Comment pondérer ces membres non-longitudinaux sans biaiser l'estimateur des poids ?

On peut montrer que la méthode dite du 'partage des poids' permet d'obtenir un poids pour ces membres sans introduire aucun biais dans l'estimateur des poids à condition que ces membres 'transversaux' reçoivent dans un premier temps **un poids initial de '0'** :

- les individus longitudinaux ont un **poids initial** déterminé par la repondération annuelle des individus longitudinaux (cf. VI.4)
- les individus non-longitudinaux reçoivent, dans un premier temps, un **poids initial** égal à '0'.
- **Le poids du ménage** est égal à la somme des poids des membres divisée par le nombre de membres, soit la valeur moyenne des poids de l'ensemble des membres.
- **Chaque membre** reçoit ensuite un **poids final** : la moyenne des poids des membres du ménage.

Au cours de cette série d'opérations, les membres longitudinaux ont transféré une partie de leur poids aux membres non-longitudinaux de leur ménage. Ils ont partagé leur poids initial avec les nouveaux cohabitants. Ceci permet de prendre en compte la richesse de l'information apportée par ces nouveaux membres et de maintenir la représentativité de l'échantillon par rapport à la population de référence dans la mesure où **les ménages ont évolué**.

2.2 Pondération des naissances : les nouveau-nés¹

Nous supposons, dans un second temps, que la population n'évolue qu'en fonction des naissances d'enfants (et des décès).

Les nouveau-nés apparaissent dans les ménages formés de membres longitudinaux.

La méthode du 'partage des poids' s'applique ici aussi. Toutefois, les membres non-longitudinaux pouvaient être présents dans la population lors de la sélection de l'échantillon initial (S). Les nouveau-nés, par contre, sont des individus supplémentaires dans l'échantillon parce qu'ils sont aussi des individus supplémentaires dans la population.

Ils doivent recevoir un poids qui fasse d'eux, des unités supplémentaires dans l'échantillon

- ils reçoivent un **poids initial** égal à ' 0 '
- le **poids du ménage** est égal à la somme des poids des individus divisée par le nombre de membres dans le ménage moins les nouveau-nés.

Leur poids final est donc égal à la moyenne des poids des autres membres du ménage.

Ces nouveau-nés restent des individus transversaux. Cette opération doit donc être réitérée chaque année pour l'ensemble des 'nouveau-nés' entrés dans l'échantillon depuis la sélection de l'échantillon initial.

Si un échantillon supplémentaire de nouveau-nés est tiré, ceux-ci assurent la représentativité de l'échantillon. Les nouveau-nés qui entrent directement par le biais des ménages devraient, alors, être pondérés différemment. Il ne leur appartiendrait plus de refléter la croissance de la population et de la répercuter dans l'échantillon. Dans ce cas, ils seraient pondérés comme tous les membres transversaux qui appartenaient déjà à la population au moment de la sélection du premier échantillon (VII.2.1).

¹ P. Lavallée: « Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode des poids partagés », Techniques d'enquête, Vol. 21, n°1, pp. 27-35, Statistique Canada, juin 1995.

2.3 Pondération des naissances : les nouveaux immigrants

Nous avons supposé jusqu'ici que la population n'évolue qu'en fonction des naissances d'enfants (et des décès). Dans un troisième temps, il faut prendre en compte le fait que la population évolue aussi en fonction de la 'naissance d'immigrants'.

Des nouveaux immigrants apparaissent spontanément dans le panel. Ils y forment un sous-groupe de l'ensemble des individus qui rejoignent les ménages. A une différence près. Ils n'étaient pas présents dans la population au moment du tirage du premier échantillon.

L'entrée de ces nouveaux immigrants dans les ménages du panel est un phénomène marginal. La plupart d'entre eux forment de nouveaux ménages. Le panel ne peut pas prendre en compte spontanément cette modification de la structure de la population.

Il est donc indispensable de sélectionner un nouvel échantillon représentatif des immigrants entrés dans le pays depuis la sélection du premier échantillon.

1. **Dans la deuxième vague d'enquête**, les nouveaux immigrants entrés 'par la porte' des ménages du panel ont été pris en compte. Ils ont été pondérés de la même manière que les nouveau-nés. Ils accroissent la taille des ménages et la taille de la population.

Ils ont reçu un **poids initial** égal à '0'.

Le **poids du ménage** est égal à la somme des poids des individus divisée par le nombre de membres dans le ménage excluant les nouveaux immigrants.

Le **poids final** des nouveaux immigrants est donc égal à la moyenne des poids des autres membres du ménage.

2. **Dans la troisième vague (1996)**, un échantillon de nouveaux immigrants entrés au pays depuis 1994 est sélectionné sur la base du fichier de la Sécurité Sociale. Ces nouveaux immigrants sont des titulaires principaux de revenus qui conduisent à des ménages. La probabilité de sélection de ces ménages est inégale.

L'équiprobabilité entre les ménages est rétablie, selon la même procédure que dans l'échantillon initial (1994), par pondération en raison inverse du nombre de titulaires principaux. Seuls les titulaires entrés au pays depuis 1994 sont pris en compte. Le poids du ménage est attribué à l'ensemble de ses membres (entrés au pays depuis 1994).

Toutefois, l'observation de ménages entiers risque d'inclure des individus de différentes origines

- déjà présents dans la population au moment de la sélection du premier échantillon,
- des nouveau-nés,
- voire, des individus longitudinaux déjà présents dans l'échantillon.

La probabilité de sélection de ces individus, sélectionnés de manière irrégulière, est mal définie. La procédure de pondération n'est pas détaillée ici.

Rappelons que cet échantillon supplémentaire de ménages et d'individus contribue simplement à maintenir la représentativité transversale des échantillons annuels. Ses membres ne sont en aucun cas des membres longitudinaux.

3. **Dans la quatrième vague d'enquête**, de nouveaux immigrants entrent 'par la porte' des ménages du panel. Ils accroissent la taille des ménages et la taille de la population (à défaut d'un nouvel échantillon d'immigrants).

Ils reçoivent un **poids initial** égal à '0'.

Le **poids du ménage** est égal à la somme des poids des individus divisée par le nombre de membres dans le ménage excluant les nouveaux immigrants.

Le **poids final** des nouveaux immigrants est donc égal à la moyenne des poids des autres membres du ménage.

On revient donc, ici, à la procédure utilisée dans la deuxième vague d'enquêtes.

3 *PROFILS DES ECHANTILLONS TRANSVERSAUX PONDERES*

Tableau 9 :

**Composition des échantillons des individus transversaux
en 1994, 1995, 1996 et 1997 (pondérés)**

| <i>STATUT</i> | <i>FREQ. 94 pondérées (wt94)</i> | <i>FREQ. 95 pondérées (wtf95)</i> | <i>FREQ. 96 pondérées (wit96tot)</i> | <i>FREQ. 97 pondérées (wit97tot)</i> |
|-------------------------|--|---|--|--|
| 1. Indiv. Longitudinaux | 10 497 | 9 844 | 9 798 | 9 744 |
| 2. Indiv. Transversaux | | 103 | 829* | 1 134* |
| Total | 10 497 | 9 947 | 10 627* | 10 878* |

* y compris les nouveaux immigrants entrés au pays en 1995 et 1996

Tableau 10 :

**Composition des échantillons des ménages transversaux
pondérés en 1994, 1995, 1996 et 1997**

| <i>STATUT</i> | <i>FREQ. 94 pondérés (mwt94)</i> | <i>FREQ. 95 pondérés (mwt95)</i> | <i>FREQ. 96 pondérés (mwt96tot)</i> | <i>FREQ. 97 Pondérés (mwt97tot)</i> |
|-------------------------|--|--|---|---|
| 1. Ménages « observés » | 5 120 | 3 918 | 4 104* | 4 186* |

* y compris les nouveaux immigrants entrés au pays en 1995 et 1996

Pour mémoire :

wt94 est le poids des individus en 1994

mwt94 est le poids des ménages en 1994

*

wil95 est le poids longitudinal des individus en 1995

wtf95 est le poids transversal des individus en 1995

mwt95 est le poids des ménages en 1995

*

wil96 est le poids longitudinal des individus en 1996

wit96tot est le poids transversal des individus en 1996
(y compris les nouveaux immigrants)

mwt96tot est le poids des ménages en 1996
(y compris les nouveaux immigrants)

*

wil97 est le poids longitudinal des individus en 1997

wit97tot est le poids transversal des individus en 1997
(y compris les nouveaux immigrants)

mwt97tot est le poids des ménages en 1997
(y compris les nouveaux immigrants)