

PANEL SOCIO-ECONOMIQUE "Liewen zu Lëtzebuerg"

Document PSELL n° 59

mai 1994

INSÉRER UN ÉCHANTILLON COMPLÉMENTAIRE DANS UN PANEL LONGITUDINAL DE MÉNAGES ET D'INDIVIDUS

SIMULATIONS
deuxième partie



B. GAILLY (CEPS/Instead)

P. LAVALLÉE (Statistique-Canada)

CEPS/Instead
Walferdange
Grand-Duché de Luxembourg

1994

Document produit par le

CEPS/Instead

**CENTRE D'ETUDES DE POPULATIONS, DE PAUVRETE
ET DE POLITIQUES SOCIO-ECONOMIQUES**

B.P. 65, L-7201 Walferdange
tél. (352) 33 32 33 - 1 - Fax. 33 27 05

Président: Gaston Schaber

INSERER UN ECHANTILLON COMPLEMENTAIRE DANS UN PANEL LONGITUDINAL D'INDIVIDUS ET DE MENAGES

SIMULATIONS deuxième partie

SOMMAIRE

Introduction.....		5
Section 1	APPORT DE NOUVEAUX MÉNAGES: propriétés de l'échantillon complémentaire	10
	1.1. L'échantillon complémentaire par rapport à l'échantillon initial.....	10
	1.2. Cette méthode a une portée restreinte	11
	1.3. La probabilité de sélection initiale des nouveaux membres est connue	12
Section 2	APPLICATION AU PANEL SOCIO-ÉCONOMIQUE LUXEMBOURGEOIS	14
	2.1. L'échantillon initial	14
	2.2. L'échantillon complémentaire	16
	2.3. Les trajectoires dans l'échantillon et leur documentation	18
	2.3.1. Les individus présents dans l'échantillon en 1985 et absents en 1991	18
	2.3.2. Les individus présents dans l'échantillon en 1985 et présents en 1991	20
	2.3.3. Les individus absents de l'échantillon en 1985 et présents en 1991	21
	Tableau 1 Trajectoires individuelles des membres légitimes et non légitimes de l'échantillon complémentaire	25

Section 3	SIMULATION: Calcul des poids des ménages et des individus en 1985 et en 1991	26
	3.1. Calcul des poids en 1985	26
	3.2. Poids individuels initiaux en 1991	26
	3.3. Poids des ménages et poids individuels finaux en 1991	29
 Section 4	 SIMULATION: Insertion dans l'échantillon initial et évaluation.....	 30
	4.1. Insertion dans l'échantillon initial	30
	4.2. Comparaisons avec les données du recensement	31
	4.3. Extension de la comparaison	32
	Tableau 2 Comparaison des caractéristiques démographiques des membres du panel en 1991: échantillon et recensement	33
	Tableau 3 Comparaison des caractéristiques démographiques des membres du panel en 1991: échantillon initial et extensions	34
 Conclusion		 37
 Annexe		 39
 Bibliographie		 43

INTRODUCTION

Une étude basée sur l'observation répétée d'un échantillon d'individus et de ménages se heurte à trois problèmes:

1. Le premier problème résulte de l'érosion ou *attrition de l'échantillon* (décès, personnes introuvables, personnes émigrées...). Ce phénomène d'attrition résulte de la lassitude des personnes. Leur participation à l'enquête est sollicitée chaque année et leur propension à refuser de répondre augmente progressivement.

Ce phénomène resterait sans importance si les refus de répondre enregistrés chaque année se répartissaient de manière aléatoire et n'affectaient l'échantillon d'aucun biais. En outre, la taille des catégories de population qui motivent l'étude doit rester suffisante pour assurer la précision des estimations statistiques.

2. L'échantillon vit en interaction permanente avec la population dont il est issu et qu'il est censé représenter au cours du temps.

A l'occasion de ces échanges, des membres de la population peuvent nouer des liens avec les membres de l'échantillon, entrer dans leur ménage ou entraîner la formation de nouveaux ménages. Ces nouveaux membres n'ayant pas été sélectionnés dans l'échantillon initial, leur probabilité de sélection n'est pas connue. Aucun poids ne peut leur être attribué alors qu'ils devraient être pris en compte pour assurer la représentativité transversale.

Toute la richesse des informations apportées par ces membres risque d'être perdue s'il n'existe aucun moyen de les prendre en compte. Cette déperdition est d'autant plus regrettable que ces informations reflètent précisément *l'évolution de l'échantillon dans la population*.

3. En règle générale, tous les phénomènes observables dans la population se reflètent spontanément dans l'évolution d'un panel longitudinal (naissances, décès, mariages, séparations, émigration). Un phénomène échappe à cette règle: *l'immigration*.

INTRODUCTION

Une procédure spécifique doit être mise en oeuvre pour sauvegarder la représentativité de l'échantillon à cet égard parce que la dynamique de son évolution ne comporte aucun mécanisme permettant de capter spontanément les nouveaux immigrants.

Le problème de l'attrition peut être géré annuellement par des procédures de pondération "classiques". Ces procédures peuvent présenter des variations selon le mode de tirage de l'échantillon initial et selon les panels mais leurs principes restent fondamentalement identiques⁽¹⁾.

Ces procédures ont cependant trois limites:

- elles ne permettent pas de prendre en compte les membres non-légitimes dont on ignore la probabilité de sélection initiale,
- elles ne contribuent pas à la lutte contre la réduction inexorable de la taille de l'échantillon et ne servent que de "béquilles" dans ce domaine,
- elles ne permettent pas de capter les nouveaux immigrants.

Dans une première étude, nous avons mis à l'épreuve les effets de deux procédures permettant de prendre en compte l'information apportée par les nouveaux membres "non-légitimes"⁽²⁾⁽³⁾.

Brièvement, cette procédure permet d'éviter tout postulat relatif au poids initial des membres non-légitimes. Ceux-ci reçoivent un poids nul. Au cours de la procédure, les membres légitimes de chaque ménage partagent leurs poids avec les membres non-légitimes du ménage. Cet *artifice* permet de prendre en compte les informations originales apportées par les membres non-légitimes dans le calcul des valeurs estimées pour les membres légitimes. Au terme de cette procédure, seuls les membres légitimes sont réellement pris en compte.

.....
(1) Voir par exemple: "Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1991" Tomes I et II, B. Gailly, PSELL n° 55, Panel Socio-Economique "Liewen zu Lëtzebuerg", CEPS, Walferdange, G.D. de Luxembourg, 1993.

(2) P. Lavallée et L. Hunter: "Weighting for the Survey of Labour and Income Dynamics", contribution au Symposium 92 de Statistique Canada "Conception et analyse des enquêtes longitudinales", J.Armstrong, N.Darcovich, P.Lavallée, Ottawa, Août 1993.

(3) B. Gailly et P. Lavallée: "Insérer des nouveaux membres dans un panel longitudinal de ménages et d'individus. Simulations" PSELL n° 54, Panel Socio-Economique "Liewen zu Lëtzebuerg", CEPS, Walferdange, G.D. de Luxembourg, 1993.

INTRODUCTION

Ce bref commentaire permet de mettre en évidence le fait que cette solution ne permet pas de modifier la taille réelle de l'échantillon des ménages et la taille de l'échantillon des individus. (Le panel allemand adopte une solution différente à cet égard⁽¹⁾).

L'objet de cette seconde étude est précisément de proposer une procédure permettant de ressourcer l'échantillon du panel en augmentant le nombre de ménages et le nombre d'individus, sans recourir au tirage d'un échantillon supplémentaire dans la population mais en utilisant plutôt un échantillon complémentaire.

De quoi s'agit-il?

L'apport d'un échantillon supplémentaire

Sélectionner un échantillon supplémentaire consisterait à fusionner

- l'échantillon initial observé depuis plusieurs années, par exemple 5 ans, et
- un échantillon sélectionné dans l'ensemble de la population telle qu'elle se présente 5 années après le tirage de l'échantillon initial.

Cet échantillon supplémentaire ne remplace pas l'échantillon initial et les deux échantillons sont fusionnés en prenant en compte le fait que l'échantillon supplémentaire contient des unités qui appartenaient déjà à la population au moment du tirage de l'échantillon initial (double probabilité de sélection). Certaines unités pourraient donc se trouver dans l'échantillon initial et réapparaître dans l'échantillon supplémentaire. L'intersection entre ces deux échantillons n'est pas vide.

Le renouvellement *partiel* de l'échantillon d'un panel, par le tirage d'un nouvel échantillon aléatoire, permet d'introduire dans le panel des nouveaux ménages composés de nouveaux individus légitimes, c'est à dire d'individus dont la probabilité de sélection est connue.

.....
(1) "Das sozio-oekonomisches Panel, Benutzerhandbuch" Deutsches Institut für wirtschaftsforschung (éd.), 1992, Version 6 - Dezember/92, Berlin, Chap. K, U. Rendtel.

INTRODUCTION

Cette méthode présente plusieurs avantages:

- elle restaure la taille de l'échantillon des ménages et des individus
- elle introduit dans l'échantillon des membres dont la présence et la durée de vie ne dépend pas des liens plus ou moins durables qu'ils ont noués avec les membres légitimes appartenant à l'échantillon initial: leur trajectoire dans le panel dépend de leurs caractéristiques personnelles
- elle préserve l'intérêt de l'observation et de l'analyse des catégories de population qui motivent l'étude
- elle permet de prendre en compte le phénomène de l'immigration et de corriger ainsi la distribution des membres selon leur nationalité
- elle permet de préserver le degré de précision des estimations statistiques.

L'échantillon initial et l'échantillon supplémentaire sont sélectionnés de manière indépendante. Par conséquent, certains membres appartenant au nouvel échantillon pouvaient être présents dans la population au moment du tirage de l'échantillon initial. Ils ont donc une double probabilité de sélection. Ces problèmes sont connus et les solutions sont largement documentées.

Cette méthode impose également des contraintes qui peuvent la rendre impraticable dans certains pays.

L'accessibilité du fichier administratif est la condition la plus élémentaire: elle n'est pas nécessairement la plus aisée à satisfaire. La répétition de cette procédure, à intervalles réguliers, est une opération qui peut s'avérer très lourde.

Une solution intermédiaire: un échantillon complémentaire

Par contre, un échantillon complémentaire est formé par un ensemble d'unités sélectionnées dans le cadre du tirage de l'échantillon initial mais restées inobservées pendant les premières années de l'étude.

Au moment de tirer l'échantillon initial, il est indispensable de tirer un nombre d'unités supérieur à la taille désirée pour l'échantillon. Ces unités sont destinées

INTRODUCTION

au remplacement des refus de répondre, des personnes introuvables, décédées ou émigrées. Le caractère aléatoire du tirage de ces "substituts" permet de sauvegarder le caractère probabiliste de l'échantillon.

Lorsque la taille souhaitée pour l'échantillon est atteinte, un certain nombre d'unités n'ont pas été utilisées. Ces unités, sélectionnées au même moment, selon la même procédure et à partir de la même liste de référence (par exemple, le même fichier administratif) que l'échantillon initial qui sera observé restent disponibles. Elles ne peuvent en aucun cas se trouver simultanément dans l'échantillon observé. L'intersection entre ces deux échantillons est vide: ces échantillons sont des échantillons complémentaires.

Cette seconde démarche doit être vue comme une solution transitoire. Elle ne peut en aucun cas remplacer définitivement le tirage d'un échantillon supplémentaire dans la population. Par contre, elle permet de pallier assez rapidement et à peu de frais aux effets de l'attrition de l'échantillon qui se manifestent souvent au cours des premières années de l'observation.

La méthode proposée ici présente deux différences par rapport à la méthode du renouvellement partiel basé sur le tirage aléatoire d'un nouvel échantillon:

- elle a une portée plus restreinte et
- le nouvel échantillon est complémentaire de l'échantillon initial. (Par la suite, nous utiliserons systématiquement le terme "d'échantillon complémentaire" pour désigner cet échantillon et le distinguer de l'échantillon initial).

Par contre, les deux méthodes présentent un point commun:

- la probabilité de sélection initiale des nouveaux membres est connue.

1

APPORT DE NOUVEAUX MÉNAGES: propriétés de l'échantillon complémentaire

1.1. L'échantillon complémentaire par rapport à l'échantillon initial

La méthode proposée ici introduit dans l'échantillon initial du panel un nouvel échantillon complémentaire de l'échantillon initial.

L'échantillon initial est souvent constitué à partir d'un *fichier* administratif de référence (PSELL, CEPS, Luxembourg). Ce n'est pas forcément toujours le cas (EDTR, Statistique-Canada).

Ce fichier a permis de sélectionner aléatoirement une *liste* d'individus repères.

L'*échantillon* initial résulte d'un tirage aléatoire d'individus repères recensés dans cette liste et conduisant chacun à l'adresse d'un ménage.

Le nombre d'individus contenus dans cette liste est généralement supérieur à la taille fixée pour l'échantillon. Ce surplus d'adresses est destiné au remplacement aléatoire des refus de répondre enregistrés au cours de l'enquête. Cette méthode est valable soit si les refus sont "ignorables", soit s'il est possible de les prendre en compte et de corriger leurs effets dès la première vague d'observation.

Lorsque la taille souhaitée est atteinte, un certain nombre d'adresses restent inutilisées: étant donné les principes de sélection de la liste des individus repères et de l'échantillon initial, ces ménages non-observés constituent toujours un échantillon aléatoire représentatif de la même population.

Cet échantillon non-observé est un échantillon complémentaire du premier: les deux sous-ensembles ne comportent aucune intersection. Les membres de l'échantillon observé ne peuvent en aucun cas appartenir à cet échantillon complémentaire.

La liste fournit donc une réserve d'individus repères dont les adresses n'ont pas été utilisées.

Ces adresses peuvent être utilisées au moment d'effectuer une première mise à jour de l'échantillon du panel et ce moment peut être déterminé en fonction de la rapidité de la dégradation de l'échantillon.

Cet élargissement de l'échantillon ne pourrait être effectué si l'on ne connaissait pas la probabilité de sélection initiale des membres de cet échantillon complémentaire. Il suppose en effet que les différences dans la composition des ménages et les modifications des caractéristiques individuelles entre l'année du tirage et l'année de la première observation soient prises en compte au moment de leur ajout dans l'échantillon initial.

Les informations indispensables à la réalisation de ces opérations seront détaillées dans le chapitre suivant.

1.2. Cette méthode a une portée restreinte

Elle présente les mêmes avantages que le renouvellement partiel de l'échantillon initial.

- Elle restaure la taille de l'échantillon des ménages et des individus.
- Elle préserve l'intérêt de l'observation et de l'analyse des catégories de la population qui motivent l'étude.
- Elle permet de préserver le degré de précision des estimations statistiques.

Par contre, elle ne permet pas de prendre en compte les personnes dont l'immigration est récente: elle ne peut donc pas être substituée au renouvellement partiel de l'échantillon à l'aide d'un échantillon supplémentaire.

Son maniement est beaucoup plus aisé: elle ne fait pas appel au fichier administratif de référence, aux autorisations et aux intermédiaires indispensables. Elle peut intervenir plus rapidement qu'une solution nécessitant le recours à un fichier administratif. Les refus de répondre étant particulièrement fréquents dans l'échantillon initial au cours des trois ou quatre premières années d'enquête, l'échantillon complémentaire corrige alors rapidement cette situation.

Il peut être utile de recourir à cette solution avant de procéder à un renouvellement partiel de l'échantillon, surtout lorsque cette procédure se heurte à des difficultés pratiques.

1.3. La probabilité de sélection initiale des nouveaux membres est connue

Cette procédure introduit dans le panel un nouvel échantillon de membres légitimes: leur probabilité de sélection au moment du tirage de l'échantillon initial est connue.

Le poids initial du ménage et le poids initial des membres peuvent donc être calculés.

Nous entendrons par "poids initial" le poids dérivé de la probabilité de sélection et servant au calcul d'un poids "final" destiné aux estimations.

Le poids "initial" est la valeur utilisée dans le calcul du "partage des poids" en vue de l'obtention d'un poids final. Il prend la valeur

- . "1/ Probabilité d'inclusion", pour les membres légitimes (soit le poids de sondage)
- . "0", pour les membres non-légitimes, entrés dans les ménages entre 1985 et la première année d'observation.

Le fait que leur probabilité de sélection au moment du tirage de l'échantillon initial soit connue est une propriété essentielle.

La composition des ménages a pu se modifier en effet durant la période où ces membres n'ont pas été observés. Ces modifications peuvent entraîner des changements dans les poids finaux des unités de l'échantillon⁽¹⁾. Comment pourrait-on calculer leur poids au moment où ils sont observés si leur probabilité de sélection initiale n'était pas connue?

.....
(1) cf. la première partie de cette étude. B. Gailly et P. Lavallée: "Insérer des nouveaux membres dans un panel longitudinal de ménages et d'individus. Simulations" PSELL n° 54, Panel Socio-Economique "Liewen zu Lëtzebuerg", CEPS, Walferdange, G.D. de Luxembourg, 1993.

La question qui se pose est donc: quelles sont les informations indispensables pour calculer les poids initiaux et quelles sont les informations nécessaires pour recalculer ces poids au moment où ces membres sont insérés dans le panel?

L'application au panel socio-économique luxembourgeois permettra de fournir des indications opérationnelles. Certains éléments peuvent varier selon les panels.

2

APPLICATION AU PANEL SOCIO-ÉCONOMIQUE LUXEMBOURGEOIS

L'échantillon initial et l'échantillon complémentaire ont été sélectionnés au même moment, à partir de la même liste d'adresses et selon la même procédure.

L'échantillon complémentaire, non-observé entre 1985 et 1991, présente donc les mêmes propriétés que l'échantillon initial. Il a également connu la même "histoire".

Il doit donc faire l'objet de la même procédure de pondération que l'échantillon initial sélectionné en 1985.

Les poids finaux doivent être corrigés en fonction des refus de répondre enregistrés en 1991 et en fonction des modifications de la composition des ménages entre 1985 et 1991.

La jonction des deux échantillons exige donc deux types d'informations:

- les informations liées au mode de pondération des unités d'analyse et
- les informations relatives à la composition des ménages en 1985 et en 1991.

2.1. L'échantillon initial

L'échantillon initial du panel socio-économique luxembourgeois est un échantillon aléatoire simple, sélectionné en 1985, à partir d'une liste d'individus repères fournie par l'Inspection Générale de la Sécurité Sociale. Cette liste formait elle-même un échantillon aléatoire simple représentatif de l'ensemble des personnes résidant au Luxembourg en 1985 et attachées d'une manière quelconque au système de sécurité ou de protection sociale.

La constitution de l'échantillon initial a été interrompue à 2012 ménages.

Au sein de chaque ménage, tous les membres ont été pris en compte; soit 6110 personnes.

Tous les ménages n'ont pas la même probabilité de sélection. En effet, le fichier des individus repères ne recense que les personnes disposant d'un revenu et les ménages peuvent être composés de plusieurs titulaires de revenus.

Plus les ménages comptent un grand nombre de titulaires de revenus, plus leur probabilité d'être sélectionnés s'élève.

Le poids de sondage du ménage est donc égal à l'inverse du nombre de personnes ayant un revenu en 1985 et le poids de sondage de chaque membre du ménage est égal au poids du ménage, puisque tous les membres du ménage ont été observés.

Il va de soi que ce critère de pondération peut varier selon les panels.

Dans le cas où le fichier recense l'ensemble des personnes physiques résidant dans le pays, un tirage aléatoire de personnes accorderait aux ménages une probabilité de sélection proportionnelle à la taille du ménage. Le poids des ménages devrait être égal à l'inverse de la taille du ménage et, dans ce cas, le nombre de personnes disposant d'un revenu n'aurait aucune pertinence.

Au cours des vagues d'enquêtes suivantes deux catégories de personnes font l'objet de procédures particulières dans le calcul des poids individuels: les enfants nés après 1985 et les membres non-légitimes.

En réalité, il existe plusieurs options. L'EDTR (Statistique-Canada) entend distinguer deux catégories d'individus non-légitimes: les individus non-légitimes "présents au départ du panel" et les individus non-légitimes "absents au départ du panel". Dans cette optique, les enfants nés après 1985 dans le PSELL seraient des individus non-légitimes absents au départ du panel.

Dans le PSELL, ces enfants nés après 1985 sont pris en compte au titre de membres légitimes à condition que l'un des deux ascendants directs soit un membre légitime du panel. Cette procédure est calquée sur la méthode utilisée par le P.S.I.D. (Panel de l'université d'Ann Arbor (Michigan)). Elle est décrite dans les documents relatifs à la pondération des individus et des ménages publiés annuellement par le C.E.P.S.: sauf exception, le "poids initial" de ces enfants est égal à la moyenne des poids des ascendants directs.

Les membres non-légitimes reçoivent dans un premier temps, un poids initial nul.

Le poids final du ménage utilisé pour les estimations transversales correspond à la valeur moyenne des poids de ses membres et dans un dernier temps ce poids est reventilé sur tous les membres du ménage.

2.2. L'échantillon complémentaire

En 1985, 230 individus repères restaient disponibles au moment où la constitution de l'échantillon a été interrompue. A cette époque, ils regroupaient autour d'eux 697 personnes pour former 230 ménages.

Entre 1985 et 1991, les personnes appartenant à ces 230 ménages ont formé 30 nouveaux ménages (séparations, mariages, co-habitations).

En 1991, ce sont 764 personnes formant donc $230 + 30$ ménages, soit 260 ménages qui acceptent de répondre. Tous ces ménages sont composés d'individus issus directement ou indirectement de la même liste d'individus repères sélectionnés en 1985.

L'échantillon complémentaire et l'échantillon initial ayant été sélectionnés suivant les mêmes règles, le calcul du poids de sondage du ménage et de ses membres s'effectue de la même manière et en fonction de l'état du ménage en 1985.

Ceci implique que les ménages observés dans l'échantillon complémentaire en 1991, puissent être reconstitués dans l'état où ils se trouvaient en 1985.

L'ensemble de ces opérations ne peut être effectué que sous quatre conditions.

1. La trajectoire suivie par chaque individu entre 1985 et 1991 doit être reconstituée afin de distinguer clairement:

- les individus qui appartenaient à l'échantillon sélectionné en 1985, mais qui sont absents de l'échantillon en 1991
- les individus qui apparaissent pour la première fois dans l'échantillon observé en 1991
- les individus présents dans l'échantillon aux deux époques.

2. Ces trajectoires doivent être reconstituées au niveau des individus. Seuls les individus sont des unités identifiables sans ambiguïté à travers les époques. Les ménages sont des unités déformables. Il est arbitraire d'établir l'identité de deux ménages observés à des époques différentes.

3. La recomposition des ménages dans l'état où ils se trouvaient en 1985 suppose que l'on dispose de quelques informations stratégiques concernant les liens entre les individus et leur ménage d'origine.

4. Les informations permettant d'établir la probabilité de sélection initiale des ménages doivent être collectées. Dans le cas du Luxembourg, il s'agit essentiellement du nombre de titulaires de revenus dans chaque ménage en 1985, mais certaines particularités du procédé de pondération exigent des informations complémentaires: l'âge des enfants et le statut de membre légitime ou non légitime de leurs ascendants directs.

Il serait préférable de procéder dès la première vague d'enquête à la collecte de certaines informations telles que: le recensement des membres du ménage présents à cette époque, les indications nécessaires au calcul de la probabilité de sondage. La nature de ces informations n'interdit pas de procéder à cette reconstitution dans un délai raisonnable au prix de quelques imperfections (ce fut le cas au Luxembourg).

2.3. Les trajectoires dans l'échantillon et leur documentation

On peut distinguer, a priori, trois grands types de trajectoires:

1. les individus présents dans l'échantillon en 1985 et absents en 1991
2. les individus présents dans l'échantillon en 1985 et en 1991
3. les individus absents de l'échantillon en 1985 et présents en 1991.

Chaque trajectoire recouvre des sous-groupes spécifiques. Ils se distinguent en fonction de leur statut dans l'ensemble du processus de pondération. L'identification de chacun de ces sous-groupes fait appel à des informations particulières.

2.3.1 les individus présents dans l'échantillon en 1985 et absents en 1991

Cette trajectoire recouvre cinq sous-groupes:

- A. *Les personnes sélectionnées en 1985 qui n'ont pas pu être contactées en 1991.*
- B. *Les personnes sélectionnées en 1985 qui ont été contactées pour la première fois en 1991 mais qui ont refusé de répondre.*
- C. *Les personnes sélectionnées en 1985 qui ont émigré entre 1985 et 1991.*
- D. *Les personnes appartenant à un ménage sélectionné en 1985 qui sont décédées entre 1985 et 1991.*
- E. *Des membres non-légitimes, n'appartenant pas à l'échantillon sélectionné en 1985, mais se trouvant dans l'un des trois derniers sous-groupes (B,C,D).*

Il est important d'isoler ce dernier sous-groupe puisqu'il ne doit pas être pris en compte dans la reconstitution des ménages sélectionnés en 1985. Il suffit que ces membres aient appartenu à un ménage entre 1985 et 1991 pour que ce type de confusion puisse se produire.

Ces cinq groupes peuvent être isolés sur la base des informations suivantes:

- le numéro d'identification du ménage de chaque individu en 1985 (on verra que dans certains sous-groupes, les individus doivent être identifiés par deux numéros d'identification de ménage)
- le numéro d'identification de l'individu (afin d'établir le lien inter-annuel)
- la présence/absence de l'individu en 1985 (afin d'identifier clairement les individus légitimes et non-légitimes)
- la raison du départ du ménage d'origine (afin d'identifier les refus de répondre, les personnes décédées et les émigrés, les sorties de l'échantillon, les déménagements ou changements d'adresses)
- l'existence ou l'absence de revenus pour chaque individu en 1985 (revenus de capitaux exclus).

Cette dernière information est capitale pour le calcul des poids initiaux dans le panel luxembourgeois (PSELL). Elle concerne également les refus de répondre, les personnes émigrées et les personnes décédées entre 1985 et 1991 puisque ces personnes doivent être prises en compte dans le calcul des poids des ménages en 1985.

Les quatre premiers sous-groupes permettent de reconstituer la composition des ménages en 1985.

La raison de l'absence en 1991 permet de classer les deux premiers groupes parmi les "refus de répondre", de calculer ainsi le taux de non-réponse et d'envisager une correction des poids initiaux en fonction de la distribution des non-réponses.

Les décédés et les émigrés sont écartés du calcul du taux de non-réponse puisqu'ils reflètent l'évolution démographique générale de la population.

2.3.2 les individus présents dans l'échantillon en 1985 et présents en 1991

Cette trajectoire correspond à deux sous-groupes.

Le second sous-groupe se distingue du premier par le fait que ses membres ont fondé de nouveaux ménages. Ils reçoivent donc un numéro de ménage différent en 1985 et en 1991. La raison de leur départ du ménage initial doit également être spécifiée.

F. Les personnes présentes dans l'échantillon en 1985 et en 1991 et appartenant au même ménage aux deux époques.

Pour isoler ce sous-groupe et reconstituer les ménages dans l'état où ils se trouvaient en 1985, il suffit de connaître:

- le numéro d'identification du ménage de chaque individu en 1985
- le numéro d'identification individuel
- l'existence/absence de revenus pour chaque individu en 1985 (critère d'identification de la probabilité de sélection)
- l'âge en 1985 (afin d'identifier les nouveaux-nés entrés dans ces ménages entre 1985 et 1991: ils n'appartenaient pas au ménage observé en 1985, bien qu'ils portent le numéro d'identification d'un ménage sélectionné en 1985).

Dans le cadre du panel luxembourgeois, seules sont retenues dans ce sous-groupe les personnes âgées de 7 ans et plus. Tous les membres nés entre 1985 et 1990 sont âgés de 0 à 6 ans.

G. Les personnes présentes dans l'échantillon en 1985 et en 1991, ayant fondé un nouveau ménage entre 1985 et 1991.

Pour identifier ces personnes, il est indispensable de connaître:

- le numéro d'identification du ménage de chaque individu en 1985
- le numéro d'identification du nouveau ménage auquel chacune de ces personnes appartient en 1991
- le numéro d'identification individuel

- l'existence/absence de revenus pour chaque individu en 1985 (critère d'identification de la probabilité de sélection)
- la raison du départ du ménage d'origine, afin de distinguer ces personnes des autres "départs" (cf. la première trajectoire: décès, émigration).

Ces personnes, observées pour la première fois en 1991, doivent être resituées dans leur ménage d'origine au moment de calculer le poids initial des ménages en 1985. Il est donc absolument indispensable qu'elles soient identifiées par deux numéros d'identification de ménage différents, sans quoi il serait impossible de les classer correctement, soit en 1985, soit en 1991.

2.3.3 les individus absents de l'échantillon en 1985 et présents en 1991

En principe, tous ces individus peuvent être pris en compte au titre d'individus non-légitimes puisqu'ils n'ont pas été inclus dans l'échantillon initial sélectionné en 1985. Toutefois, il peut être utile de distinguer certains sous-groupes en fonction de la méthode de pondération adoptée dans le cadre de chaque panel.

H. Les personnes absentes de l'échantillon en 1985 et présentes dans un nouveau ménage constitué entre 1985 et 1991 et composé d'au moins un membre légitime.

Le cas le plus typique correspond aux nouveaux membres ayant fondé un nouveau ménage avec une personne ayant quitté un ménage existant en 1985.

Ces personnes n'entrent pas dans la composition des ménages sélectionnés en 1985.

Il est donc indispensable de les identifier et de connaître:

- le numéro d'identification du ménage de chaque individu en 1991
- le numéro d'identification individuel
- la raison de leur entrée dans ce ménage (afin de les distinguer des autres "entrées")
- leur âge en 1991 (afin de les distinguer des enfants âgés de moins de 7 ans et plus particulièrement des enfants dont l'un des ascendants directs est un membre légitime du panel).

2

APPLICATION AU PANEL SOCIO-ÉCONOMIQUE LUXEMBOURGEOIS

1. Les enfants nés dans l'échantillon entre 1985 et 1991

Ces enfants sont observés pour la première fois en 1991.

Pour le PSELL, ces enfants sont des membres légitimes du panel. Ils sont entrés simultanément dans la population et dans l'échantillon et l'un au moins de leurs ascendants directs est un membre légitime du panel.

Ils sont tous âgés de moins de 7 ans en 1991.

Leur probabilité d'inclusion dans le panel dépend directement de la probabilité de sélection de leurs ascendants directs; soit, leur probabilité de sélection initiale (1985).

L'identification des ascendants permet d'accéder à leur poids initial et de calculer le poids initial de l'enfant; soit la valeur moyenne des poids des parents.

Les enfants nés dans des ménages qui existaient déjà en 1985 reçoivent un poids égal à cette valeur moyenne (*groupe 1.1*).

Les enfants nés dans des ménages formés *entre* 1985 et 1991 reçoivent un poids égal à la moitié du poids de l'ascendant qui appartenait à l'échantillon initial (*groupe 1.2*). Ces enfants sont entrés dans leur ménage avant 1991 mais, à la différence des enfants appartenant au sous-groupe précédent (*1.1*), ils y sont entrés après la formation d'un nouveau ménage. Leur poids initial doit donc être calculé après l'arrivée d'un membre non-légitime dans le ménage donc, logiquement, après 1985.

La reconstitution des ménages sélectionnés en 1985 ne doit pas prendre en compte ces enfants âgés de moins de 7 ans en 1991 et (lorsque le système de pondération prévoit de leur réserver un traitement particulier) ces enfants doivent être isolés afin de calculer leur poids initial en 1991.

Il est indispensable de connaître:

- le numéro d'identification de leur ménage en 1985 (pour les enfants nés dans des ménages existant en 1985)
- le numéro d'identification de leur ménage en 1991
- le numéro d'identification individuel
- la raison de leur "entrée" dans un nouveau ménage

- leur âge
- le numéro d'identification individuel du père et de la mère. Le calcul préalable du poids initial des ascendants directs permet d'établir leur statut (légitime/non-légitime). Le numéro matricule des ascendants permet d'accéder à ces poids et de calculer le poids initial de l'enfant.

J. Les autres individus non-légitimes entrés et toujours présents dans des ménages qui existaient déjà en 1985 (et les enfants qui les accompagnent).

Ces personnes doivent être identifiées parce qu'elles ne sont pas prises en compte au moment de reconstituer la composition du ménage en 1985.

Le dispositif de pondération du PSELL exige également qu'ils soient, en 1991, isolés des membres légitimes. Dans un premier temps, ils reçoivent un poids nul, avant d'être pris en compte, dans un second temps, dans le calcul du poids du ménage.

Pour isoler ce sous-groupe, il suffit de connaître

- le numéro d'identification de leur ménage en 1991
- le numéro d'identification individuel
- la raison de leur "entrée" dans le ménage
- leur âge (afin de pouvoir les distinguer éventuellement des membres du sous-groupe précédent).

En résumé, ces différentes trajectoires peuvent être distinguées à partir de dix informations. Seules les 5 premières peuvent être utilement collectées dès la première enquête du panel:

- le numéro d'identification du ménage de chaque individu en 1985
- le numéro d'identification de l'individu
- la présence/absence de l'individu dans l'échantillon en 1985
- l'âge en 1985 (qui ne peut être déduit de l'âge en 1991 que pour les membres présents en 1991...).

- l'existence ou absence de revenus pour chaque individu présent en 1985 (ou tout autre critère important pour le calcul des poids initiaux)
- le numéro d'identification du ménage de chaque individu en 1991
- l'âge en 1991
- la raison de l'entrée dans un nouveau ménage
- la raison du départ du ménage d'origine
- le numéro d'identification individuel de la mère et du père.

L'observation rétrospective du nombre de personnes ayant un revenu dans le ménage au moment de la sélection de l'échantillon peut être une source d'erreurs. Mais ces erreurs se cantonnent dans la catégorie des ménages dont la composition s'est modifiée entre 1985 et 1991.

Il est possible également que tous les membres du ménage ne puissent être identifiés. Le calcul des poids de sondage est impossible et ces ménages doivent être écartés de l'échantillon (en 1991). Certains cas peuvent être détectés grâce à l'âge des personnes. Certaines personnes, mariées en 1991, étaient nécessairement mineures en 1985. Si leur ménage, en 1985, ne comporte aucune personne adulte, il est probable que tous les membres n'ont pas pu être identifiés.

Le Tableau 1 présente la synthèse des trajectoires observées dans l'échantillon complémentaire du panel luxembourgeois, en 1991.

A aucun moment, l'échantillon n'a compté 818 personnes. Ces 818 personnes correspondent au total des personnes observées soit dans le cadre de l'échantillon initial, soit dans l'échantillon de 1991.

Seules les personnes appartenant aux groupes A, B, F, G, C et D appartiennent à l'échantillon sélectionné en 1985, soit 697 personnes.

Entre 1985 et 1991, 54 personnes quittent l'échantillon (groupes A, B, C et D).

Aux 643 membres des groupes F et G viennent se joindre 121 personnes qui seront prises en compte soit au titre de membres légitimes (groupes I.1 et I.2) soit au titre de membres non-légitimes (groupes H et J).

Tableau 1

**Trajectoires individuelles des membres légitimes et non-légitimes
de l'échantillon complémentaire**

TRAJECTOIRES INDIVIDUELLES ⁽¹⁾		Fréquence absolue	Fréquence relative
A.B.	Refus de répondre 91	34	4.1 %
F.	Présents 85 - 91	611	74.6%
G.	Membres légitimes fondent un ménage	32	3.9%
I.1.	Enfants < de 7 ans dans ménage initial	13	1.1 %
I.2.	Enfants < de 7 ans dans nouveau ménage	13	1.7%
H.	Membres non-légitimes fondent un ménage	30	3.7%
J.	Membres non-légitimes dans ménage initial	65	8.5%
C.D.E.	Emigrés / Décédés	20	2.4%
Total individus observés		818	100.0%

Source: PSELL - CEPS/Instead

(1) Les lettres renvoient aux sous-groupes décrits ci-avant dans le texte.

3

SIMULATION:

Calcul des poids des ménages et des individus en 1985 et en 1991

3.1. Calcul des poids en 1985

L'échantillon complémentaire sélectionné en 1985 compte 697 membres.

Les adresses sélectionnées en 1985 permettent de reconstituer 230 ménages (2 ménages n'ont pas pu être reconstitués de manière exhaustive).

Le nombre de titulaires de revenus présents dans chaque ménage peut être recensé. Il détermine la probabilité de sélection de chaque ménage et permet de calculer son poids de sondage qui dépend du nombre de titulaires de revenus dans le ménage.

Tous les membres du ménage étant observés, le poids du ménage peut être attribué à chacun de ses membres.

Chaque membre représente le ménage au même titre puisque les poids individuels sont identiques.

3.2. Poids individuels initiaux en 1991

L'observation effectuée en 1991 montre que l'échantillon complémentaire a subi des modifications, tant au niveau des ménages qu'au niveau des individus. Toutes les modifications observées dans l'échantillon initial du panel se retrouvent dans l'échantillon complémentaire: elles impliquent donc la mise en place du même dispositif de pondération.

1. Les refus de répondre en 1991

Au moment où la première enquête se déroule soit en 1991, 34 personnes refusent de répondre. Les critères disponibles permettent d'écarter l'hypothèse de la présence de biais systématiques liés à ces refus de répondre et leur proportion est aussi trop faible.

SIMULATION:
Calcul des poids des ménages et des individus
en 1985 et en 1991

3

2. Les personnes décédées ou émigrées entre 1985 et 1991

Ces personnes reflètent l'évolution démographique de la population: elles n'introduisent aucun biais dans l'échantillon.

3. Les membres légitimes n'ayant pas changé de ménage

Le groupe F regroupe 611 membres légitimes. Ils ont répondu en 1985 et en 1991. Ils n'ont pas changé de ménage. Les refus de répondre n'ayant eu qu'un impact insignifiant, ces membres gardent, dans un premier temps, le même poids qu'en 1985.

4. Les membres légitimes ayant fondé un nouveau ménage

Le groupe G regroupe également 32 membres légitimes: ils ont répondu en 1985 et en 1991 mais ils ont quitté leur ménage d'origine pour fonder un nouveau ménage.

Ces membres, ayant quitté leur ménage initial, reçoivent deux numéros d'identification du ménage: le premier les rattache à leur ménage d'origine; le second permet d'identifier leur nouveau ménage; ils partagent éventuellement ce second matricule avec d'autres membres parmi lesquels on trouvera généralement des membres non-légitimes.

Les refus de répondre n'ayant eu qu'un impact insignifiant, les membres légitimes gardent, dans un premier temps, le même poids qu'en 1985.

Les membres non-légitimes reçoivent un poids initial nul.

5. Les nouveaux enfants des membres légitimes

Fils et filles de membres légitimes, ces enfants sont nés simultanément dans la population de référence et dans l'échantillon: ils héritent de la probabilité d'inclusion et de la moyenne des poids de leurs ascendants directs.

Ils sont nés avant 1991, c'est à dire avant que l'on puisse observer des modifications éventuelles de la composition des ménages. Leur probabilité

3

SIMULATION:

Calcul des poids des ménages et des individus en 1985 et en 1991

d'inclusion et leur poids sont donc liés à l'état du ménage avant 1991. Ils doivent être calculés en fonction de la composition du ménage en 1985: ce sont donc les probabilités d'inclusion et les poids du père et de la mère en 1985 qui doivent être pris en compte.

6. Les membres non-légitimes

Les groupes H et J regroupent 95 membres non-légitimes. Les uns sont entrés dans l'échantillon à l'occasion de la fondation d'un nouveau ménage. D'autres ont rejoint des ménages qui existaient en 1985.

Ces membres auraient pu appartenir à la population de référence en 1985. Leur probabilité de sélection initiale est inconnue. Ils reçoivent, dans un premier temps, un poids initial nul.

7. Les enfants nés dans un nouveau ménage

Ces 13 enfants sont nés dans des ménages composés d'un ou plusieurs membres légitimes présents en 1985 et d'un ou plusieurs membres non-légitimes entrés dans l'échantillon après 1985.

Ces enfants sont nés simultanément dans la population et dans l'échantillon. Ils ne pouvaient pas appartenir à la population de référence au moment où l'échantillon initial a été tiré.

Ils reçoivent un poids dont la valeur est égale à la valeur moyenne des poids des ascendants directs. Ce poids ne pouvait en aucun cas leur être attribué au préalable: en 1985, le membre légitime qui leur a donné naissance appartenait à un autre ménage et le membre non-légitime n'appartenait pas à l'échantillon.

Le membre non-légitime a reçu un poids initial nul. La valeur moyenne des poids des ascendants est donc égale la moitié du poids de l'ascendant membre légitime du panel.

En 1991, 669 membres légitimes ($611 + 32 + 13 + 13$) bénéficient d'un poids initial différent de "0" et 95 membres non-légitimes ($30 + 65$) ont reçu un poids initial nul.

En 1991, l'échantillon complémentaire compte donc 764 personnes.

SIMULATION:
Calcul des poids des ménages et des individus
en 1985 et en 1991

3

3.3. Poids des ménages et poids individuels finaux en 1991

Le poids des 260 ménages observés en 1991 est calculé selon les règles adoptées pour l'échantillon initial du panel: il correspond à la valeur moyenne des poids des individus qui composent le ménage en 1991.

Les membres non-légitimes interviennent dans le calcul du poids du ménage munis d'un poids initial nul.

Le poids final du ménage est reventilé sur chacun de ses membres, légitime ou non. Chaque membre représente le ménage au même titre (muni du même poids).

Rappelons que le poids final attribué aux membres non-légitimes n'est qu'un artifice permettant de transférer les informations qu'ils apportent aux membres légitimes.

Ce procédé n'implique aucun postulat relatif à la probabilité de sélection initiale des membres non-légitimes.

On trouvera, en annexe à cette étude, la démonstration en vertu de laquelle la prise en compte des individus non-légitimes ne modifie pas finalement les poids des individus légitimes appartenant aux ménages de l'échantillon.

L'insertion de cet échantillon complémentaire dans l'échantillon du panel modifie-t-elle profondément le profil de l'échantillon? Cet effet peut être estimé en procédant à quelques comparaisons.

Le dernier recensement de la population a été effectué au Luxembourg en 1991. Il permet de vérifier la conformité des distributions de quelques variables démographiques telles qu'elles se présentent dans le panel, en 1991 pour les cas suivants:

- lorsque seuls les membres légitimes appartenant à l'échantillon initial du panel sont pris en compte et pondérés selon la procédure classique (sans les membres non-légitimes);
- lorsque les membres légitimes et non-légitimes de l'échantillon initial du panel sont pris en compte et pondérés selon la méthode du partage des poids;
- lorsque les membres légitimes et non-légitimes de l'échantillon initial du panel et de l'échantillon complémentaire sont pris en compte et pondérés selon la méthode du partage des poids.

La comparaison entre ces trois cas est, ensuite, étendue à un plus grand nombre de variables sans référence aux données du recensement.

4.1. Insertion dans l'échantillon initial

La jonction des deux échantillons a pour effet d'augmenter la taille de l'échantillon initial (individus et ménages).

Cet échantillon final n'est pas un échantillon stratifié. Les deux échantillons sont issus d'une même source. Ils sont strictement dépendants: il est impossible qu'un individu appartienne aux deux échantillons. Ils ont connu la (les) même(s) histoire(s). Ils ont été pondérés selon les mêmes règles. Il s'agit d'un seul échantillon dont la taille est simplement accrue.

Le calcul de la variance n'est donc pas modifié mais l'erreur des estimations diminue puisque la variance est divisée par un échantillon de plus grande taille. Les estimations sont plus précises et situées dans un intervalle de confiance plus étroit.

Cette caractéristique marque une différence essentielle entre cette procédure de ressourcement de l'échantillon et la procédure envisagée dans l'étude précédente. Rappelons que cette dernière procédure prenait en compte les individus non-légitimes mais ne modifiait pas réellement la taille de l'échantillon.

La taille de l'échantillon était accrue de manière artificielle (cf. Annexe: seule la valeur moyenne des variables est influencée par la prise en compte des individus non-légitimes. En réalité, les individus non-légitimes ont et gardent un poids nul).

La prise en compte des individus non-légitimes enrichit l'information contenue dans l'échantillon. Cependant l'erreur des estimations se calcule toujours par référence au nombre d'individus légitimes puisque les membres non-légitimes ont toujours un poids nul.

L'insertion d'un échantillon complémentaire, contrairement à la première procédure, augmente le nombre d'individus légitimes et le nombre de ménages: l'erreur des estimations, calculée selon la méthode classique, s'en trouve réduite, aussi bien au niveau des individus que des ménages.

4.2. Comparaisons avec les données du recensement

Les distributions de fréquences obtenues dans les différentes versions de l'échantillon ne présentent que des écarts insignifiants.

Le parallélisme est identique avec les données du recensement des personnes physiques effectué en 1991 (Tableau 2).

Quelques écarts minimes peuvent provenir

- du fait que le recensement prend en compte les ménages collectifs et les fonctionnaires internationaux.
- du fait que les différents échantillons du panel souffrent d'un biais systématique: les nouveaux immigrants n'ont pas été assimilés dans les enquêtes effectuées depuis 1985.

L'émigration n'a pas été compensée. Les autochtones sont largement sur-représentés, aussi bien chez les hommes que chez les femmes. Ce résultat était prévu. La comparaison avec les données du recensement le confirme. La prise en

compte des membres non-légitimes et l'extension de l'échantillon initial ne comblent pas cette carence.

4.3. Extension de la comparaison

La comparaison entre les trois versions de l'échantillon peut être étendue à 14 variables supplémentaires (Tableau 3).

Le parallélisme des distributions se confirme systématiquement.

Ceci signifie

- d'une part, que l'échantillon complémentaire (pondéré) a évolué de la même manière que l'échantillon initial du panel pondéré annuellement entre 1985 et 1991;
- d'autre part, que cette évolution semble conforme à l'évolution de la population (le phénomène d'immigration mis à part);
- et, enfin, que la prise en compte des membres non-légitimes n'introduit aucune perturbation dans le profil de l'échantillon.

SIMULATION:

Insertion dans l'échantillon initial et évaluation

4

Tableau 2

**COMPARAISON DES CARACTÉRISTIQUES DÉMOGRAPHIQUES
DES MEMBRES DU PANEL EN 1991:**

échantillon et recensement

	col.1	col.2	col.3	col.4
Echantillon	Echantillon initial membres légitimes	Echantillon initial membres légitimes et non-légit.	Echantillon initial et complément. membres légitimes et non-légit.	Recensement
Pondération	classique	poids partagés	poids partagés	
N=	4 046	4 590	5 354	384 633

Sexe

1. homme	49.0	49.1	49.2	49.1
2. femme	51.0	50.9	50.8	50.9

Age

1. < 18 ans	22.8	22.1	22.0	20.7
2. 18 - 24 ans	10.5	10.3	10.4	9.8
3. 25 - 34 ans	14.7	14.8	14.5	17.5
4. 35 - 44 ans	15.2	15.4	15.6	15.4
5. 45 - 54 ans	13.1	13.4	13.3	12.2
6. 55 - 64 ans	11.5	11.6	11.9	11.2
7. > 64 ans	12.2	12.4	12.3	13.2

Etat civil

1. célibataire	41.4	40.6	40.4	40.1
2. marié	48.7	49.0	49.4	48.5
3. séparés divorcés	3.2	3.3	4.1	3.9
5. veuf	6.6	7.0	6.0	7.5
6. N.R.	0.1	0.1	0.1	

Nationalité

1. Luxembourg.	82.0	81.9	82.1	70.0
2. autres C..E.E.	16.5	16.6	16.6	26.6
3. hors C.E.E.	1.4	1.5	1.3	3.4

4

SIMULATION:

Insertion dans l'échantillon initial et évaluation

Tableau 3

COMPARAISON DES CARACTÉRISTIQUES DÉMOGRAPHIQUES DES MEMBRES DU PANEL EN 1991:

échantillon initial et extensions

	col.1	col.2	col.3
Echantillon	Echantillon initial membres légitimes	Echantillon initial membres légitimes et non-légit.	Echantillon initial et complément. membres légitimes et non-légit.
Pondération	classique	poids partagés	poids partagés
N=	4 046	4 590	5 354

Caractéristiques personnelles

Age par sexe

1. Hommes			
1.1. < 18 ans	11.7	11.7	11.6
1.2. 18 - 64 ans	32.7	32.9	33.0
1.3. > 64 ans	4.6	4.5	4.6
2. Femmes			
2.1. < 18 ans	11.1	10.5	10.4
2.2. 18 - 64 ans	32.3	32.5	32.7
2.3. > 64 ans	7.6	7.9	7.7

Emploi

1. oui	40.9	40.9	40.3
2. non	33.7	34.3	35.1
3. enfants	25.4	24.8	24.6

Adultes/enfant

1. adulte	74.6	75.2	75.4
2. enfant	25.4	24.8	24.6

Lien avec le chef de ménage

1. chef de ménage	37.2	37.1	36.9
2. époux(se)	23.8	24.2	24.4
3. "époux(se)"	1.0	1.1	1.1
4. "ami(e)"	0.1	0.2	0.2
5. fils, fille	34.4	33.3	33.2
6. autres	3.5	4.1	4.2

Caractéristiques du ménage**Nombre de personnes dans le ménage**

1.	8.4	8.4	7.9
2.	20.4	20.4	20.2
3.	22.2	22.2	22.1
4.	31.3	31.3	31.7
5.	10.8	10.8	11.5
6.	5.4	5.4	4.8
7.	0.9	0.9	1.2
8.	0.6	0.6	0.6

Nombre d'enfants dans le ménage

0.	43.5	43.5	43.6
1.	21.7	21.7	21.9
2.	24.5	24.5	24.7
3.	7.4	7.4	7.2
4.	2.7	2.7	2.3
5.	0.3	0.3	0.3

Nombre d'adultes dans le ménage

1.	11.3	11.3	10.8
2.	58.6	58.6	58.9
3.	18.8	18.8	18.3
4.	8.7	8.7	9.0
5.	1.8	1.8	2.2
6.	0.5	0.5	0.6
7.	0.3	0.3	0.2

Nombre d'emplois dans le ménage

0.	16.4	16.4	16.7
1.	42.7	42.7	43.9
2.	30.9	30.9	29.2
3.	7.0	7.0	6.9
4.	2.5	2.5	2.7
5.	0.3	0.3	0.3
6.	0.1	0.1	0.2
7.	0.1	0.1	0.1

Caractéristiques du chef de ménage

Sexe

1. homme	85.8	85.8	86.5
2. femme	14.2	14.2	13.5

Age

1. < 24 ans	1.4	1.4	1.3
2. 25 - 34 ans	14.1	14.1	13.8
3. 35 - 44 ans	28.4	28.4	29.0
4. 45 - 54 ans	24.1	24.1	23.5
5. 55 - 64 ans	18.4	18.4	18.9
6. > 64 ans	13.6	13.6	13.5

Etat civil

1. célibataire	6.3	6.3	6.2
2. marié	79.6	79.6	81.1
3. séparé	1.2	1.2	1.1
4. divorcé	3.9	3.9	3.8
5. veuf	8.8	8.8	7.6
6. N.R.	0.2	0.2	0.2

Emploi

1. oui	71.2	71.2	70.9
2. non	28.8	28.8	29.1

Etat civil par Sexe

1. homme non-marié	6.9	6.9	6.9
2. homme marié	78.7	78.7	80.4
3. femme non-mariée	13.4	13.4	11.6
4. femme mariée	0.8	0.8	1.0
5. N.R.	0.2	0.2	0.1

Emploi par Sexe

1. homme sans emploi	20.9	20.9	21.6
2. homme avec emploi	64.9	64.9	64.9
3. femme sans emploi	7.8	7.8	7.5
4. femme avec emploi	6.4	6.4	6.0

CONCLUSION

Le problème de l'attrition peut être géré annuellement par des procédures de pondération classiques. Ces procédures présentent trois limites:

- elles ne permettent pas de prendre en compte les membres non-légitimes dont on ignore la probabilité de sélection,
- elles ne permettent pas de lutter contre l'érosion de l'échantillon,
- elles ne permettent pas de prendre en compte les nouveaux immigrants.

L'étude précédente comparait différentes solutions permettant de prendre en compte l'information apportée par les membres non-légitimes. Cette procédure n'a aucun effet sur l'érosion de l'échantillon. Cette seconde étude propose une solution plus efficace à cet égard. *Elle propose d'insérer un échantillon complémentaire de l'échantillon initial*, issu de la liste d'adresses qui a permis de sélectionner l'échantillon initial du panel.

Cette solution ne présente pas les mêmes avantages qu'un renouvellement de l'échantillon par l'insertion d'un *échantillon supplémentaire* tiré aléatoirement de la population de référence après quelques années d'enquêtes. Elle ne permet pas de prendre en compte les nouveaux immigrants entrés dans le pays depuis le tirage de l'échantillon initial.

Cette solution ne peut donc être qu'un palliatif provisoire.

Elle présente néanmoins certains avantages:

- elle restaure la taille de l'échantillon des ménages et des individus
- elle introduit dans l'échantillon des membres dont la présence et la durée de vie ne dépend pas des liens plus ou moins durables qu'ils ont noués avec les membres légitimes appartenant à l'échantillon initial: leur trajectoire dans le panel dépend de leurs caractéristiques personnelles
- elle préserve l'intérêt de l'observation et de l'analyse des catégories de population qui motivent l'étude
- elle permet de préserver le degré de précision des estimations statistiques.

Cette procédure ne pose pas le problème de l'accès aux fichiers administratifs. Elle exige néanmoins que certaines informations soient disponibles. L'histoire des

CONCLUSION

individus membres de cet échantillon complémentaire doit être reconstituée puisqu'ils n'ont pas été observés pendant une certaine période.

Ces informations doivent permettre

- de calculer la probabilité de sélection des ménages et des individus au moment de la sélection de la liste des adresses;
- de retracer la trajectoire de chaque individu et notamment de repérer et de distinguer les membres légitimes et non-légitimes qui sont entrés dans les ménages au cours de la période;
- de repérer la formation de nouveaux ménages au cours de cette période.

Les essais effectués dans le cadre du panel luxembourgeois semblent indiquer que l'insertion de cet échantillon complémentaire dans l'échantillon du panel ne modifie guère le profil des variables qui ont été testées.

On en tirera donc le bénéfice d'estimations plus précises puisque la taille de l'échantillon est accrue mais on ne résoud pas le problème de la prise en compte des nouveaux immigrants.

ANNEXE

On montre ici que l'estimation du poids des membres des ménages est équivalente lorsque les membres non-légitimes ne sont pas pris en compte et lorsqu'ils le sont. Seule, l'estimation de la valeur moyenne des variables se trouve modifiée parce que les valeurs des individus non-légitimes sont prises en compte et transférées aux individus légitimes.

Soit: w'_{ij} estimation du poids initial des individus j appartenant aux ménages i .

w'_{ij} = soit $1/p_j$ pour j^L où p_j est la probabilité de sélection de j^L
soit 0 pour j^{NL} et j^L est tout individu légitime
où j^{NL} est tout individu non-légitime.

Les poids initiaux des membres j du ménage i servent à calculer le poids final du ménage w_i :

$$w_i = 1/M_i \sum_{j=1}^{M_i} w'_{ij} = 1/M_i \sum_{j=1}^{M_i} 1/p_j \quad (1)$$

où M_i = le nombre d'individus membres du ménage
et M_i^L = le nombre d'individus légitimes membres du ménage.

Le poids final du ménage w_i est égal à la moyenne ($1/M_i \sum_{j=1}^{M_i}$) des poids des individus membres du ménage (w'_{ij}). La première équation est équivalente à la seconde: seuls les poids des membres légitimes (M_i^L) entrent dans le calcul de la somme des poids puisque les membres non-légitimes ont un poids de zéro; mais cette somme est divisée par le nombre total de membres du ménage (M_i), qu'ils soient légitimes ou non.

Le poids final du ménage (w_i) est ensuite assigné à tous les membres du ménage (légitimes ou non);

soit: $w_{ji} = w_j$ pour $j=1,\dots,M_i$

le poids de chaque individu est égal au poids du ménage pour tous les membres du ménage.

Pour estimer un total tel que Y , on utilise

$$Y = \sum_{i=1}^N \sum_{j=1}^{M_i} w_{ij} Y_{ij} \quad (2)$$

l'estimation du total Y est égale à la somme sur l'ensemble des ménages de la somme sur l'ensemble des individus appartenant à chaque ménage de leur valeur pondérée.

Si k est une autre manière de désigner l'ensemble des individus (légitimes ou non) on peut aussi écrire:

$$Y = \sum_{i=1}^N \sum_{j=1}^{M_i} (1/M_i \sum_{k=1}^{M_i} w'_{ik}) Y_{ij} \quad (3)$$

où les j sont remplacés par les k et les w_{ij} de l'équation (2) remplacés par leur définition donnée dans l'équation (1) ... puisque $w_{ij} = w_i$ (les poids des individus appartenant à chaque ménage sont égaux au poids du ménage).

Puisque la valeur entre parenthèses est indépendante de $\sum_{j=1}^{M_i}$ (dans l'équation (3)), on peut écrire de manière équivalente:

$$Y = \sum_{i=1}^N 1/M_i \sum_{k=1}^{M_i} w'_{ik} \sum_{j=1}^{M_i} Y_{ij}$$

ou encore

$$Y = \sum_{i=1}^N \sum_{k=1}^{M_i} w'_{ik} [1/M_i \sum_{j=1}^{M_i} Y_{ij}]$$

où la valeur entre les crochets correspond exactement à $\bar{Y}_i$, c'est à dire la valeur moyenne de Y . Soit:

$$Y = \sum_{i=1}^N \sum_{k=1}^{M_i} w'_{ik} \bar{Y}_i \quad (4)$$

Puisque j comprend tous les membres (légitimes ou non) du ménage, on voit que les valeurs des membres non-légitimes influencent la valeur moyenne de Y_i .

Par contre, w'_{ik} ne peut prendre que deux valeurs. Puisque $w'_{ik} = w'_{ji}$:

w'_{jk} = soit $1/p_j$ pour j^L où p_j est la probabilité de sélection de j^L
 et j^L est tout individu légitime
 soit 0 pour j^{NL} où j^{NL} est tout individu non-légitime.

L'équation (4) devient donc:

$$Y = \sum_{i=1}^n \sum_{k=1}^{M_i L} 1/p_i \bar{Y}_i.$$

La valeur moyenne de Y_i , enrichie de l'information apportée par les individus non-légitimes, ne concerne que les individus légitimes puisque les membres non-légitimes ont toujours un poids nul ($w'_{jk} = 0$ pour j^{NL}).

L'estimation du poids des membres des ménages est équivalente que les membres non-légitimes soient pris en compte ou non. CQFD.

Liste des publications du panel socio-économique des ménages

"Liewen zu Lëtzebuerg"

CEPS/Instead, Walferdange, Grand-Duché de Luxembourg

Série: *Techniques et Méthodologie*

Méthodologie générale & répertoire des variables - Année d'enquête: 1985 (Première vague). Document PSELL n° 1 - P.DICKES, P.HAUSMAN, A. KERGER -1987.

Pratique de l'échelonnement multidimensionnel. Document PSELL n° 7. P.DICKES, J.TOURNOIS (1989).

Logistique & documentation - Principes d'organisation de la documentation dans le panel. Document PSELL n° 9 J. TOURNOIS (1988).

Documentation transversale des variables 1985: première vague. Document PSELL n° 10. J. TOURNOIS (1988).

Description statistique des variables du questionnaire -1986- (deuxième vague). Document PSELL n°12. A. KERGER, R. DE WEVER (1988).

Le mode d'échantillonnage du panel "Liewen zu Lëtzebuerg" Bilan des deux premières vagues. Document PSELL n°14. P.HAUSMAN (1990).

La collecte des données en 1986. Elaboration du questionnaire, déroulement de l'enquête, opérations de chiffrement. Document PSELL n° 16. A.KERGER (1989).

Organisation der Daten des Luxemburger Haushaltspanels. (Eingabe, Speicherung und Analyse von Paneldaten). Document PSELL n° 17. G.SCHMAUS (1990) (version anglaise: 17a).

NDR, partition évaluée selon la méthode de Roubens et Libert. Document PSELL n° 18. B.GAILLY (1989).

Dispositif des pondérations individuelles et des pondérations des ménages en 1985 et 1986. Document PSELL n°19; B.GAILLY, P.HAUSMAN - 1990.

Examen des effets du phénomène d'attrition sur l'étude des revenus et de l'emploi - Années de références: 1985, 1986 et 1987. Document PSELL n° 23 P.HAUSMAN, B.GAILLY (1990).

La constitution des fichiers de référence, nécessaire à l'étude du phénomène d'attrition. Document PSELL n° 24 R.DE WEVER (1990).

Dispositif des pondérations individuelles et des pondérations des ménages en 1985 et 1987. Document PSELL n° 25. B.GAILLY, P.HAUSMAN (1990).

Bilan de l'attrition au cours des trois premières vagues d'enquêtes: 1985/ 1986/ 1987. Document PSELL n° 26 B.GAILLY, P.HAUSMAN (1990).

Imputation des revenus manquants dans le panel socio-économique luxembourgeois. Document PSELL n°27 P.HAUSMAN (1990).

"PSELLDOC" Système documentaire pour le panel Luxembourgeois. Document PSELL n°28. J.J.WESTER avec la collaboration de A. AUBRUN (1990).

Le déroulement de la collecte en 1987. Elaboration du questionnaire, déroulement de l'enquête, opération de chiffrement. Document PSELL n°29. A.KERGER (1990).

La production des données: Vague 1988-1990. Document PSELL n° 30. A.KERGER (1990).

Description statistique des variables du questionnaire -1987- Troisième vague. Document PSELL n° 31. R.DE WEVER, A.KERGER (1991).

Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1988. Document PSELL n° 40. B.GAILLY (1991).

Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1989. Document PSELL n°48. B.GAILLY (1992).

Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1990. Document PSELL n°53. B.GAILLY (1993).

Insérer des nouveaux membres dans un panel longitudinal de ménages et d'individus: simulations. Document PSELL n°54. B.GAILLY (CEPS/Instead) avec la collaboration de P.LAVALLÉE (Statistics-Canada) (1993).

Dispositif des pondérations individuelles et des pondérations des ménages de 1985 à 1991. Tome II. Document PSELL n°55. B.GAILLY (1993).

Insérer un échantillon complémentaire dans un panel longitudinal d'individus et de ménages: simulations - 2e partie. Document PSELL n°59. B.GAILLY (CEPS/ Instead), P.LAVALLÉE (Statistique-Canada) (1994).

*