

N° de catalogue 93-18

**INSÉRER DES NOUVEAUX MEMBRES DANS UN PANEL
LONGITUDINAL DE MÉNAGES ET D'INDIVIDUS :
SIMULATIONS**

Février 1993

Bernard Gailly, CEPS/INSTEAD

Pierre Lavallée, Division des méthodes d'enquêtes sociales, Statistique Canada

La série de documents de recherche de l'EDTR est conçue en vue de communiquer les résultats des études ainsi que les décisions importantes ayant trait à l'Enquête sur la dynamique du travail et du revenu. Ils sont offerts en français et en anglais. Pour obtenir une description sommaire des documents disponibles ou un exemplaire de ces documents, communiquez avec Philip Giles, EDTR, par la poste à Édifice Jean-Talon, 11^{ième} étage, section D8, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6; par INTERNET: GILES@STATCAN.CA; par téléphone au (613) 951-2891; ou par télécopieur au (613) 951-3253.

SOMMAIRE

La comparaison des résultats obtenus dans le cadre des simulations effectuées sur l'échantillon observé par le PSELL (Panel Socio-économique *Liewen zu Lëtzebuerg+), en 1990, permet de constater trois choses essentielles.

- ! Les différents modèles présentent des résultats très comparables : les écarts entre les distributions obtenues sous chaque modèle et les distributions obtenues sous le modèle en vigueur actuellement sont très limités.

- ! Le modèle présenté par Ernst cumule trois avantages :
 - ! L'estimateur des poids est sans biais;
 - ! Il présente systématiquement les résultats les plus conformes aux distributions obtenues sous le modèle classique;
 - ! Il repose sur un postulat dont la portée est plus limitée que le postulat sous-jacent au modèle de Rao. Les nouveaux membres "non-légitimes" permettent d'accroître la variabilité intra-ménage et, donc, l'information à propos du ménage qui les accueille, mais ils partagent aussi certains traits communs avec les membres du ménage d'accueil.

Ils ne ressemblent pas nécessairement à l'ensemble de l'échantillon dans lequel ils entrent et ils ne sont pas représentatifs de la population-cible prise dans son ensemble à un moment donné.

Ils permettent de mieux cerner l'évolution de ces ménages d'accueil en apportant des informations originales.

Ils n'ont pas d'existence individuelle indépendante du ménage qui les accueille.

La régulation des poids s'effectue au sein de chaque ménage. Elle ne s'effectue pas au niveau de l'ensemble de l'échantillon (Rao).

! Ces conclusions rejoignent strictement les résultats obtenus par Lavallée et Hunter (1992). Ceci contribue à leur donner une certaine généralité.

On trouvera en Annexe 1, les résultats d'une seconde comparaison effectuée sur l'échantillon observé par le PSELL en 1987. Ils confirment les résultats obtenus sur l'échantillon observé en 1990.

Ce document est également publié comme Document PSELL n° 54, une publication du panel socio-économique luxembourgeois *Liewen zu Lëtzebuerg+, CEPS/Instead, Walferdange, Grand-Duché de Luxembourg, 1993.

TABLE DES MATIÈRES

	Page
1. INTRODUCTION	1
1.1. Évolution démographique générale	2
1.2. Effets d'échantillonnage	3
1.2.1. Effets démographiques internes	3
1.2.2. Effets de l'observation répétée	4
1.3. Simulations	5
2. ÉCHANTILLONNAGE ET PONDÉRATION "CLASSIQUE"	6
2.1. L'échantillon initial	6
2.2. Les statuts individuels dans l'échantillon	7
2.3. Calcul des pondérations	9
2.4. Implications	10
2.5. Inconvénient de ce procédé	13
3. LE "PARTAGE DES POIDS" (ERNST)	14
3.1. Principes	14
3.2. Implications	15
3.3. Simulation	19
4. "L'ESTIMATEUR COMPOSITE" (RAO)	29
4.1. Principes	29
4.2. Implications	32
4.3. Simulation	38
5. CONCLUSION	53
BIBLIOGRAPHIE	55
ANNEXE 1 : Comparaison des estimations selon le mode de pondération en 1987 (Classique et Partage du poids du ménage)	56
ANNEXE 2 : La procédure de Ernst ne comporte aucun biais	65

1. INTRODUCTION

Le PSELL, " Panel Socio-économique *Liewen zu Lëtzebuerg+ " a débuté en 1985. Il observe chaque année les mêmes personnes et les mêmes ménages. Il poursuit, comme toutes les études longitudinales, un objectif essentiel : recueillir des informations inter-annuelles reflétant l'évolution des niveaux de vie, des modes de vie et des conditions de vie des ménages.

L'observation annuelle fournit des images successives de l'état de la situation dans la population-cible. Chaque échantillon annuel peut donc être utilisé comme une enquête ponctuelle. Mais le caractère longitudinal de l'observation ne facilite pas la réalisation de ces études en coupes synchroniques.

En effet, l'échantillon initial n'a jamais été renouvelé au cours des années suivantes.

Il est donc soumis à des déformations progressives : il prend en compte certains aspects de la démographie générale mais il évolue également selon des tendances spécifiques.

- ! Il prend en compte l'évolution démographique générale de la population à laquelle il appartient : naissances, mariages, séparations, émigrations, décès.
- ! Il subit des effets spécifiques liés à l'évolution de l'échantillon observé. Ces effets spécifiques sont de deux ordres : d'une part, l'échantillon a sa propre démographie et d'autre part, sa composition est influencée par le caractère répétitif des observations.

Ces effets spécifiques peuvent être gérés de telle manière que l'échantillon garde ses propriétés statistiques fondamentales. Les procédures de pondération

permettent généralement de gérer certains aspects de cette évolution spécifique de l'échantillon.

Il existe plusieurs méthodes de pondération. L'objectif de ce document est de mettre en évidence les avantages comparatifs de différentes procédures de pondération et de proposer une solution qui permette d'accroître la sensibilité de l'échantillon aux phénomènes démographiques généraux.

1.1. Évolution démographique générale

L'échantillon reflète l'évolution démographique de la population parce qu'il est "représentatif" de la population de référence. Les naissances, les décès, les mariages, les divorces, les séparations, l'émigration sont des événements qui peuvent être observés aussi bien dans l'échantillon que dans la population.

Mais tous ces événements ne sont pas pris en compte au même titre dans l'évolution de l'échantillon du PSELL.

Ils sont toujours pris en compte lorsqu'il s'agit de décrire les caractéristiques des ménages. L'arrivée d'une personne supplémentaire accroît la taille du ménage. L'apport d'un nouveau revenu est répercuté dans le montant du revenu global du ménage. Le mariage d'un enfant donne lieu à la constitution d'un nouveau ménage.

Par contre, ces événements démographiques ne sont pas toujours reflétés dans l'échantillon des individus : par exemple, seuls les descendants des membres de l'échantillon initial sont admis au titre de membres légitimes de l'échantillon. Tous les autres nouveaux membres sont déclarés "non-légitimes" et exclus des analyses de l'échantillon annuel parce qu'il est délicat de leur attribuer une probabilité de sélection individuelle sans biaiser l'échantillon. Comment calculer la probabilité de

sélection initiale d'un jeune marié qui entre dans l'échantillon en 1989? Il appartenait déjà à la population-cible en 1985 mais il n'a pas été sélectionné dans l'échantillon initial.

1.2. Effets d'échantillonnage

1.2.1. Effets démographiques internes

Dans le cadre du PSELL, l'échantillon individuel ne reflète pas, actuellement, tous les aspects de l'évolution démographique générale. Les nouveaux membres, dits "non-légitimes", apportent le témoignage de ces évolutions mais ils ne sont pas pris en compte dans les analyses ponctuelles de l'échantillon individuel.

Cette position provoque une perte d'information considérable.

Les personnes et les ménages sélectionnés pour former l'échantillon vivent, en réalité, au sein d'une population plus large. Ils entretiennent constamment des échanges avec d'autres personnes appartenant à la "population-cible". Ils s'associent parfois avec des personnes qui n'appartiennent pas au panel pour former de nouvelles cellules familiales. Ils accueillent un ou plusieurs nouveaux membres "non-légitimes" dans une cellule existante au sein de l'échantillon.

Tous ces échanges entre l'échantillon et la population sont observés dans le cadre du panel. En effet, ces nouveaux membres se prêtent généralement aux entretiens proposés par les observateurs. Ils entrent donc dans l'échantillon annuel.

Leurs caractéristiques sont prises en compte au moment où l'analyse porte sur les conditions de vie des ménages mais elles restent inexploitées lorsque l'analyse porte sur les individus.

Cette perte d'information nuit à la représentativité et à la longévité de l'échantillon.

Certains phénomènes démographiques ne pourront pas être pris en compte dans les études ponctuelles, aussi longtemps que ces nouveaux membres du panel seront tenus à l'écart des analyses ponctuelles, bien qu'ils participent à l'observation.

1.2.2. Effets de l'observation répétée

La répétition annuelle de l'observation influence également l'évolution de l'échantillon.

Certains membres de l'échantillon initial refusent de répondre après un certain nombre d'années de participation; ces refus peuvent être temporaires et certains membres reprennent parfois leur collaboration après une période d'absence plus ou moins prolongée.

Une procédure de pondération annuelle permet de corriger les effets liés aux refus de répondre des membres de l'échantillon initial et de préserver ainsi les qualités statistiques de l'échantillon (estimateurs sans biais, précision des valeurs estimées).

Mais cette procédure exclut, actuellement, l'introduction des nouveaux membres "non-légitimes" dans les échantillons. Ce dispositif gère l'évolution interne du panel sans résoudre le problème qui met le plus en danger sa longévité : inexorablement, la taille de l'échantillon se réduit sous l'effet des décès partiellement compensés par les naissances, des émigrations non-compensées par les immigrations et des refus de répondre non-compensés par l'entrée de nouveaux membres.

La pondération des échantillons est devenue une procédure courante mais elle pose des problèmes particuliers lorsqu'elle doit s'appliquer à des échantillons qui s'inscrivent dans le cadre d'une observation longitudinale. Ce document propose et compare plusieurs solutions à ce problème.

1.3. Simulations

La procédure de pondération mise en oeuvre jusqu'ici sera décrite brièvement dans la première partie de ce document sous l'appellation "pondération classique".

Plusieurs modèles de pondération permettent de résoudre le problème de l'attribution d'un poids effectif aux nouveaux membres "non-légitimes". Les modèles présentés respectivement par Ernst (1989) et par J.N.K. Rao (papier non publié) ont été testés et discutés par P. Lavallée et L. Hunter (1992).

La seconde partie de ce document sera consacrée à une réplique de ces travaux.

Le mode de détermination des probabilités de sélection des nouveaux membres sera présenté dans le cadre de chaque modèle. Des variantes seront également proposées afin de mettre en évidence les avantages et les désavantages comparatifs de chaque modèle.

Chaque modèle de pondération sera appliqué à l'échantillon observé en 1990. Les résultats seront comparés aux estimations obtenues sous le modèle de pondération "classique", sachant que ce dernier ne prend pas en compte les nouveaux membres "non-légitimes".

Enfin, le modèle qui nous paraît être le plus approprié, sera appliqué à l'échantillon observé en 1987. Les résultats seront comparés aux estimations obtenues dans le

cadre du modèle de pondération "classique". Cette dernière simulation permettra de confirmer la stabilité des effets du modèle privilégié et d'observer éventuellement certains effets liés au caractère longitudinal de l'étude.

2. ÉCHANTILLONNAGE ET PONDÉRATION "CLASSIQUE"

2.1. L'échantillon initial

L'échantillon initial du PSELL est un échantillon aléatoire simple. Il représente les personnes résidant au Luxembourg et attachées au système de sécurité sociale ou à un autre système de protection sociale. Chacune de ces personnes est titulaire officiel d'un revenu. Ce fichier n'est donc pas un fichier couvrant l'ensemble des ménages ou l'ensemble des personnes résidant au Luxembourg.

Chaque personne sélectionnée conduit à une adresse. Cette adresse correspond à un ménage.

La procédure de sélection permet à plusieurs titulaires de revenus de conduire à la même adresse. Certains ménages ont donc une probabilité de sélection plus élevée que d'autres : plus le nombre de titulaires de revenus présents dans le ménage est élevé, plus la probabilité de sélection du ménage est élevée.

Pour corriger ce biais, il faut identifier le nombre de titulaires de revenus présents dans chaque ménage. Tous les membres du ménage étant interrogés, cette opération peut être effectuée a posteriori.

Le poids initial de chaque ménage correspond à l'inverse du nombre de membres titulaires d'un revenu.

Le poids initial des membres du ménage est égal au poids du ménage puisque tous les membres du ménage ont participé à l'enquête .

- ! Le poids du ménage est forcément égal à la moyenne des poids des individus.
- ! Tous les membres du ménage ont le même poids, ce qui confère à ces poids une cohérence interne ("consistency"). Cette situation ne sera plus respectée par la suite pour des raisons qui seront expliquées.

Il ne semble pas utile d'introduire d'autres corrections dans cet échantillon initial parce que les refus de répondre ne modifient pas de manière significative la structure de l'échantillon par rapport à la population-cible.

2.2. Les statuts individuels dans l'échantillon

À partir de la deuxième vague d'enquêtes, les poids sont d'abord calculés au niveau individuel et non plus au niveau du ménage.

D'une part, cette procédure permet de prendre en compte le fait que les personnes peuvent connaître des destins plus ou moins autonomes par rapport au ménage initial.

D'autre part, au fil des années, il est de plus en plus malaisé d'identifier les ménages qui se correspondent. Ils se transforment sous l'effet des entrées et des sorties, des mariages et des séparations.

Avant de calculer les poids individuels, l'échantillon doit être décomposé en 6 sous-groupes parce que tous les membres ne sont pas pris en compte au même titre.

- ! Les membres du panel qui refusent de répondre : ce sont des personnes sélectionnées dans le premier échantillon (t_0) et qui refusent de répondre au cours des années suivantes (t_{0+n}).

- ! Les membres du panel présents dans l'échantillon annuel : ce sont les personnes sélectionnées dans le premier échantillon (t_0) et qui sont présentes dans l'échantillon observé au cours des années suivantes (t_{0+n}). Ces personnes peuvent être absentes ou refuser de répondre à une ou plusieurs enquêtes. Leur poids individuel peut être recalculé chaque fois qu'elles rejoignent l'échantillon.

- ! Les nouveaux membres "légitimes" : ce sont des enfants nés simultanément dans la population et dans l'échantillon (entre (t_{0+n}) et ($t_{0+(n-1)}$)). Parmi ces enfants, seuls les descendants directs de membres du panel sont admis au titre de membres "légitimes". Il suffit que l'un des deux ascendants soit présent dans l'échantillon et membre du panel.

Ces nouveaux membres légitimes reçoivent un poids égal au poids moyen de leurs ascendants directs présents dans l'échantillon au moment où ils font leur entrée (différents cas peuvent se présenter).

- ! Les nouveaux membres "non-légitimes" présents dans l'échantillon : ces personnes entrent dans l'échantillon au temps (t_{0+n}). La plupart d'entre elles pouvaient appartenir à la population-cible au moment où l'échantillon initial a été sélectionné (t_0). Elles auraient pu faire partie de cette sélection mais

ce ne fut pas le cas. Elles entrent dans des ménages de l'échantillon au hasard des circonstances. Il est donc malaisé de leur attribuer une probabilité de sélection.

Elles reçoivent un poids de "0", de telle sorte qu'elles ne seront jamais prises en compte dans les analyses synchroniques de l'échantillon des individus.

- ! Les nouveaux membres "non-légitimes" quittant l'échantillon : ces personnes contribuaient à définir les caractéristiques des ménages lorsqu'elles étaient présentes dans l'échantillon. Lorsqu'elles le quittent, elles sont rangées dans cette catégorie parce qu'elles ne sont plus prises en compte, ni dans le calcul annuel des taux de réponse individuels ni dans la définition des caractéristiques des ménages.

- ! Les émigrés et les personnes décédées forment la sixième catégorie : ces personnes reflètent des mouvements démographiques généraux. Elles ne doivent pas être prises en compte dans le calcul des taux de réponse.

Seules les deux premières catégories interviennent dans le calcul des pondérations individuelles.

2.3. Calcul des pondérations

Chaque année, l'échantillon initial observé en (t_0) est partitionné en sous-catégories correspondant à des taux de réponses particulièrement contrastés. Ces taux de réponse sont calculés par rapport à l'échantillon initial et non par rapport à l'échantillon de l'année précédente.

Le poids de chaque membre présent dans l'échantillon correspond à l'inverse du taux de réponse de la sous-catégorie de population à laquelle il appartient. Ce poids obtenu au temps t_{0+n} est corrigé par le poids individuel initial au temps t_0 .

Les nouveaux membres "non-légitimes" reçoivent un poids de "0".

Les nouveaux membres légitimes reçoivent un poids égal à la moyenne des poids de leurs ascendants directs (les poids "0" sont pris en compte).

Les ménages reçoivent un poids égal au poids moyen de leurs membres.

2.4. Implications

Ce procédé implique certaines conséquences.

1. Les poids des membres d'un ménage peuvent être différents les uns des autres.

Chaque membre a suivi, suit, ou suivra une trajectoire autonome qui ne correspond pas nécessairement à celle des autres membres du ménage auquel il appartient à un moment donné. Il est donc assez logique que les membres d'un ménage puissent recevoir des poids différents à un moment donné.

Cette situation présente un inconvénient : des individus appartenant à un même ménage ne pèsent pas d'un même poids dans l'estimation des valeurs d'une caractéristique. Certains membres "représentent" plus le ménage que d'autres. L'uniformité des poids conférerait sans doute plus de cohérence aux estimations des valeurs individuelles et des ménages.

2. Les poids des membres d'un ménage peuvent être différents du poids du ménage.

Chaque membre suit une trajectoire qui ne se résout pas entièrement dans la trajectoire du ménage. Il peut refuser individuellement de répondre à l'enquête, ou appartenir en tant qu'individu à une sous-catégorie de l'échantillon qui reste particulièrement bien représentée, ou quitter le ménage et fonder son propre ménage ou, encore, se joindre à des nouveaux membres "non-légitimes" pour former un nouveau ménage.

Envisageons le cas d'un membre du panel appartenant à un groupe de personnes particulièrement peu enclin à participer aux enquêtes.

Il reçoit un poids individuel très élevé parce qu'il doit représenter un grand nombre de personnes qui lui ressemblent.

Cette personne rejoint un ménage dont les membres n'appartiennent pas à l'échantillon initial. Le poids de ce nouveau ménage est inversement proportionnel au nombre de membres "non-légitimes" qu'il contient. Ces membres "non-légitimes" ayant reçu un poids individuel nul ("0"), seul le poids du membre légitime du panel est divisé par le nombre total de personnes que contient le nouveau ménage.

Plus le ménage contient des membres "non-légitimes", plus le poids de ce ménage est faible, mais ce poids n'affecte en rien le poids individuel du membre légitime du panel qui se trouve dans ce ménage : il garde son poids individuel élevé.

Le nombre de membres "non-légitimes" présents dans le ménage influence le poids du ménage mais ne modifie pas le poids individuel de la personne membre légitime du panel.

Si le poids du ménage était ensuite redistribué à tous les membres du ménage, le membre légitime du panel recevrait un poids individuel réduit par rapport à son poids initial. Cette opération a été effectuée la première année parce que tous les membres du ménage appartenaient à la population-cible au moment du tirage de l'échantillon et parce que tous participaient à l'enquête.

Pour effectuer la même opération au cours des années suivantes, il faut accepter certaines hypothèses dont le paragraphe suivant souligne l'importance.

3. Le départ d'un membre "non-légitime" n'a pas d'effet sur le poids des autres membres du ménage.

Lorsqu'un nouveau membre "non-légitime" quitte le ménage d'accueil, le panel continue à le suivre. Il conserve son poids individuel initial ("0") et il forme un nouveau ménage dont le poids reste égal à "0".

Le départ de ce membre a généralement pour effet d'accroître le poids du ménage d'accueil ("ceteris paribus") : les poids individuels sont divisés par un nombre inférieur de membres ayant un poids nul.

Par contre, ce départ reste sans effet sur le poids individuel des autres membres de ce ménage. Ceci signifie que le poids nul affecté aux nouveaux membres "non-légitimes" dispense d'émettre toute hypothèse concernant le fait qu'ils "ressemblent" ou non aux autres membres du ménage d'accueil.

Si le départ d'un membre "non-légitime" devait accroître le poids des membres légitimes appartenant au même ménage, cela signifierait que ces autres membres "compensent" le départ du membre "non-légitime". Tout se passerait "comme si" les membres légitimes appartenant au ménage se trouvaient sous-représentés suite au départ d'une personne qui leur ressemble au moins sur certains traits.

4. Le rapport entre le poids du ménage et le poids de ses membres reste constant : par construction, le poids du ménage est toujours égal au poids moyen de ses membres.

Le ménage est une entité composite dont le contenu peut varier; la probabilité qu'il appartienne à l'échantillon dépend de la probabilité que ses membres restent dans l'échantillon et dans le ménage. Le ménage n'a pas d'existence autonome, il évolue en fonction des membres qui le composent. Le poids du ménage n'est donc que la résultante du poids des individus qui le composent.

L'évolution de l'échantillon ne modifie pas cette propriété.

2.5. Inconvénient de ce procédé

Ce procédé de pondération présente un inconvénient majeur.

Il écarte systématiquement les nouveaux membres "non-légitimes" des analyses synchroniques au plan individuel. Il entrave l'adaptation de l'échantillon initial aux phénomènes démographiques dont les nouveaux membres "non-légitimes" portent précisément la trace.

Une grande quantité d'informations reste inexploitée.

Cette procédure de pondération est conçue en vue de résoudre les problèmes de représentativité de l'échantillon sur le plan strictement synchronique. Elle ne prend pas en compte les conséquences de l'observation diachronique de l'échantillon et elle ne permet pas de retarder les effets de la détérioration de l'échantillon des individus.

Les simulations qui suivent illustrent les possibilités de résoudre ce problème en appliquant deux procédés différents.

3. LE "PARTAGE DES POIDS" (ERNST)

3.1. Principes

L'approche décrite par Ernst (1989) et testée par Lavallée et Hunter (1992) correspond à "un partage des poids" des individus légitimes avec les membres "non légitimes" ("a share approach"(Lavallée et Hunter)). Elle a pour effet d'attribuer un poids différent de "0" aux nouveaux membres "non-légitimes" et de permettre ainsi qu'ils soient pris en compte dans l'analyse synchronique des caractéristiques individuelles.

Dans un premier temps,

- ! le poids de chaque membre légitime du panel correspond à l'inverse de sa probabilité de sélection, et
- ! le poids des nouveaux membres "non-légitimes" est fixé à "0".

Le poids du ménage correspond à la moyenne des poids de ses membres.

Dans un second temps, le poids du ménage est redistribué à chaque membre du ménage.

Ce dernier élément est le seul qui distingue le modèle de Ernst et le modèle "classique". Mais il entraîne des conséquences importantes.

3.2. Implications

Les deux premières propriétés de cette procédure ont été signalées par Lavallée et Hunter.

1. On peut montrer que l'estimateur est sans biais (parce que le poids initial des membres "non-légitimes" est et reste égal à "0") (cf. Annexe 2).

Cette opération reproduit chaque année la procédure adoptée la première année. En 1985, les titulaires de revenus sont des "individus repères". Ils permettent de calculer le poids du ménage bien que la probabilité de sélection des autres membres interrogés soit inconnue. Le poids du ménage est ensuite attribué à tous les membres du ménage. En effet, les membres sont sélectionnés de manière exhaustive dans chaque ménage.

Au cours des années suivantes, les membres légitimes du ménage sont les "individus repères". Leur probabilité de sélection étant connue, ils permettent de calculer le poids du ménage. Le poids du ménage peut être ensuite attribué à tous les membres du ménage, bien que la probabilité de sélection initiale des membres "non-légitimes" soit inconnue. En effet, tous les membres sont observés dans chaque ménage.

2. Tous les individus appartenant à un ménage ont le même poids individuel.

Tous les membres représentent le ménage au même titre. Cette uniformité confère une cohérence interne à l'estimation des valeurs individuelles et des valeurs des ménages ("consistency").

3. Cette propriété ne contredit pas fondamentalement le principe de l'autonomie des trajectoires individuelles par rapport aux autres membres du ménage.

Chaque année, le poids initial de chaque membre du ménage reflète sa trajectoire individuelle dans le panel : il prend en compte son statut - membre du panel - membre "non-légitime" - descendant direct d'un membre "légitime" du panel - taux de réponse du sous-échantillon auquel il appartient en tant qu'individu.

Ce poids peut donc être différent du poids des autres membres du ménage (cf. section 2.4).

Lorsqu'une personne quitte un ménage, s'installe seule ou rejoint un autre ménage, elle prend une trajectoire différente des autres membres du ménage qu'elle quitte. Elle emporte son poids initial et non le poids du ménage qu'elle quitte. Son poids initial et son poids final sont à nouveau indépendants des poids des membres du ménage qu'elle quitte.

Dans le cas où elle s'installe seule, son poids initial et son poids final sont identiques.

Dans le cas où elle fonde ou rejoint un autre ménage, son poids initial ne s'en trouve pas modifié. Seul son poids final prendra en compte la composition de ce nouveau ménage.

4. L'autonomie de la trajectoire individuelle par rapport au ménage est également respectée.

Selon le modèle classique, le poids individuel des membres légitimes n'est pas influencé par la présence de membres "non-légitimes" au sein du ménage.

Par contre, dans le cadre du modèle présenté par Ernst, le poids du ménage est plus faible lorsque le ménage contient un grand nombre de membres "non-légitimes" puisque le poids des membres légitimes doit être divisé par un plus grand nombre de membres ayant un poids nul. Il s'en suit logiquement que le poids final des membres légitimes s'en trouve diminué : ils héritent du poids du ménage.

Mais ce poids final des individus est strictement lié à la constitution, plus ou moins provisoire, du ménage auquel ils appartiennent.

Cette conséquence du modèle présenté par Ernst repose sur un postulat.

Le passage du poids individuel initial au poids individuel final par l'intermédiaire du poids du ménage fonctionne comme si les nouveaux membres "non-légitimes" ressemblaient au(x) membre(s) légitime(s) qui composent le ménage. À certains égards, ces nouveaux membres présenteraient certains traits communs avec les membres légitimes appartenant au ménage qui les accueille bien qu'à d'autres égards, ils en soient différents.

Les membres "non-légitimes" apportent, bien entendu, des informations originales et le partage des poids leur permet précisément de manifester cet enrichissement. Mais l'opération du partage des poids entraîne également des effets que l'on ne peut admettre qu'à la condition d'admettre simultanément une ressemblance partielle entre les membres légitimes et les membres "non-légitimes" qui se joignent à eux au sein d'un même ménage.

Ce postulat se manifeste de la manière suivante.

Normalement, les personnes qui représentent une catégorie de population bien représentée ou sur-représentée au sein d'un échantillon reçoivent un poids relativement faible. C'est le principe fondamental de toute procédure de pondération.

Or, le partage du poids des membres légitimes avec les membres "non-légitimes" appartenant au même ménage réduit le poids individuel final des membres légitimes. Il faut donc admettre que c'est bien la présence de ces membres "non-légitimes" dans le ménage qui permet aux membres légitimes de représenter un plus petit nombre de personnes avec lesquelles elles partagent certains traits communs.

5. Un nouveau membre "non-légitime" est entré dans le panel parce qu'il est entré dans un ménage donné. Son existence dans le panel est et restera totalement dépendante du ménage dans lequel il se trouve.

Il est impossible de calculer la probabilité de sélection initiale des membres "non-légitimes". Ils reçoivent donc un poids initial nul ("0").

Leur poids final est le poids du ménage d'accueil. S'ils quittent ce ménage et s'installent isolément, ou dans un ménage composé exclusivement de membres "non-légitimes", leur poids initial est nul, le poids du ménage est nul et leur poids final est nul.

6. Le modèle présenté par Ernst et le modèle classique écartent automatiquement des analyses synchroniques, les individus et les ménages constitués uniquement de nouveaux membres "non-légitimes" (leur poids est nul).

En résumé :

1. L'originalité du modèle présenté par Ernst tient donc essentiellement au passage d'un poids individuel initial vers un poids individuel final par l'intermédiaire du poids du ménage.
2. Cette caractéristique du modèle permet de prendre en compte l'originalité de l'information apportée par les nouveaux membres mais elle introduit simultanément un présupposé : les membres "non-légitimes" présentent des traits communs avec les membres légitimes de leur ménage d'accueil.
3. Toutefois ce présupposé a une portée plus étroite que le postulat sous-jacent au modèle de Rao.

3.3. Simulation

Le modèle classique et le modèle présenté par Ernst calculent le poids du ménage de la même manière. La simulation peut donc se limiter à une comparaison des valeurs estimées au niveau individuel.

Vingt variables ont été sélectionnées. Les huit premières décrivent des caractéristiques individuelles; les douze dernières décrivent des caractéristiques du ménage et du chef du ménage auquel l'individu appartient. Ces caractéristiques du ménage et du chef de ménage sont attribuées à l'individu. Elles décrivent le contexte familial dans lequel il est situé.

Caractéristiques individuelles	Caractéristiques du ménage	Caractéristiques du chef de ménage
=====	=====	
sexe	nombre de personnes	sexe
âge	nombre d'enfants	âge
âge par sexe	nombre d'adultes	état civil
état civil	nombre d'emplois	emploi
nationalité	type d'habitation	état civil selon le
adulte/enfant	canton de résidence	sexe
lien avec le chef de ménage		emploi selon le sexe
position sur le marché de l'emploi		

Le tableau 1 compare les estimations de ces variables sous chaque modèle de pondération.

La première colonne présente les distributions sous le modèle de pondération en vigueur actuellement dans le PSELL.

Il eut été préférable de comparer les estimations obtenues sous chaque procédure et les données "vraies", observables dans la population. Ces données ne sont pas disponibles. Les données du recensement permettront de procéder à ces tests sur l'échantillon observé en 1991.

Les estimations obtenues sous le système de pondération classique constituent les seules valeurs repères disponibles. Nous admettons qu'elle sont fiables, à ceci près

qu'elles ne prennent pas suffisamment en compte les modifications démographiques de la population de référence. Une procédure alternative prenant en compte ces modifications peut donc être préférable, si elle fournit des estimations sans biais, équivalentes (ou identiques) aux estimations fournies par la procédure classique.

La deuxième colonne présente les estimations après pondération selon le modèle présenté par Ernst.

La prise en compte des membres "non-légitimes" (après pondération) accroît la taille de l'échantillon de 11,6%.

En 1990, 92 personnes appartiennent à des ménages constitués exclusivement de membres "non-légitimes". Leur poids final reste égal à "0". Elles sont automatiquement exclues de l'estimation des valeurs individuelles. La taille réelle de l'échantillon est donc 4641 personnes et non 4735 personnes.

La comparaison des résultats montre essentiellement deux choses.

1. L'écart entre les pourcentages est généralement inférieur à 1 point pour les caractéristiques individuelles : la prise en compte des membres "non-légitimes" aurait une légère tendance à réduire le poids relatif des enfants du chef de ménage au profit des autres positions par rapport au chef de ménage. En d'autres termes, les membres "non-légitimes" seraient plus souvent des adultes que des enfants.
2. Les caractéristiques du ménage et du chef de ménage, définies comme caractéristiques de l'environnement individuel, présentent des distributions rigoureusement identiques.

TABLEAU 1

Comparaison des estimations selon le mode de pondération en 1990

(Classique et Partage du poids du ménage)

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Caractéristiques individuelles

Sexe

1. masculin	48.1	48.4
2. féminin	51.9	51.6

Âges

1. < 18 ans	22.4	21.8
2. 18 - 24 ans	10.7	10.9
3. 25 - 34 ans	14.8	14.9
4. 35 - 44 ans	14.5	14.7
5. 45 - 54 ans	13.1	13.2
6. 55 - 64 ans	10.8	10.9
7. > 64 ans	13.7	13.6

Âges par sexe

1. homme < 18 ans	11.1	11.1
2. homme 18 - 64 ans	32.0	32.4
3. homme > 64 ans	5.0	4.9
1. femme < 18 ans	11.2	10.8
2. femme 18 - 64 ans	32.0	32.1
3. femme > 64 ans	8.7	8.7

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

État civil

1. célibataire	40.3	39.7
2. marié	48.3	49.0
3. veuf	8.3	8.2
4. divorcé	2.1	2.1
5. séparé	1.0	1.0

Nationalité

1. Luxembourgeois	79.7	79.5
2. C.E.E.	18.5	18.7
3. hors C.E.E.	1.8	1.8

Adulte/ enfant

1. adulte	74.9	75.4
2. enfant	25.1	24.6

Lien avec le chef de ménage

1. chef du ménage	37.4	37.4
2. épouse	23.9	24.1
3. concubin	0.9	1.1
4. ami(e)	0.1	0.3
5. fils/fille	33.8	32.7
6. autres	3.9	4.4

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Position sur le marché du travail

1. enfant	25.1	24.6
2. pension: invalidité	2.9	2.9
3. pension: retraite	9.2	9.2
4. pension: survie	5.7	5.6
5. travaille ou maladie	39.4	39.5
6. demandeur d'emploi	0.6	0.6
7. ménagère et autres	17.1	17.6

Caractéristiques du ménage

Nombre de personnes dans le ménage

1 personne	8.9	8.9
2 personnes	20.1	20.1
3 personnes	23.6	23.6
4 personnes	27.7	27.7
5 personnes	12.7	12.7
6 personnes	5.0	5.0
7 personnes	1.3	1.3
8 personnes	0.7	0.7

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Nombre d'enfants dans le ménage

pas d'enfant	43.8	43.8
1 enfant	22.3	22.3
2 enfants	22.4	22.4
3 enfants	8.5	8.5
4 enfants	2.7	2.7
5 enfants	0.3	0.3

Nombre d'adultes dans le ménage

1 adulte	11.4	11.4
2 adultes	59.6	59.6
3 adultes	17.4	17.4
4 adultes	8.2	8.2
5 adultes	2.2	2.2
6 adultes	0.7	0.7
7 adultes	0.5	0.5
8 adultes	0.0	0.0

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Nombre de personnes ayant un emploi dans le ménage

pas d'emploi	18.2	18.2
1 emploi	43.1	43.1
2 emplois	28.5	28.5
3 emplois	7.1	7.1
4 emplois	2.5	2.5
5 emplois	0.3	0.3
6 emplois	0.3	0.3

Type d'habitation

1. maison rurale	7.7	7.7
2. maison unifamiliale individuelle	34.0	34.0
3. maison unifamiliale jumelée	10.3	10.3
4. maison unifamiliale en série	25.3	25.3
5. 2 à 4 logements	13.4	13.3
6. 5 à 19 logements	6.3	6.3
7. plus de 20 logements	0.7	0.7
8. habitat précaire	0.1	0.1
9. maison de retraite	0.2	0.2
10. N.R.	2.0	2.1

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Canton

1. Luxembourg - ville	15.4	15.4
2. Capellen	7.7	7.7
3. Esch sur Alzette	33.7	33.7
4. Luxembourg campagne	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

Caractéristiques du chef de ménage

Sexe du chef de ménage

1. masculin	85.9	85.9
2. féminin	14.1	14.1

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Âge du chef de ménage

1. < 24 ans	1.7	1.7
2. 25 - 34 ans	16.2	16.2
3. 35 - 44 ans	26.5	26.5
4. 45 - 54 ans	23.8	23.8
5. 55 - 64 ans	17.1	17.1
6. > 64 ans	14.7	14.7

État civil du chef de ménage

1. célibataire	6.4	6.4
2. marié	79.1	79.1
3. séparé	1.3	1.3
4. divorcé	3.1	3.1
5. veuf	10.1	10.1

Emploi du chef de ménage

1. sans emploi	29.4	29.4
2. emploi	70.6	70.6

État civil du chef de ménage selon le sexe

1. homme non marié	7.3	7.3
2. homme marié	78.7	78.7
3. femme non mariée	13.6	13.6
4. femme mariée	0.4	0.4

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4159)	(n=4641)

Emploi du chef de ménage selon le sexe

1. homme sans emploi	20.3	20.3
2. homme avec emploi	65.6	65.6
3. femme sans emploi	9.1	9.1
4. femme avec emploi	5.0	5.0

4. "L'ESTIMATEUR COMPOSITE" (RAO)

4.1. Principes

Lavallée et Hunter décrivent le modèle proposé par J.N.K. Rao de la manière suivante :

- ! Il se focalise en priorité sur la formulation de la procédure de calcul des poids plutôt que sur la valeur proprement dite des poids des individus sélectionnés (sur l'estimateur, plutôt que sur les valeurs estimées). Il s'agit de construire un modèle de l'estimateur des poids des membres "non-légitimes".

- ! Ce modèle part du principe en vertu duquel les membres légitimes (sélectionnés en 1985) et les membres "non-légitimes" (entrés dans les échantillons des années suivantes) forment deux sous-échantillons distincts tirés d'une même population.

- ! Chaque sous-échantillon fournit une estimation des valeurs de la population. Ceci implique que les estimations du sous-groupe des membres "non-légitimes" reposent sur un postulat : les membres "non-légitimes" sont traités comme si ils avaient été sélectionnés selon un échantillonnage aléatoire stratifié.

- ! Ce détour permet
 - ! de calculer les probabilités de sélection des membres dans les deux sous-groupes

 - ! de composer un estimateur unique des poids, qui réunit le calcul des poids dans chacun des deux sous-groupes

 - ! tout en respectant l'importance relative de chaque sous-groupe au sein de l'échantillon total.

- ! Le poids de chaque membre légitime du panel correspond à l'inverse de sa probabilité de sélection (modèle classique, modèle de Ernst).

- ! Les membres "non-légitimes" sont classés par rapport à une variable de référence (par exemple : "les cantons de résidence"). Ces membres sont sur-représentés dans certains cantons et sous-représentés dans d'autres cantons. Ces taux de représentation définissent les probabilités de sélection des membres "non-légitimes".

(En l'absence de données relatives à la population-cible, en 1990, l'échantillon des membres légitimes fait office de population de référence).

Les membres "non-légitimes" sont pondérés en raison inverse de leur probabilité de sélection.

- ! L'estimateur final des poids, applicable à l'échantillon total, est composé des deux facteurs de pondération correspondant respectivement au sous-échantillon des membres légitimes et au sous-échantillon des membres "non-légitimes". Les poids des uns et des autres sont pris en compte proportionnellement à leur importance relative dans l'échantillon total.

- ! La moyenne des poids des membres du ménage définit le poids du ménage. Cette phase est commune aux trois modèles.

- ! Le poids du ménage peut être ensuite redistribué à chaque membre (modèle de Ernst), mais cette dernière étape n'est pas obligatoire. Les poids individuels pourraient être utilisés en tant que tels puisque les membres "non-légitimes" sont dotés d'un poids individuel initial différent de "0".

L'attribution du poids du ménage à ses membres a la propriété souhaitable de conférer une cohérence interne aux estimations des valeurs individuelles et des valeurs des ménages. Tous les membres du ménage ont le même poids et représentent le ménage au même titre.

Cette opération est fondamentale dans le modèle de Ernst, puisqu'elle établit le vecteur des poids des membres "non-légitimes". Dans le cadre du modèle de Rao, cette opération pourrait être omise puisque les membres "non-légitimes" ont reçu un poids initial indépendant du poids du ménage.

Cette distinction entre les deux modèles résulte du fait qu'ils sont fondés sur des présupposés différents qui entraînent des implications très différentes.

4.2. Implications

Les deux premières propriétés de ce modèle ont été signalées par Lavallée et Hunter.

1. L'estimateur composite comporte un biais. Ce biais provient de l'estimateur des poids des membres "non-légitimes". En effet, ces membres ne forment pas un échantillon probabiliste, ils ne sont pas tirés aléatoirement et la variable de stratification est choisie arbitrairement. L'importance de ce biais s'élève parallèlement à la taille relative de ce sous-échantillon dans l'échantillon total.
2. Tous les membres d'un ménage peuvent avoir le même poids individuel, à condition que le poids du ménage soit redistribué à tous les membres. Cette uniformité des poids permet de restaurer la cohérence des estimations des valeurs individuelles et des valeurs des ménages.
3. Cette propriété ne contredit pas fondamentalement le principe de l'autonomie des trajectoires individuelles par rapport aux trajectoires des autres membres du ménage.

! Si le poids du ménage n'est pas redistribué à ses membres, chaque individu (membre légitime ou non) conserve son poids annuel initial.

Chaque année, le poids initial de chaque membre du ménage reflète sa trajectoire individuelle dans le panel : il prend en compte son statut - membre du panel - membre "non-légitime" - descendant direct d'un membre "légitime" du panel - taux de réponse annuel de la sous-catégorie de population à laquelle il appartient en tant que membre légitime du panel - poids de la strate à laquelle il appartient en tant que membre "non-légitime".

Les poids des membres du ménage sont d'autant plus hétérogènes que leurs origines et leurs trajectoires sont plus hétéroclites.

- ! Si le poids du ménage est redistribué à chacun des membres, le poids annuel initial d'un membre reste indépendant du poids des autres membres du ménage.

Lorsqu'une personne quitte un ménage, s'installe seule ou rejoint un autre ménage, elle prend une trajectoire différente des trajectoires des membres du ménage qu'elle quitte. Son poids initial et son poids final sont à nouveau indépendants des poids des membres du ménage qu'elle quitte.

Si elle s'installe seule, son poids initial et son poids final sont identiques.

Si elle fonde ou rejoint un autre ménage, son poids initial ne s'en trouve pas modifié, mais son poids final, en tant que membre de ce nouveau ménage, prend en compte la nouvelle composition de ce ménage.

4. L'autonomie de la trajectoire individuelle par rapport au ménage est également respectée. Lorsqu'un individu quitte le ménage, il n'emporte que son poids individuel initial.

Le calcul du poids du ménage s'effectue de la même manière dans le modèle de Ernst et dans le modèle "classique", mais il existe une différence importante entre le modèle de Rao et le modèle de Ernst.

Dans le cadre du modèle présenté par Ernst, le poids individuel final des membres du ménage sera plus faible lorsque le ménage contient un plus grand nombre de membres "non-légitimes" puisque seul le poids des membres légitimes devra être divisé par un plus grand nombre de membres (cf. Section 3.2). La répartition des poids s'effectue au sein de chaque ménage selon la configuration propre à ce ménage.

Ce ne sera pas le cas, dans le cadre du modèle de Rao. Les membres légitimes ne sont plus nécessairement amenés à représenter un nombre inférieur d'individus qui leur ressemblent.

La régulation des poids des membres du ménage n'obéit plus à une logique interne au ménage. Elle est soumise à une logique externe : elle est commandée par la variable qui a été choisie pour stratifier et pondérer le sous-échantillon des membres "non-légitimes".

Le postulat de Ernst n'est plus applicable : les nouveaux membres "non-légitimes" permettent de mieux cerner les modifications qui se sont produites dans les ménages d'accueil sans présupposer qu'ils "ressemblent" en aucune manière aux membres du ménage qu'il rejoignent.

Le postulat de Rao a une portée plus générale.

Les nouveaux membres "non-légitimes" sont pondérés par rapport à une population-cible (une caractéristique quelconque de cette population). On présume donc qu'ils représentent cette population. Or, cette population-cible est également la population que les membres "légitimes" sont censés représenter. Ceci revient à supposer que les deux sous-échantillons, tirés d'une population-mère commune, se ressemblent (moyennant un terme d'erreur).

On peut montrer que cette hypothèse est trop générale. La ressemblance entre les deux sous-échantillons dépend des caractéristiques envisagées.

Si les deux sous-échantillons se ressemblent, ils doivent présenter des distributions parallèles sur les variables de comparaison.

Dans le cas contraire, l'échantillon des membres "non-légitimes" doit sur-représenter ou sous-représenter certaines catégories de population et deux solutions peuvent être envisagées :

- ! soit, il faut admettre que les deux sous-échantillons ne se ressemblent pas, refuser le postulat de départ et renoncer à insérer les membres "non-légitimes" dans l'échantillon total,

- ! soit, il faut corriger ces effets de sur-représentation en pondérant le sous-échantillon des membres "non-légitimes", annuler ses particularités et renoncer ainsi à l'apport des nouvelles informations qui sont précisément censées régénérer l'échantillon initial. (On peut s'interroger sur la signification d'une telle pondération dans le cadre d'un échantillon non-probabiliste).

Les deux sous-échantillons se ressemblent effectivement sur certains points : le sexe, le sexe du chef de ménage et le canton de résidence sont les trois critères qui assurent la plus grande ressemblance entre les deux sous-échantillons (cf. Tableau 3).

Les trois variables qui traduisent les plus fortes divergences entre les deux sous-échantillons sont l'âge des individus, l'âge du chef de ménage et la nature du lien avec le chef de ménage (cf. Tableau 3). Les nouveaux membres sont sur-représentés parmi :

- ! les personnes âgées de 25 à 34 ans,
- ! les concubin(e)s, les ami(e)s, et les autres relations du chef de ménage au détriment des enfants du chef de ménage,
- ! les membres qui appartiennent à des ménages dont le chef de ménage a moins de 34 ans.

Les deux sous-échantillons ne se ressemblent pas systématiquement. Les autres variables présentent des contrastes plus ou moins accentués entre les deux sous-échantillons. L'intégration des membres "non-légitimes" suppose donc que leur échantillon soit redressé afin d'annuler les zones de sur (sous)-représentation.

Avant d'envisager les effets de cette procédure de pondération de l'échantillon des membres "non-légitimes", il reste à souligner trois points.

5. Le modèle classique et le modèle de Ernst avaient un point commun : les nouveaux membres "non-légitimes" n'ont d'existence dans le panel qu'en raison de leur appartenance à un ménage. Leur poids individuel initial est

nul. Lorsqu'ils quittent le ménage, s'installent seuls ou rejoignent un ménage formé exclusivement par des membres "non-légitimes", leur poids individuel est nul, le poids du ménage est nul et leur poids final est nul.

Le modèle de Rao se distingue des deux premiers sur trois points.

- ! Le poids initial des membres "non-légitimes" n'est plus "0". Un poids initial leur est attribué en fonction de leur position par rapport à la variable de stratification.

Ceci implique que les membres "non-légitimes" ont une existence dans le panel qui est indépendante du ménage auquel ils appartiennent. Lorsqu'ils quittent le ménage d'accueil et s'installent seuls, ils continuent à bénéficier d'un poids individuel initial lié à leur position par rapport à la variable de stratification de l'échantillon des membres "non-légitimes".

- ! Les nouveaux membres "non-légitimes" reçoivent a priori un poids initial plus élevé dans la cadre du modèle de Rao que dans le cadre du modèle de Ernst .

Les nouveaux membres sont déjà affectés d'un poids supérieur à "0" lorsqu'ils entrent dans le calcul du poids du ménage, : les ménages qui accueillent de nouveaux membres reçoivent donc a priori un poids plus élevé dans le cadre de ce modèle.

- ! Les deux premiers modèles excluent automatiquement tout membre "non-légitime" appartenant à un ménage formé exclusivement par des membres "non-légitimes". Ces membres n'interviennent donc pas dans l'analyse synchronique des caractéristiques individuelles.

Le modèle de Rao n'exclut pas automatiquement ces membres. Ils ont un poids initial différent de "0". Ils peuvent être pris en compte dans l'analyse synchronique des caractéristiques individuelles, même s'ils appartiennent à un ménage composé exclusivement de membres 'non-légitimes'.

4.3. Simulation

Le tableau 2 reprend les vingt variables utilisées dans la première simulation et les compare aux estimations obtenues sous le modèle de pondération classique.

La première colonne présente les distributions sous le modèle de pondération "classique" en vigueur actuellement dans le PSELL.

La deuxième colonne présente les distributions sous une première version du modèle de "l'estimateur composite" de Rao. Conformément aux suggestions de l'auteur, la variable de stratification de l'échantillon des membres "non-légitimes" présente un caractère très général : les cantons de résidence.

Les deux sous-échantillons se répartissent de manière similaire dans les différents cantons (cf. Tableau 3). Par conséquent, l'effet de la pondération est très limité et les points de divergence entre les deux sous-échantillons restent entièrement perceptibles.

Le choix de cette variable a le mérite de ne pas annuler l'originalité des caractéristiques des nouveaux membres. Elle permet d'attribuer un poids initial aux membres "non-légitimes". Mais on discerne mal l'intérêt d'une pondération qui n'a guère d'effets sur le profil de l'échantillon (outre le fait qu'on peut s'interroger sur l'opportunité de pondérer un échantillon non-probabiliste).

La troisième colonne propose une alternative. L'échantillon des membres "non-légitimes" est pondéré en fonction des deux variables qui présentent les contrastes les plus significatifs entre les deux sous-échantillons : l'âge du chef de ménage et le lien avec le chef de ménage.

L'objectif de cette opération a le mérite d'être clair : redresser l'échantillon des membres "non-légitimes" afin qu'il ressemble autant que possible à l'échantillon des membres légitimes. La ressemblance entre les deux sous-échantillons est mieux assurée ... au détriment de l'originalité des informations apportées par les nouveaux membres.

La quatrième colonne présente également les distributions après stratification de l'échantillon des membres "non-légitimes" selon l'âge du chef de ménage et selon le lien avec le chef de ménage. Une précaution supplémentaire permet d'obtenir une meilleure comparaison.

Les membres "non-légitimes" ayant quitté le ménage d'accueil et se trouvant dans un ménage formé uniquement par des membres "non-légitimes" sont exclus volontairement, bien que le modèle leur attribue un poids.

La comparaison des résultats montre essentiellement quatre choses.

1. Les différentes méthodes conduisent à des résultats très semblables.
2. Lorsque l'échantillon des membres "non-légitimes" est stratifié par cantons, 23 modalités présentent des écarts égaux ou supérieurs à 1 point (18 écarts compris entre 1 et 2 points; 5 écarts supérieurs à 2 points). Ces écarts sont probablement négligeables sur le plan statistique, mais ils sont plus nombreux sous ce modèle que sous les autres.

3. Lorsque l'échantillon des membres "non-légitimes" est stratifié par le lien avec le chef de ménage et par l'âge du chef de ménage, 11 modalités présentent des écarts égaux ou supérieurs à 1 point.

4. Lorsque l'échantillon des membres "non-légitimes" est stratifié par le lien avec le chef de ménage et par l'âge du chef de ménage, en excluant volontairement les membres qui n'ont plus aucun lien avec un membre légitime, 9 modalités présentent des écarts égaux ou supérieurs à 1 point.

TABLEAU 2

Comparaison des estimations selon le mode de pondération, en 1990

- ! Classique
- ! Poids composite A (stratification par cantons)
- ! Poids composite B (stratification par lien au chef de ménage et âges)
- ! Poids composite B (exclus : les ménages sans membre légitime)

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: ménages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Caractéristiques individuelles

Sexe

1. masculin	48.1	48.6	48.6	48.8
2. féminin	51.9	51.4	51.4	51.2

Âges

1. < 18 ans	22.4	21.7	21.9	21.8
2. 18 - 24 ans	10.7	11.9	11.3	11.2
3. 25 - 34 ans	14.8	16.9	15.4	15.5
4. 35 - 44 ans	14.5	14.3	14.9	14.7
5. 45 - 54 ans	13.1	12.4	12.9	13.2
6. 55 - 64 ans	10.8	10.2	10.6	10.5
7. > 64 ans	13.7	12.6	13.0	13.1

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Âges par sexe

1. homme < 18 ans	11.1	11.0	11.0	11.1
2. homme 18 - 64 ans	32.0	33.1	32.8	32.8
3. homme > 64 ans	5.0	4.6	4.8	4.9
1. femme < 18 ans	11.2	10.8	10.9	10.7
2. femme 18 - 64 ans	32.0	32.5	32.3	32.2
3. femme > 64 ans	8.7	8.0	8.2	8.3

Nationalité

1. Lux- embourgeois	79.7	79.2	79.5	79.7
2. C.E.E.	18.5	19.1	18.9	18.7
3. hors C.E.E.	1.8	1.7	1.6	1.6

État civil

1. célibataire	40.3	40.1	40.0	39.8
2. marié	48.3	48.6	48.2	48.8
3. veuf	8.3	7.8	8.1	7.9
4. divorcé	2.1	2.3	2.4	2.3
5. séparé	1.0	1.2	1.3	1.2

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Adulte/ enfant

1. adulte	74.9	75.8	75.5	75.6
2. enfant	25.1	24.2	24.5	24.4

Lien avec le chef de ménage

1. chef du ménage	37.4	37.0	36.8	36.6
2. épouse	23.9	23.8	23.7	23.9
3. concubin	0.9	1.7	1.4	1.4
4. ami(e)	0.1	0.5	0.4	0.3
5. fils/fille	33.8	31.5	32.1	32.2
6. autres	3.9	5.5	5.6	5.6

Position sur le marché du travail

1. enfant	25.1	24.2	24.5	24.4
2. pension: invalidité	2.9	2.8	2.9	2.8
3. pension: retraite	9.2	8.6	8.8	9.0
4. pension: survie	5.7	5.3	5.6	5.5
5. travaille ou maladie	39.4	41.6	40.6	40.4
6. demandeur d'emploi	0.6	0.7	0.7	0.7
7. ménagère et autres	17.1	16.9	16.9	17.2

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: ménages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Caractéristiques du ménage

Nombre de personnes dans le ménage

1 personne	8.9	8.1	8.3	8.0
2 personnes	20.1	21.0	19.7	20.0
3 personnes	23.6	24.2	24.2	23.9
4 personnes	27.7	26.8	27.2	27.3
5 personnes	12.7	12.1	12.2	12.2
6 personnes	5.0	5.4	5.7	5.8
7 personnes	1.3	1.7	2.0	2.1
8 personnes	0.7	0.7	0.7	0.7

Nombre d'enfants dans le ménage

pas d'enfant	43.8	43.9	43.4	43.5
1 enfant	22.3	23.6	23.4	23.4
2 enfants	22.4	21.2	21.6	21.4
3 enfants	8.5	8.2	8.4	8.4
4 enfants	2.7	2.6	2.5	2.6
5 enfants	0.3	0.5	0.7	0.7

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Nombre d'adultes dans le ménage

1 adulte	11.4	10.4	10.7	10.3
2 adultes	59.6	60.2	58.7	58.8
3 adultes	17.4	17.2	18.2	18.2
4 adultes	8.2	8.3	8.1	8.3
5 adultes	2.2	2.5	2.8	2.9
6 adultes	0.7	0.9	1.0	1.0
7 adultes	0.5	0.4	0.5	0.4
8 adultes	0.0	0.1	0.0	0.1

Nombre de personnes ayant un emploi dans le ménage

pas d'emploi	18.2	16.5	17.0	16.9
1 emploi	43.1	42.2	42.6	43.0
2 emplois	28.5	30.8	29.6	29.2
3 emplois	7.1	7.1	7.2	7.1
4 emplois	2.5	2.8	3.1	3.2
5 emplois	0.3	0.3	0.3	0.3
6 emplois	0.3	0.3	0.2	0.3

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Type d'habitation

1. maison rurale	7.7	7.4	7.3	7.5
2. maison unifamiliale individuelle	34.0	33.5	33.5	33.6
3. maison unifamiliale jumelée	10.3	10.2	10.3	10.5
4. maison unifamiliale en série	25.3	25.4	25.8	25.6
5. 2 à 4 logements	13.4	14.1	13.9	13.8
6. 5 à 19 logements	6.3	6.5	6.2	6.1
7. plus de 20 logements	0.7	0.7	0.7	0.7
8. habitat précaire	0.1	0.1	0.1	0.1
9. maison de retraite	0.2	0.1	0.1	0.1
10. N.R.	2.0	2.0	2.0	2.0

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Canton

1. Luxembourg - ville	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 campagne	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

Caractéristiques du chef de ménage

Sexe du chef de ménage

1. masculin	85.9	85.9	85.9	86.4
2. féminin	14.1	14.1	14.1	13.6

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
	Fréquences (%)			
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

Âge du chef de ménage

1. < 24 ans	1.7	2.8	2.0	2.0
2. 25 - 34 ans	16.2	19.2	16.7	16.6
3. 35 - 44 ans	26.5	25.9	27.5	26.9
4. 45 - 54 ans	23.8	22.5	23.4	24.0
5. 55 - 64 ans	17.1	16.2	16.6	16.6
6. > 64 ans	14.7	13.4	13.8	13.9

État civil du chef de ménage

1. célibataire	6.4	7.5	6.9	6.7
2. marié	79.1	77.7	77.9	78.8
3. séparé	1.3	1.6	1.5	1.4
4. divorcé	3.1	3.5	3.7	3.3
5. veuf	10.1	9.8	10.0	9.8

Emploi du chef de ménage

1. sans emploi	29.4	27.8	28.5	28.6
2. emploi	70.6	72.2	71.5	71.4

	Pondération classique Membres	Poids composite A (stratification par cantons)	Poids composite B (str.: lien C.M. âge du C.M.)	Poids composite B (exclus: mé- nages N.L.)
Fréquences (%)				
	(n=4159)	(n=4735)	(n=4735)	(n=4641)

État civil du chef de ménage selon le sexe

1. homme non marié	7.3	8.7	8.4	8.1
2. homme marié	78.7	77.2	77.5	78.3
3. femme non mariée	13.6	13.7	13.7	13.2
4. femme mariée	0.4	0.4	0.4	0.4

Emploi du chef de ménage selon le sexe

1. homme sans emploi	20.3	19.1	19.7	19.8
2. homme avec emploi	65.6	66.8	66.2	66.6
3. femme sans emploi	9.1	8.7	8.9	8.8
4. femme avec emploi	5.0	5.4	5.3	4.8

TABLEAU 3
COMPARAISON DES SOUS-ÉCHANTILLONS
SELON SIX VARIABLES

Comparaison des échantillons selon le canton de
résidence (fréquences en %)

Cantons	Échantillons	
	Membres légitimes	Membres non-légitimes
Luxembourg	15.4%	13.4%
Capellen	7.7%	5.2%
Esch/Alzette	33.7%	42.5%
Lux. campag.	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

Comparaison des échantillons selon le sexe des membres (fréquences en %)

Sexe	Échantillons	
	Membres légitimes	Membres non-légitimes
homme	48.1%	50.9%
femme	51.9%	49.1%
Total	100.0%	100.0%
N=	4159	576

Comparaison des échantillons selon le sexe du ch. de ménage (fréquences en %)

Sexe du chef de ménage	Échantillons	
	Membres légitimes	Membres non-légitimes
homme	85.9%	85.6%
femme	14.1%	14.4%
Total	100.0%	100.0%
N=	4159	576

Comparaison des échantillons selon les catégories d'âge (fréquences en %)

Catégories d'âge	Échantillons	
	Membres légitimes	Membres non-légitimes
< 18 ans	22.4%	16.8%
18 à 24 ans	10.7%	21.9%
25 à 34 ans	14.8%	35.1%
35 à 44 ans	14.5%	10.2%
45 à 54 ans	13.1%	5.9%
55 à 64 ans	10.8%	4.2%
> 64 ans	13.7%	5.9%
Total	100.0%	100.0%
N=	4159	576

Comparaison des échantillons selon le lien avec le chef de ménage (fréquences en %)

Liens avec le chef de mén.	Échantillons	
	Membres légitimes	Membres non-légitimes
chef de mén.	37.4%	32.1%
épouse	23.9%	21.7%
concubin(e)	0.9%	7.6%
ami(e)	0.1%	3.6%
fil(s)/fil(le)	33.8%	16.1%
autres	3.9%	18.9%
Total	100.0%	100.0%
N=	4159	576

Comparaison des échantillons selon l'âge du chef de ménage (fréquences en %)

Âge du chef de ménage	Échantillons	
	Membres légitimes	Membres non-légitimes
< 24 ans	1.6%	11.1%
25 à 34 ans	16.2%	40.3%
35 à 44 ans	26.5%	21.7%
45 à 54 ans	23.9%	12.7%
55 à 64 ans	17.1%	9.5%
> 64 ans	14.7%	4.7%
Total	100.0%	100.0%
N=	4159	576

5. CONCLUSION

La comparaison des résultats permet de constater trois choses essentielles.

1. Les différents modèles présentent des résultats très comparables : les écarts entre les distributions obtenues sous chaque modèle et les distributions obtenues sous le modèle en vigueur actuellement sont très limités.
2. Le modèle présenté par Ernst cumule trois avantages :
 - ! l'estimateur des poids est sans biais,
 - ! il présente systématiquement les résultats les plus conformes aux distributions obtenues sous le modèle classique,

- ! il repose sur un postulat dont la portée est plus limitée que le postulat sous-jacent au modèle de Rao. Les nouveaux membres "non-légitimes" permettent de mieux cerner les modifications qui se sont produites dans les ménages d'accueil, mais cet enrichissement présuppose que l'on admette simultanément que ces nouveaux membres partagent certains traits communs avec les membres du ménage d'accueil.

Ils ne ressemblent pas nécessairement à l'ensemble de l'échantillon dans lequel ils entrent et ils ne sont pas représentatifs de la population-cible prise dans son ensemble à un moment donné, mais ils rendent compte des modifications qui se sont produites dans les ménages qui les accueillent.

Ils n'ont pas d'existence individuelle indépendante du ménage qui les accueille.

La régulation des poids s'effectue au sein de chaque ménage. Elle ne s'effectue pas au niveau de l'ensemble de l'échantillon (Rao).

3. Ces conclusions rejoignent strictement les résultats obtenus par Lavallée et Hunter (1992). Ceci contribue à leur donner une certaine généralité.

On trouvera en Annexe 1, les résultats d'une comparaison effectuée sur l'échantillon observé en 1987, par le PSELL. Ils confirment également ces conclusions et les résultats obtenus sur l'échantillon observé en 1990.

BIBLIOGRAPHIE

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 n° 48, CEPS/Insead, Walferdange, 1992.

Lavallée P., Hunter L. (1992) "Weighting for the Survey of Labour and Income Dynamics", contribution au symposium "Conception et analyse des enquêtes longitudinales", Ottawa, 1992 (à paraître).

Rao J.N.K., papier non publié.

ANNEXE 1

Comparaison des estimations selon le mode de pondération en 1987

(Classique et Partage du poids du ménage)

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Caractéristiques individuelles

Sexe

1. masculin	48.2	48.2
2. féminin	51.8	51.8

Âges

1. < 18 ans	23.6	23.3
2. 18 - 24 ans	11.1	11.3
3. 25 - 34 ans	15.2	15.6
4. 35 - 44 ans	14.6	14.6
5. 45 - 54 ans	12.6	12.4
6. 55 - 64 ans	10.5	10.4
7. > 64 ans	12.4	12.4

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Âges par sexe

1. homme < 18 ans	12.0	11.8
2. homme 18 - 64 ans	31.5	31.6
3. homme > 64 ans	4.7	4.7
1. femme < 18 ans	11.6	11.5
2. femme 18 - 64 ans	32.5	32.7
3. femme > 64 ans	7.7	7.7

État civil

1. célibataire	40.0	40.0
2. marié	48.8	49.0
3. veuf	8.1	8.0
4. divorcé	2.3	2.2
5. séparé	0.8	0.8

Nationalité

1. Luxembourgeois	77.8	77.7
2. C.E.E.	20.3	20.4
3. hors C.E.E.	1.9	1.9

Adulte/ enfant

1. adulte	74.7	75.0
2. enfant	25.3	25.0

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Lien avec le chef de ménage

1. chef du ménage	36.6	36.4
2. épouse	23.9	24.1
3. concubin(e)	0.9	0.9
4. ami(e)	0.2	0.2
5. fils/fille	34.2	33.6
6. autres	4.2	4.8

Position sur le marché du travail

1. enfant	25.3	25.0
2. pension: invalidité	3.0	3.0
3. pension: retraite	7.8	7.7
4. pension: survie	5.5	5.4
5. travaille ou maladie	39.7	39.9
6. demandeur d'emploi	0.8	1.0
7. ménagère et autres	17.9	18.0

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Caractéristiques du ménage

Nombre de personnes dans le ménage

1 personne	8.0	8.0
2 personnes	19.6	19.6
3 personnes	22.8	22.8
4 personnes	28.7	28.7
5 personnes	12.9	12.9
6 personnes	4.9	4.9
7 personnes	2.1	2.1
8 personnes	0.7	0.7
9 personnes et plus	0.3	0.3

Nombre d'enfants dans le ménage

pas d'enfant	41.7	41.7
1 enfant	23.6	23.6
2 enfants	22.8	22.8
3 enfants	9.1	9.1
4 enfants	2.7	2.7
5 enfants	0.1	0.1

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Nombre d'adultes dans le ménage

1 adulte	10.4	10.4
2 adultes	58.4	58.4
3 adultes	18.9	18.9
4 adultes	7.8	7.8
5 adultes	2.8	2.8
6 adultes	1.4	1.4
7 adultes	0.2	0.2
9 adultes	0.1	0.1

Nombre de personnes ayant un emploi dans le ménage

pas d'emploi	17.4	17.4
1 emploi	42.2	42.2
2 emplois	30.0	30.0
3 emplois	6.9	6.9
4 emplois	2.5	2.5
5 emplois	0.9	0.9
6 emplois	0.1	0.1

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Type d'habitation

1. maison rurale	9.3	9.3
2. maison unifamiliale individuelle	34.6	34.6
3. maison unifamiliale jumelée	9.1	9.1
4. maison unifamiliale en série	25.3	25.3
5. 2 à 4 logements	13.1	13.1
6. 5 à 19 logements	5.7	5.7
7. plus de 20 logements	1.1	1.1
8. habitat précaire	0.1	0.1
9. N.R.	1.7	1.7

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Canton

1. Luxembourg - ville	17.7	17.7
2. Capellen	6.9	6.9
3. Esch sur Alzette	34.3	34.3
4. Luxembourg campagne	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

Caractéristiques du chef de ménage

Sexe du chef de ménage

1. masculin	86.3	86.3
2. féminin	13.7	13.7

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Âge du chef de ménage

1. < 24 ans	1.8	1.8
2. 25 - 34 ans	18.7	18.7
3. 35 - 44 ans	27.3	27.3
4. 45 - 54 ans	23.0	23.0
5. 55 - 64 ans	15.8	15.8
6. > 64 ans	13.4	13.4

État civil du chef de ménage

1. célibataire	6.5	6.5
2. marié	79.6	79.6
3. séparé	1.2	1.2
4. divorcé	2.7	2.7
5. veuf	10.0	10.0

Emploi du chef de ménage

1. sans emploi	26.7	26.7
2. emploi	73.3	73.3

État civil du chef de ménage selon le sexe

1. homme non marié	7.0	7.0
2. homme marié	79.3	79.3
3. femme non mariée	13.4	13.4
4. femme mariée	0.3	0.3

	Pondération classique Membres	Poids moyen du ménage M./non - M.
	Fréquences (%)	
	(n=4648)	(n=4914)

Emploi du chef de ménage selon le sexe

1. homme sans emploi	17.6	17.6
2. homme avec emploi	68.7	68.7
3. femme sans emploi	9.1	9.1
4. femme avec emploi	4.6	4.6

ANNEXE 2

La procédure de Ernst ne comporte aucun biais

La démonstration a été présentée par Ernst (1992, pp. 143 et ss.). On en trouve également une version dans Lavallée et Hunter (1992).

D'une manière générale, une variable aléatoire quelconque X caractérisant une population de taille N peut être décrite de la manière suivante :

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

où x_i est la valeur (x) de cette variable pour tout $i^{\text{ème}}$ unité appartenant à la population.

Une enquête classique peut estimer la variable X en sélectionnant un échantillon probabiliste tel que tout individu i appartenant à la population N , a une probabilité positive et connue d'être sélectionné dans l'échantillon, soit p_i (la probabilité de sélection).

Deux cas peuvent se présenter.

1. Soit, tous les individus ont la même probabilité d'être sélectionnés. Dans ce cas, l'estimation de X peut s'écrire :

$$X^{\text{est}} = \frac{1}{n} \sum_{i=1}^n x_i$$

où X^{est} est l'estimation de X à partir des valeurs (x) observées sur l'ensemble des individus (i) appartenant à l'échantillon qui compte n individus. Chaque individu est pondéré en raison inverse de sa probabilité de sélection, soit $1/p_i$.

On peut écrire de manière strictement équivalente :

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

avec $w_i = 1/p_i$ pour tout individu appartenant à l'échantillon et $w_i = 0$ pour tout individu n'appartenant pas à l'échantillon, w_i étant le poids attribué à chaque individu. Ce poids (w_i) égale $1/p_i$ si tous les individus ont la même probabilité de sélection. Ce poids est nul pour tout individu n'appartenant pas à l'échantillon.

2. Soit, tous les individus n'ont pas la même probabilité de sélection (p_i). Ces différentes probabilités sont connues. L'estimation de X peut s'écrire de la même manière, soit :

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

Toutefois, w_i (le poids des individus) doit être calculé : $w_i = 1/p_i$, le poids d'un individu i correspond à l'inverse de sa probabilité de sélection s'il appartient à l'échantillon et à 0 s'il n'appartient pas à l'échantillon.

Le problème qui se pose est le suivant : peut-on insérer dans l'échantillon du panel des individus qui n'ont pas été sélectionnés au cours de la première sélection de l'échantillon du panel? Ils sont entrés dans un ménage au cours du panel et la probabilité de leur entrée est inconnue.

Ne risque-t-on pas de biaiser l'estimation de X en accordant un poids positif à ces individus?

En fait, l'estimation de X reste sans biais, si l'espérance mathématique de l'estimateur des poids reste égale à 1 :

$$E(X) = X \quad (1)$$

$$\text{si } E(w_i) = 1 \quad (2)$$

l'espérance mathématique de X_{II} est un estimateur sans biais de X (1) si l'espérance mathématique de w_i est un estimateur sans biais de 1 et cette condition est réalisée si l'espérance mathématique de l'estimateur des poids (w_i) égale 1 (2).

Définir une procédure de pondération des ménages sans biais revient donc à définir une variable aléatoire W_i qui puisse satisfaire à la condition de l'équation (2), (Ernst, 1992, p. 144).

Or, l'espérance mathématique d'une variable aléatoire quelconque Y est donnée par :

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

l'espérance mathématique de la variable Y correspond à la somme des produits des valeurs de y par la probabilité que cette valeur se réalise (soit la probabilité P que Y soit égal à y, étant entendu que y peut prendre n'importe quelle valeur).

Par conséquent, l'espérance mathématique de l'estimateur des poids est donnée par la même équation, soit :

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

L'espérance mathématique de l'estimateur des poids (W) des individus (i) appartenant aux ménages ($_{mi}$), soit $E(w_{mi})$, correspond à une somme de produits : les produits des valeurs possibles des poids des individus appartenant aux ménages (w) par les probabilités d'obtenir ces valeurs $P(W_{mi}=w)$.

Dans notre cas, W_{mi} (les poids des individus appartenant aux ménages) peut présenter deux valeurs initiales puisque nous voulons prendre en compte les membres légitimes et les membres "non-légitimes" :

1. soit $W_{mi} = 1/p_{mi}$ pour les membres légitimes dont la probabilité de sélection (p_{mi}) est connue, parce qu'ils appartiennent à un ménage (m) sélectionné dans le cadre de l'échantillon initial,
2. soit $W_{mi} = 0$ pour les membres "non-légitimes" dont la probabilité de sélection ne peut pas être calculée, parce qu'ils n'appartiennent pas à un ménage sélectionné dans le premier échantillon, alors qu'ils appartiennent probablement à la population-cible au moment de ce premier tirage.

Par application de l'équation (3) :

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

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

où $0 \cdot P(W_{mi}=0)$ est nécessairement égal à 0 et $P(W_{mi}=1/p_{mi})$ est la probabilité que le poids initial prenne la valeur $1/p_{mi}$; ce qui correspond très exactement à la probabilité que l'individu soit sélectionné dans l'échantillon. Donc :

$$E(W_{mi}) = (1/p_{mi} \cdot p_{mi}) + 0 = 1, \text{ ce qu'il fallait démontrer.}$$

On notera qu'il n'est donc pas nécessaire de connaître la valeur des probabilités des membres "non-légitimes" puisqu'elle est simplement annulée.