

**PANEL SOCIO-ECONOMIQUE  
"LIEWEN ZU LETZEBUERG"**

---

Document PSELL n° 23

**Examen des effets du phénomène  
d'attrition sur l'étude des  
revenus et de l'emploi**

Années de référence: 1985, 1986 et 1987

P. HAUSMAN  
B. GAILLY

---

Document produit par le

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

**C.E.P.S./INSTEAD  
B.P.65 L-7201 WALFERDANGE  
Tél. (352) 33 32 33 - 1**

**Président: Gaston Schaber**

---

1990

**PANEL SOCIO-ECONOMIQUE  
"LIEWEN ZU LETZEBUERG"**

---

Document PSELL n° 23

**Examen des effets du phénomène  
d'attrition sur l'étude des  
revenus et de l'emploi**

Années de référence: 1985, 1986 et 1987

P. HAUSMAN  
B. GAILLY

---

Document produit par le

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

**C.E.P.S./INSTEAD  
B.P.65 L-7201 WALFERDANGE  
Tél. (352) 33 32 33 - 1**

**Président: Gaston Schaber**

---

1990

## Introduction

L'objectif de l'étude présentée dans ce document concerne les conséquences possibles d'un phénomène normal dans une recherche longitudinale, phénomène que l'on désigne par le terme "attrition".

Dans ce type de recherche, l'échantillon observé lors de la première enquête (première vague) va subir des modifications au fil des années;

*par exemple, une partie des personnes interrogées au cours de la première vague ne le sera plus au moment de la seconde, pour différentes raisons et, inversement, des personnes qui ne figuraient pas dans l'échantillon de la première vague, apparaissent dans celui de la seconde vague.*

Lorsque l'on parle d'attrition, on se réfère à ce phénomène d'usure de l'échantillon.

Le chercheur qui s'intéresse, par exemple, à l'analyse diachronique des revenus ne peut ignorer l'existence de ce phénomène et, surtout, les conséquences que celui-ci peut entraîner au niveau de la qualité des travaux envisagés.

*\* l'évolution de l'échantillon initial ne rend-t-elle pas caduque l'analyse de certains groupes dont l'effectif se serait fortement réduit au fil du temps?  
\* certaines variables centrales pour la réalisation d'une étude ne sont-elles pas biaisées suite aux déformations subies par l'échantillon initial?*

Les réponses à ces questions sont indispensables pour l'analyste. Il faut même considérer que les informations relatives à la déformation globale de l'échantillon font partie du bagage de tout analyste qui souhaite exploiter les données d'un panel.

Et, même si cet analyste ne s'intéresse qu'à une problématique spécifique -comme celle des revenus- il ne devrait pas s'engager dans une telle voie sans une connaissance minimale des mécanismes qui, dans une étude longitudinale, vont éventuellement modifier la composition de certains groupes.

L'équipe de recherche, responsable du panel, connaît les problèmes posés par ce phénomène d'attrition. C'est d'ailleurs pour cela que sont menés, chaque année, des travaux longs et précis au cours desquels ce phénomène est examiné sous différents aspects.

A l'issue de ces travaux, une série de paramètres sont proposés afin de gérer les conséquences liées à l'usure inévitable de l'échantillon.

Le système de pondérations adopté dans le cadre du programme PSELL exclut volontairement certaines variables. Il n'existe pas

de mesure spécifique visant à corriger les défections observées éventuellement dans certaines tranches de revenus: de telles mesures comporteraient trop de risques non- maîtrisés , même à moyen terme et se heurteraient en outre à la multiplicité des choix possibles; ainsi:

- *quel type de revenu faudrait-il sélectionner?*
- *faut-il préférer le revenu global du ménage au revenu exprimé par unité de consommation?*
- *doit-on envisager l'ensemble des revenus cumulés au niveau du ménage ou les revenus individuels?*

Selon le choix adopté dans le programme PSELL, les revenus n'interviennent pas directement dans la correction des effets de l'usure de l'échantillon. Cela ne signifie pas que le système de pondérations soit sans effet sur l'évolution de certaines catégories de revenus. Mais nous ignorons de quelle manière et dans quelle mesure il influence la mesure de la distribution des revenus.

L'objectif de ce rapport sera précisément d'éclairer les analystes sur l'état de ce problème au cours des trois premières vagues du panel luxembourgeois (1985, 1986, 1987).

La démarche s'appliquera ici à deux thèmes d'étude: les revenus et l'emploi; elle pourra servir de référence pour les examens qui seraient nécessaires pour d'autres thèmes de recherche.

**PREMIERE PARTIE**

**Le phénomène de l'attrition et  
l'étude des revenus.**

## Plan de la démarche

### Les variables de base

L'examen proposé porte sur les deux mesures du revenu suivantes :

- \* le revenu net disponible du ménage [RD1]
- \* le revenu net disponible par équivalent-adulte ou unité de consommation [RDIFA1] dans le ménage; ce revenu a été ajusté à la taille du ménage par l'application d'une échelle d'équivalence affectant les poids suivants aux membres du ménage :
  - 1.0 pour le premier adulte
  - 0.7 pour tout adulte supplémentaire
  - 0.5 pour chaque enfant, présent dans le ménage.

Les valeurs de ces deux mesures du revenu ont été attribuées aux individus interrogés au cours de la première vague [1985]. Chacun des 6110 individus de cet échantillon est donc caractérisé par ces deux variables; les individus appartenant à un même ménage reçoivent une valeur de revenu identique.

L'échantillon des individus fait ensuite l'objet d'un décilage selon chaque type de revenu [une précaution particulière a été prise afin d'éviter de classer les membres d'un même ménage au sein de deux déciles différents].

Les travaux présentés dans ce rapport portent donc sur les déciles du revenu net disponible du ménage [PDRD1] ainsi que sur les déciles du revenu net disponible par unité de consommation (U.C) [PDRDIFA1].

### Les objectifs

Nous partons de la distribution des revenus par déciles de l'échantillon de la première vague [1985].

Cet échantillon connaît, au cours du temps, une série de déformations dont les causes principales peuvent être résumées comme suit :

- \* des individus ont refusé de répondre au questionnaire de la seconde vague et/ou de la troisième vague [1];
- \* des individus n'ont pas pu être localisés pour répondre au questionnaire de seconde et/ou troisième vague(s) [2];
- \* des individus sont décédés entre les vagues d'enquêtes [3];
- \* des individus sont partis à l'étranger [4];
- \* de nouveaux individus apparaissent dans l'échantillon au cours de la seconde ou troisième vague, qui n'étaient pas présents au cours de la vague précédente [5].

Si l'on examine l'évolution de notre échantillon sur une période relativement courte, les motifs n°3, 4 et 5 peuvent être considérés comme des mouvements "naturels" qui se produisent aussi au sein de la population-mère de notre échantillon. Par conséquent, ils n'affectent pas la représentativité de l'échantillon.

Par contre, les départs selon les motifs n°1 et 2 peuvent entraîner des biais qu'il conviendra, le cas échéant, de redresser grâce au système de pondération.

Il s'agit donc de répondre à un certain nombre de questions:

- *Comment se répartissent les effets de ces différents mouvements de transformation observés dans l'échantillon au cours du temps et selon les déciles du revenu établis en 1985 ?*
- *Quels sont les effets du système de pondération à cet égard?*

En dernière instance, cette approche doit déterminer dans quelle mesure un échantillon cylindré sur plusieurs vagues constitue encore le "prolongement" de l'échantillon initial complet. La conclusion de cet examen éclairera donc l'analyste sur les conditions dans lesquelles il pourra pratiquer une étude diachronique sur les revenus.

Le premier chapitre traitera ces problèmes au niveau des deux premières vagues ( 85-86).

Le second chapitre prolongera cet examen sur une période couvrant trois vagues (85,86, 87).

---

1. Dans cette catégorie, seules les naissances peuvent être assimilées à des mouvements naturels (enfants nés d'un parent considéré comme full-member panel, c.à.d. d'une personne présente dans l'échantillon initial).



## Chapitre 1

### Examen de l'effet de l'attrition sur les deux mesures du revenu

#### 1.1 Organisation de la démarche

L'objectif poursuivi dans ce premier chapitre peut être résumé par la question suivante :

- Dans quelle mesure, les taux de réponses enregistrés au cours de la seconde vague d'observation diffèrent-ils selon les catégories de revenus définies au cours de la première vague ?

Pour répondre à cette question, la procédure suivante a été retenue :

1° Tous les individus de la première vague sont, tout d'abord, classés selon les déciles des deux mesures du revenu sélectionnées pour ce test :

- le revenu disponible de leur ménage
- le revenu disponible de leur ménage exprimé par équivalent-adulte.

2° La distribution des taux de réponses enregistrés au cours de la seconde vague d'observation sont ensuite présentés en fonction des revenus décilés observés au cours de la première vague.

3° Les non-réponses recouvrent plusieurs types de phénomènes. La répartition de chaque catégorie de non-réponses sera examinée en fonction des déciles des deux mesures du revenu.

4° Ces analyses étant pratiquées sur l'ensemble de l'échantillon des individus "membres du panel" en 1985, il convient de corriger le biais de sélection initial qui affecte cet échantillon (voir à ce propos : PSELL n°19, la construction de la variable WGT85).

5° Un dernier examen sera appliqué au seul sous-échantillon des personnes interrogées au cours des deux premières vagues d'enquêtes; ce type d'analyse exige que l'échantillon soit pondéré. Le système de pondération développé pour la seconde vague tient compte de deux facteurs (cf. PSELL n°19, variable WGT86) :

- les probabilités de sélection initiales,
- et l'ajustement différentiel des non-réponses entre les deux premières vagues.

Cette approche permettra d'évaluer l'effet du système de pondération sur les variables de revenus.

## 1.2 Bilan de l'attrition 85-86

Au cours de la première vague du panel, l'échantillon comprenait 6110 personnes.

De la première à la seconde vague, cet échantillon a subi un certain nombre de modifications :

- \* de nouveaux individus apparaissent en 1986
- \* des individus présents en 1985 ne figurent plus dans l'échantillon interviewé lors de la seconde vague.

L'arrivée de nouveaux individus au cours de la seconde vague constitue un aspect fort intéressant, mais il ne sera pas envisagé dans le cadre de cette étude.

L'usure de l'échantillon représente un phénomène plus préoccupant; sachant que la taille de l'échantillon de la première vague va diminuer au fil des années, quelles seront les conséquences de ce phénomène au niveau des principaux types d'analyses diachroniques.

Le tableau suivant présente l'état dans lequel se trouve l'échantillon initial après la seconde vague d'observation. Comment les différentes trajectoires des individus ont-elles affecté leur distribution initiale sur la variable des revenus?

Table 1.1: BILAN DES NON-REponses LORS DE LA SECONDE VAGUE : ETAT DE L'ECHANTILLON INTERVIEWE EN 1986

Diagnostic 85-86	Echantillon brut		Echantillon pondéré*	
	C.A	%	%	C.A
- individus présents en 85 et 86	5167	84.6	85.0	5192
- refus d'enquête en 1986	833	13.6	13.2	804
- départs naturels (décès- émigrés)	110	1.8	1.8	114
TOTAL échantillon 85	6110	100.0	100.0	6110

(\*) Echantillon 1985, corrigé du biais de sélection initial [variable de pondération : WGT85]

Selon la variable "diagnostic 85-86", seuls 85 % des personnes constituant l'échantillon original ont été interrogées en 1986 (réf. échantillon pondéré).

Les "non-réponses" correspondent donc à 15% de cet échantillon originel.

### 1.3 Examen du taux de survie [85-86] dans les déciles du revenu

Table 1.2: TAUX DE SURVIE SELON LES DECILES DU REVENU DISPONIBLE ET DU REVENU DISPONIBLE PAR UNITE DE CONSOMMATION.

Déciles du revenu disponible global	Diagnostic 85-86 simplifié		Déciles du revenu disponible par U.C	Diagnostic 85-86 simplifié		Total échantillon
	Présents 85-86	Disparus 86		Présents 85-86	Disparus 86	
	% col.1	% col.2		% col.3	% col.4	
1er décile	8.9	16.0	1er décile	9.5	12.8	10.0
2e décile	10.2	8.8	2e décile	10.5	7.3	10.0
3e décile	10.4	7.9	3e décile	10.0	9.8	10.0
4e décile	10.4	7.8	4e décile	10.5	7.0	10.0
5e décile	9.8	10.9	5e décile	10.3	8.1	10.0
6e décile	10.3	8.5	6e décile	9.6	12.8	10.0
7e décile	9.8	10.6	7e décile	9.7	11.3	10.0
8e décile	10.7	6.3	8e décile	9.9	10.7	10.0
9e décile	9.5	12.9	9e décile	10.0	10.0	10.0
10e décile	9.9	10.3	10e décile	9.9	10.1	10.0
Ensemble (éch. pondéré WGT85)	100.0	100.0		100.0	100.0	100.0

(Coeff. cont. = 0.11 ; signif. à p = 0.001)

(coeff. cont. = 0.08 ; signif. à p = 0.001)

Ce tableau montre de quelle façon le décilage des revenus (établi en 1985) peut être affecté par l'évolution de l'échantillon initial après la seconde vague d'enquêtes.

En théorie, chaque colonne devrait être composée de 10 modalités contenant chacune un effectif égal à 10 % de l'échantillon de référence.

Si l'érosion de l'échantillon affectait les déciles de manière uniforme, le tableau montrerait cette absence totale de biais et les effectifs resteraient répartis également dans tous les déciles. Inversement, l'écart entre les pourcentages présentés dans les différentes colonnes et le seuil de 10 % indique l'importance relative du biais introduit dans chaque décile par l'usure de l'échantillon.

L'examen de la colonne '1' permet de répondre à la question suivante: le sous-groupe des personnes interrogées en 1985 ET en 1986 (n = 5192) est-il réparti sur la variable "revenus décilés" conformément à la distribution de l'ensemble de l'échantillon de la première vague (n= 6110)?

La colonne '3' permet de répondre à la même question appliquée aux revenus disponibles par unité de consommation.

Selon les % portés dans les colonnes 1 et 3, il apparaît que l'évolution de l'échantillon modifie peu l'équilibre du décilage établi en 1985; en effet, la répartition de l'effectif des individus présents en 1985 et 1986 n'est guère éloignée des 10% attendus.

On notera uniquement que, dans le groupe des personnes observées deux fois, en 1985 et en 1986, le premier décile est légèrement sous-représenté. Ce phénomène touche les deux variables de revenu bien qu'il soit moins aigu dans le cas du revenu disponible par unité de consommation.

Les déciles du revenu disponible par unité de consommation ne sont d'ailleurs pratiquement pas affectés par l'évolution de l'échantillon; il s'agit là d'un résultat plutôt satisfaisant.

La répartition des individus qui n'ont pas été interrogés deux fois traduit les mêmes tendances. L'examen des colonnes '2' et '4' conduit aux conclusions suivantes:

- les défections sont sur-représentées dans le premier et le neuvième décile du revenu disponible;
- elles sont sur-représentées dans le premier et le sixième décile du revenu disponible par unité de consommation.

#### Premier bilan

A la suite de ce premier examen, nous pouvons déjà dégager quelques éléments susceptibles d'éclairer les conséquences de l'attrition sur l'étude des revenus.

Si l'on se place d'un point de vue très général, on peut affirmer que l'évolution de l'échantillon n'a pratiquement aucun effet sur les deux variables du revenu considérées; le faible niveau des deux coefficients de contingence le confirme. Une approche plus détaillée montre cependant que les pertes enregistrées au cours de la seconde vague concernent un peu plus souvent des individus appartenant au premier décile et cette tendance est légèrement plus accentuée dans le cas du revenu disponible global (16% des individus de l'échantillon de 1985 qui ne sont pas interrogés en 1986 appartiennent au premier décile du revenu disponible contre 12.8% pour le revenu disponible par unité de consommation).

Ce premier bilan met en évidence une situation qui n'est, en aucun cas, de nature à compromettre les analyses en matière de revenus.

Le léger déficit enregistré au niveau du premier décile ne se révèle pas significatif. Il ne compromet pas l'étude de la pauvreté; en effet, le revenu disponible par unité de consommation - qui mesure avec plus de précision le niveau de vie - est à peine affecté par l'évolution de l'échantillon observée au cours des deux premières vagues.

La section suivante envisagera, de façon plus nuancée, la répartition du sous-groupe des personnes membres de l'échantillon de 1985 qui n'ont pas été interrogées l'année suivante.

## 1.4 Différenciation des non-réponses en 1986

Le tableau suivant permet de se faire une idée plus précise des mécanismes qui ont infléchi le taux de réponses ainsi que la répartition de l'échantillon subsistant selon les déciles du revenu.

Table 1.3: REPARTITION DES CATEGORIES DE NON-REPONSES OBSERVEES EN 1986, SELON LES DECILES DU REVENU (ECH. PONDERE)

Déciles du revenu disponible global	Diagnostic 85-86			Déciles du revenu disponible par U.C	Diagnostic 85-86			Total échantillon
	Présents 85-86 (1)	Refus	Départs naturels		Présents 85-86(1)	Refus	Départs naturels	
	col.1 %	col.2 %	col.3 %		col.4 %	col.5 %	col.6 %	
1er décile	8.9	<u>14.4</u>	<u>27.6</u>	1er décile	9.5	10.9	<u>25.8</u>	10.0
2e décile	10.2	7.9	15.3*	2e décile	10.5	6.9	10.4*	10.0
3e décile	10.4	7.9	7.8*	3e décile	10.0	8.8	16.8*	10.0
4e décile	10.4	7.4	10.4*	4e décile	10.5	7.2	5.2*	10.0
5e décile	9.8	11.3	7.4*	5e décile	10.3	8.4	5.9*	10.0
6e décile	10.3	9.2	4.1*	6e décile	9.6	<u>14.0</u>	4.1*	10.0
7e décile	9.8	11.3	5.5*	7e décile	9.7	12.1	5.6*	10.0
8e décile	10.7	6.5	5.1*	8e décile	9.9	11.5	5.4*	10.0
9e décile	9.5	<u>13.4</u>	9.3*	9e décile	10.0	9.9	10.9*	10.0
10e décile	9.9	10.7	7.5*	10e décile	9.9	10.2	10.0*	10.0
Ensemble (éch. pondéré WGT85)	100.0 (n=5167)	100.0 (n=833)	100.0 (n=110)		100.0 (n=5167)	100.0 (n=833)	100.0 (n=110)	100.0 (n=6110)

(Coeff.cont. = 0.13 )

(coeff.cont.= 0.12)

(\*) moins de 20 unités

(1) Comme l'information consignée dans les colonnes '1' et '4' a déjà été commentée dans la section précédente (cfr colonnes '1' et '3' du tableau '2'), nous n'en ferons donc plus état ici.

Les colonnes 2-3 et 6-7 fournissent les éléments de base permettant d'analyser de manière plus précise les mécanismes qui sont à l'origine de la réduction de l'effectif de l'échantillon initial (1985).

Les "départs naturels" représentent quelque 2% de l'échantillon de la première vague du panel. Il s'agit là d'une catégorie d'événements que l'on peut qualifier de rares; ceci n'exclut pas que ces départs puissent jouer un rôle déterminant dans l'évolution de certaines zones de l'échantillon initial.

C'est effectivement ce que montre l'examen des colonnes '3' et '6'. Les départs "naturels", d'ordre démographique, ne s'effectuent pas au hasard: ils touchent plus particulièrement le premier décile (du RD et RD/UC<sup>2</sup>).

2. revenu disponible global et revenu disponible par unité de consommation

Dans le premier décile de l'échantillon selon le revenu disponible en 1985, les départs naturels sont 2.7 fois plus nombreux que ceux que l'on attendrait dans l'hypothèse où seul le hasard interviendrait dans leur répartition (ce rapport est égal à 2.6 dans le cas du premier décile selon le RD/UC).

Tous les autres déciles semblent relativement moins concernés par ces départs naturels (compte tenu, surtout, du fait qu'en chiffres absolus, ces départs représentent des effectifs très faibles n'excédant jamais 20 cas).

La mise-en-évidence de cette distribution très inégalitaire des départs naturels permet de mieux comprendre l'évolution observée dans le premier décile du revenu disponible et du revenu disponible par unité de consommation.

La répartition des départs naturels explique, en effet:

- toute la réduction de l'effectif du premier décile selon le RD/UC;
- une partie de la réduction de l'effectif du premier décile selon le revenu disponible.

Dans ce dernier cas, les départs naturels ne sont pas seuls en cause. Ils ajoutent leurs effets à l'influence des "refus" pour déterminer la réduction différentielle de la composition du premier décile dans la seconde vague.

L'intervention très localisée de cette composante de l'attrition (décès/émigration) fournit un élément capital pour notre étude. Bien que les écarts enregistrés au niveau du premier décile soient encore minimes (surtout pour le revenu disponible par unité de consommation), nous sommes néanmoins en mesure de préciser que ces écarts ne représentent pas des biais liés à certains événements spécifiques au dispositif d'enquête, mais correspondent à des modifications que l'on peut aussi observer dans l'ensemble de la population.

Dans ce cas précis, on postule qu'une modification enregistrée dans l'échantillon n'affecte pas sa représentativité lorsque cette modification peut aussi être observée dans la population-mère de l'échantillon.

C'est dans cette perspective que nous pouvons interpréter la dégradation de l'effectif observé dans le premier décile du revenu disponible et du revenu disponible par unité de consommation. Cette usure de l'échantillon n'introduit pas en elle-même un biais; elle correspond à un phénomène qui s'est aussi produit dans l'ensemble de la population.

Cette dernière remarque suggère que l'analyse de l'évolution de l'échantillon doit prendre en compte d'une part, les personnes présentes en 85 et 86 et, d'autre part, les refus de répondre. Les départs naturels ne doivent pas être pris en compte. (L'effectif de l'échantillon est porté à 5994 sujets. Les poids des personnes doivent être ajustés afin que la somme des poids reste égale à la taille de l'échantillon).

Dans ces conditions la valeur du coefficient de contingence passe de 0.13 à 0.10 pour le test concernant le revenu disponible et de 0.12 à 0.08 pour le revenu disponible par unité de consommation. Il n'y a pratiquement aucune relation entre les refus enregistrés en 1986 et les deux décilages de l'échantillon selon les revenus.

L'examen des col.'2' et '5' confirment globalement cette conclusion. Toutefois quelques écarts peuvent être notés dans la répartition des refus:

- déciles 1, 4, 8, et 9 pour le revenu disponible
- déciles 2, 4, 6 pour le revenu disponible par unité de consommation.

Les écarts relatifs les plus importants peuvent être observés dans les mêmes déciles, pour les personnes ayant participé aux deux vagues d'enquêtes. Toutefois, l'écart le plus important entre le taux de survie observé et le taux de survie théorique est de l'ordre de 1%; tous les autres écarts sont inférieurs à 0.5%.

## 1.5 L'effet des pondérations de la seconde vague sur les déciles du revenu

Lorsque les "départs naturels" ne sont plus pris en compte, l'association entre les trajectoires des personnes (variable "diagnostic 85-86") et le décilage de l'échantillon selon les deux mesures du revenu (RD et RD/UC) tend à se réduire.

Néanmoins, l'examen des refus de répondre enregistrés en 1986 indique une répartition relativement inégalitaire de ceux-ci parmi les déciles. Ces écarts sont minimes, mais ils sont susceptibles d'introduire des biais dans les analyses des revenus appliquées aux personnes présentes en 1985 et en 1986.

Le système de pondération mis au point pour la seconde vague du panel (cfr. PSELL n°19) est précisément conçu dans le but de maintenir la "représentativité" d'un tel échantillon cylindré; en d'autres termes, ce système de pondération s'efforce de réduire les effets de l'évolution de l'échantillon dans la mesure ou cette évolution est liée au phénomène des refus de répondre enregistrés au cours de la seconde vague d'enquêtes.

Quel est l'impact de ce système de pondération sur les variables du revenu, sachant que ces variables n'ont pas été prises en compte dans l'élaboration de ce système de pondération? (Cfr. liste des variables utilisées dans le système de pondération de la seconde vague, in: PSELL n°19).

Le tableau suivant permet de comparer les effets des différents éléments du dispositif de pondération sur l'échantillon des personnes présentes dans le panel en 1985 et 1986.

Table 1.4: COMPARAISON<sup>3</sup> DES EFFETS DE CHAQUE MODE DE PONDERATION SUR LES DECILES DU REVENU<sup>4</sup>

Répartition de l'échantillon des présents 85 et 86							
Déciles selon le revenu disponible	Hors pondération	Pondération= biais de sélection initial 1e vague: 85	Pondération 2e vague: 86	Déciles selon le revenu disponible par U.C	Hors pondération	Pondération = biais de sélection initial 1e vague:85	Pondération 2e vague:86
	Col.1. %	Col.2. %	Col.3. %		Col.4. %	Col.5. %	Col.6. %
1er décile	6.9	8.9	9.3	1er décile	8.1	9.5	9.4
2e décile	7.7	10.2	10.3	2e décile	8.9	10.5	10.2
3e décile	8.3	10.4	10.4	3e décile	9.2	10.0	9.8
4e décile	8.5	10.4	10.3	4e décile	10.8	10.5	10.5
5e décile	8.2	9.8	9.7	5e décile	10.4	10.3	10.2
6e décile	10.3	10.3	10.3	6e décile	10.4	9.6	9.7
7e décile	10.0	9.8	9.8	7e décile	10.9	9.7	9.7
8e décile	12.6	10.7	10.5	8e décile	11.3	9.9	10.1
9e décile	13.1	9.5	9.5	9e décile	10.6	10.0	10.2
10e décile	14.3	9.9	9.9	10e décile	9.3	9.9	10.1
Ensemble*	100.0	100.0	100.0		100.0	100.0	100.0

\* n= 5167

3. cette comparaison est réalisée sur l'échantillon des personnes présentes en 1985 et 1986

4. ces déciles ont été établis sur l'ensemble de l'échantillon observé en 1985

Il est important de rappeler que cette comparaison s'applique uniquement aux personnes qui ont été interrogées à la fois en 1985 et en 1986 (soit: n=5167) et non à l'ensemble de l'échantillon observé en 1985.

Les déciles de revenus ont été établis sur la base de l'ensemble de l'échantillon observé en 1985.

Les colonnes '1' et '4' présentent la répartition de l'échantillon cylindré (85-86) avant l'intervention de toute forme de pondération; il s'agit donc d'un échantillon "brut".

Les colonnes '2' et '5' montrent de quelle façon la correction du biais de sélection initial (1985) modifie la répartition des personnes ayant répondu à deux reprises, selon les déciles des deux revenus. Ces pourcentages ont déjà été présentés dans le tableau n°3.

La comparaison des résultats présentés dans les colonnes 2-3 et dans les colonnes 5-6 sont particulièrement intéressants.

La colonne '3' et la colonne '6' présentent la répartition par déciles lorsque l'échantillon est contrôlé par le système de pondération en vigueur pour la seconde vague: la variable de pondération utilisée en 1986 prend en compte, à la fois, le biais de sélection initial (1985) et les corrections effectuées suite aux refus enregistrés au cours de la seconde vague (1986).

La conclusion est claire: la variable de pondération "WGT86" ne modifie pas la répartition de l'échantillon cylindré. La correction du biais de sélection initial (WGT85) redresse légèrement l'échantillon mais la variable de pondération calculée pour la seconde vague (WGT86) n'intervient pas malencontreusement là où il n'y a plus lieu d'intervenir.

En d'autres termes:

\* le biais de sélection initial était corrélé avec les variables du revenu: la correction de ce biais (WGT85) (col. '2' et col. '5') rétablit l'équiprobabilité entre les déciles.

\* les refus de répondre en 1986 ne modifient pas sensiblement la distribution des personnes dans les déciles et l'effet de correction de la variable de pondération calculée en 1986 tient essentiellement au fait que cette variable prend en compte et redresse le biais de sélection initial.



## Chapitre 2

### L'attrition 1985/1987 et les mesures du revenu

#### 2.1 INTRODUCTION

L'usure de l'échantillon ne semble pas entraîner des conséquences fâcheuses pour l'analyse des revenus des personnes qui "survivent" dans le panel après deux vagues d'enquête. Le chapitre précédent a été consacré à ce premier point.

Cette conclusion est particulièrement importante puisque la première transition (85-86) est un moment critique dans la vie d'une étude longitudinale: les refus de répondre sont particulièrement nombreux à ce moment. D'autres travaux montrent effectivement que les refus de répondre ont tendance à diminuer au fil des vagues d'enquêtes: le taux d'attrition s'élève à 13.4% au cours de la vague d'enquêtes de 1986; mais la réduction de l'échantillon suite aux refus a été progressivement contenue au fil des années; c'est ainsi qu'en 1987, les pertes consécutives aux refus ne représentent plus que 11% de l'ensemble des enquêtes distribuées (ce taux spécifique d'attrition est égal au rapport "refus+introuvables" / "total enquêtes distribuées" (hors départs naturels); ce taux était égal à 35% lors de la première vague et à 13% pour la seconde vague).

La fidélité des personnes à l'enquête s'accroît encore par la suite. Le taux "spécifique" de refus ne s'élève plus qu'à 6.2% des enquêtes distribuées (hors départs naturels) lors de la quatrième vague (1988); et, lors de la campagne suivante (1989), ce taux de refus descend à 5.6%.

Ces différents éléments semblent indiquer que les biais introduits par les refus risquent de diminuer "quantitativement" au fil du temps.

Il n'en reste pas moins que l'échantillon s'use. Cette usure est, partiellement, compensée par l'arrivée de nouveaux membres dans le panel; mais nous avons délibérément laissé de côté ce processus de régénération afin de nous placer dans la perspective la plus exigeante, c'est à dire la perspective d'un échantillon "cylindré" sur plusieurs années où toute modification correspond à une "réduction" de l'effectif des personnes observées.

Dès que cette usure de l'échantillon ne revêt plus un caractère aléatoire, le système de pondération s'efforce de sauvegarder la "représentativité" de l'échantillon "subsistant" par rapport à l'échantillon originel.

Cette notion de "représentativité" doit être bien comprise: quelle que soit la méthode utilisée, ce caractère de "représentativité" ne peut couvrir tout le champ d'observation. Cette représentativité ne peut être assurée -à la limite- que pour les variables qui sont prises en compte dans le système de pondération et pour les autres variables qui leur sont associées de manière significative.

L'influence de l'usure de l'échantillon dans le temps est un aspect du panel qui doit être contrôlé en permanence. Plus la période de référence sera longue, plus ces contrôles seront importants et détaillés.

La suite de cette étude permet de franchir une étape supplémentaire dans cette direction. L'examen entrepris au cours du premier chapitre est poursuivi sur une période qui couvre trois vagues du panel, à savoir 1985, 1986 et 1987. Il s'agit de vérifier si l'échantillon cylindré sur ces trois années peut encore être considéré comme une base valable pour étudier l'évolution des revenus durant cette période.

Les références des deux mesures du revenu utilisées restent identiques: il s'agit de deux décilages de l'échantillon initial (des individus), selon le revenu disponible et le revenu disponible par U.C.

## 2.1 Bilan de l'attrition 85, 86, 87.

L'échantillon initial du panel (1985) comprenait 6110 individus. Au cours des deux vagues suivantes, des mouvements d'entrées et de sorties ont modifié la structure de cet échantillon:

- des individus n'ont plus été interrogés au cours d'une vague subséquente.
- de nouveaux individus apparaissent dans l'échantillon.

Après trois années, de tels mouvements conduisent à un répertoire "total" qui n'est plus composé de 6110 personnes mais de 6542 personnes.

Il n'est pas nécessaire de décrire ici ces différents mouvements de manière détaillée. Mais le seul fait de mentionner leur importance permet de rappeler que si le panel perd une partie de son effectif (refus/départs naturels), il s'enrichit aussi de nouveaux effectifs.

L'échantillon initial des 6110 unités reste le cadre de référence de cette étude. De 1985 à 1986, cet échantillon avait déjà subi d'importantes transformations et cette tendance s'est poursuivie en 1987.

Le tableau suivant présente le bilan de cette évolution au moment de la seconde et de la troisième vague. Les % reproduits dans ce tableau ont été calculés par référence à l'ensemble de l'échantillon initial, corrigé du biais de sélection initial par la variable WGT85.

Table 2.1: BILAN DES NON-REPNSES EN 1986 ET 1987: ETAT DE L'ECHANTILLON INTERROGE EN 1985.

Etat de l'échantillon initial			
DIAGNOSTIC	2e vague 86 %	3e vague 87 %	3e vague C.A hors pondération
- individus présents - en 85 et 86 - en 85 et 87	85.0	74.7	(4511)
- refus d'enquête	13.2	22.4	(1427)
- départs naturels	1.8	2.9	(172)
Ensemble (n=6110, éch. pondéré)	100.0	100.0	(6110)

La proportion des personnes qui répondent à l'enquête s'est encore amenuisée lors de la troisième vague du panel. Cette réduction est, certes, plus faible qu'en 1986; mais le bilan cumulé indique néanmoins qu'il ne reste plus -en 1987- que trois personnes sur quatre ayant été interrogées en 1985.

Les refus de répondre atteignent maintenant 22.4% de l'échantillon de base.

Un tel bilan peut être commenté de nombreuses façons.

D'une manière générale il convient de rappeler que cette évolution s'inscrit dans l'ordre des choses; un panel est une entreprise "vivante", soumise -par là même- à des lois que le chercheur ne peut ignorer.

La seule question importante est celle de savoir si l'échantillon s'inscrit toujours à l'intérieur des limites indispensables pour que la poursuite de l'entreprise soit jugée raisonnable.

Avant d'examiner les effets de l'attrition sur les variables de revenu il convient de souligner le fait que certains mouvements observés au sein de l'échantillon ne doivent pas nécessairement être considérés comme irréversibles. Le tableau suivant en atteste, même si de tels cas sont encore peu fréquents durant les trois premières vagues.

Table 2.2: EVOLUTION DE L'ECHANTILLON INITIAL EN 1987, SELON LE BILAN ETABLI EN 1986.

Diagnostic 1987	Refus	Présents	Départs naturels	Total ech. non pond.
1986				
Refus	828	5	-	833
Présents	599	4502	66	5167
Départs naturels	-	4	106	110
total ech. non pond.	1427	4511	172	6110

Ce tableau permet de repérer deux types de mouvements:

- 5 individus refusent l'enquête en 1986, mais l'acceptent en 1987.
- 4 individus qui avaient quitté le pays en 1986, reviennent au Luxembourg en 1987 et acceptent l'enquête.

De tels cas font évidemment figure d'exception parmi les mouvements d'entrée/sortie du panel. Il serait tout à fait déraisonnable de compter sur ces quelques exceptions pour endiguer -même partiellement- les mouvements de sortie qui se manifestent dans l'échantillon. Ces quelques cas symbolisent cependant une opportunité effective: certains départs ne doivent pas être considérés comme définitifs.

Si, d'un côté, il faut bien admettre que la lutte contre l'érosion de l'échantillon ne conduira jamais à une stabilisation de l'effectif initial, il ne faut pas négliger l'utilité de certaines stratégies visant à résister contre cette érosion. L'équipe de recherche n'est pas complètement désarmée face à cette situation. Certaines mesures ont été prises mais elles ne manifestent pas leurs effets au cours des trois premières années du panel.

C'est en 1988 que, pour la première fois, l'équipe chargée de la collecte des données dans le panel luxembourgeois s'est trouvée dans une situation favorable pour la récupération d'enquêtes qui avaient été refusées auparavant (cette opération est désignée par l'expression "track system").

Parmi l'ensemble des enquêtes refusées en 1987, 150 ont été à nouveau mises en circulation l'année suivante. Grâce à cette opération, 50 ménages ont accepté de participer à l'enquête de la quatrième vague (soit environ 150 personnes). Cette opération a encore été répétée en 1990 à partir d'un effectif plus important d'enquêtes enregistrées comme refus en 1986.

Cette stratégie illustre bien le fait que les refus répertoriés lors d'une vague du panel ne doivent pas être interprétés comme des refus définitifs, du moins pour une fraction d'entre eux. Tous les efforts déployés pour minimiser l'importance de ceux-ci concourent à maintenir la qualité du panel au fil des ans.

## 2.2 Examen du taux de survie (85-87) dans les déciles du revenu

Le tableau suivant indique que la variable "diagnostic 85-87" est pratiquement indépendante des deux décilages du revenu (cfr. les coefficients de contingence).

Table 2.3: TAUX DE SURVIE (85-87) SELON LES DECILES DU R.D ET R.D/U.C ETABLIS EN 1985.

Déciles du revenu disponible	Diagnostic 85-87 simplifié		Déciles du revenu disponible par U.C	Diagnostic 85-87 simplifié		Total échantillon
	Présents 85-87	Disparus 85-87		Présents 85-87	Disparus 85-87	
	% col.1	% col.2		% col.3	% col.4	
1er décile	9.0	12.8	1er décile	9.8	10.7	10.0
2e décile	10.2	9.5	2e décile	10.5	8.5	10.0
3e décile	9.8	10.4	3e décile	10.3	9.2	10.0
4e décile	10.5	8.6	4e décile	9.7	10.7	10.0
5e décile	9.8	10.5	5e décile	10.5	8.5	10.0
6e décile	10.4	9.0	6e décile	9.4	12.0	10.0
7e décile	9.9	10.1	7e décile	9.7	10.4	10.0
8e décile	10.6	8.4	8e décile	10.1	9.1	10.0
9e décile	10.1	9.8	9e décile	10.2	9.7	10.0
10e décile	9.7	10.9	10e décile	9.8	10.5	10.0
Ensemble (éch. pondéré WGT85)	100.0	100.0		100.0	100.0	100.0 (n=6110)

coefficient de contingence = 0.07

coefficient de contingence = 0.06

D'une façon générale, les pourcentages enregistrés dans chaque décile sont très proches de ceux que nous avons déjà notés lors du bilan 85-86; la répartition de l'effectif "disparus 85-87" est cependant plus uniforme (cfr col. 2 et 4) que lors du bilan précédent.

Vu l'écart très faible (moins de 0.5% dans la plupart des cas) entre le taux observé et le taux théorique pour chaque décile et pour l'échantillon cylindré 85-87 (cfr col. 1 et 3), il serait déjà possible de conclure que le phénomène d'attrition n'a aucune incidence sur l'étude des revenus au cours de la période 85-87.

Nous allons cependant envisager d'autres éléments afin de mieux saisir les mécanismes susceptibles d'altérer l'échantillon cylindré.

## 2.3 La différenciation des non-réponses en 1987.

Le tableau suivant différencie les non-réponses selon deux rubriques: les refus et les départs naturels. Ceci permet de préciser l'origine de certains écarts observés dans la répartition de l'échantillon cylindré 85-87 sur les déciles du revenu définis en 1985.

Table 2.4: REPARTITION DES CATEGORIES DE NON-REponses OBSERVEES EN 1987, SELON LES DECILES DU REVENU (ECH.PONCERE)

Déciles du revenu disponible R.D	Diagnostic 85-87			Déciles du revenu disponible par U.C R.D/U.C	Diagnostic 85-86			Total échantillon
	Présents 85-87	Refus	Départs naturels		Présents 85-87	Refus	Départs naturels	
	col.1 %	Col.2 %	Col.3 %		Col.4 %	Col.5 %	Col.6 %	
1er décile	9.0	11.5	23.5	1er décile	9.8	9.9	16.7	10.0
2e décile	10.2	8.8	15.1	2e décile	10.5	8.3	10.2*	10.0
3e décile	9.8	10.7	8.2*	3e décile	10.3	8.6	14.1	10.0
4e décile	10.5	8.4	9.9*	4e décile	9.7	11.4	5.7*	10.0
5e décile	9.8	11.2	5.2*	5e décile	10.5	8.4	9.4*	10.0
6e décile	10.4	9.4	6.2*	6e décile	9.4	12.8	5.5*	10.0
7e décile	9.9	10.6	5.6*	7e décile	9.7	10.9	6.9*	10.0
8e décile	10.6	8.7	5.7*	8e décile	10.1	10.2	6.7*	10.0
9e décile	10.1	10.1	8.0*	9e décile	10.2	9.5	11.7	10.0
10e décile	9.7	10.7	12.5	10e décile	9.8	10.2	13.1	10.0
Ensemble (éch. pondéré WGT85)	100.0	100.0	100.0		100.0	100.0	100.0	100.0 (n=6110)

Coefficient de contingence = 0.11

Coefficient de contingence = 0.09

(\*) moins de 20 cas

L'interprétation du tableau '9' est très claire:

- On notera, tout d'abord, que la différenciation des non-réponses conduit à élever le degré d'association entre les mesures du revenu et le diagnostic 85-87; cette augmentation est, infime, mais elle peut être attribuée à la modalité des "départs naturels"; les colonnes 3 et 6 montrent que les décilages du revenu ne sont sensibles qu'à cette seule modalité "départs naturels" de la variable "diagnostic 85-87".

Un examen plus détaillé permet ensuite d'observer certains faits intéressants:

- En ce qui concerne le R.D., seul le premier décile se trouve légèrement sous-représenté (- 1%) dans l'échantillon cylindré 85-87 (cfr col.1); les contrôles effectués à partir des colonnes 2 et 3 indiquent que cette sous-représentation dépend entièrement des départs naturels observés dans le premier décile.

- La situation est encore plus favorable pour la variable R.D/U.C. puisqu'aucun écart supérieur à 0.6% n'est enregistré dans l'échantillon cylindré. Cet écart (0.6 %) observé dans le sixième décile semble davantage attribuable aux refus. L'intervention du système de pondération développé pour 1987 permet de rectifier légèrement cet écart. Les départs naturels ont plutôt tendance à se concentrer dans le premier et le troisième décile du R.D/U.C.

En comparant ce bilan établi en 1987 avec celui présenté pour 1986 (cfr chap.1, sect. 4, tabl.3), nous pouvons mettre en évidence un autre fait important:

- En dépit du fait que l'usure de l'échantillon se soit maintenue à un niveau d'intensité relativement élevé entre la seconde et la troisième vague, ce phénomène ne semble pas avoir affecté la répartition de l'échantillon cylindré 85-87 selon les déciles du revenu (et ce, dès avant que le système de pondération de la troisième vague n'entre en jeu).

- Au contraire, il semblerait plutôt que les risques de biais tendent à s'égaliser entre les différents déciles.

Ce phénomène peut être observé particulièrement au niveau de l'évolution de la répartition des refus selon les déciles: de la seconde à la troisième vague, les refus se répartissent de façon plus uniforme.

Même si l'on observe la même tendance pour les "départs naturels" ces derniers demeurent encore fort concentrés dans certains déciles.

Ces deux mécanismes opposés, à savoir l'égalité de répartition des refus et l'inégalité de répartition des départs naturels produisent un état de l'échantillon cylindré 85-87 qui, du point de vue de l'analyse du revenu, peut être considéré comme très satisfaisant.

Les écarts, minimes, qui sont enregistrés dépendent principalement des variations des "départs naturels". Aucune intervention particulière ne doit être envisagée puisque de tels changements se sont aussi produits dans la population-mère de l'échantillon initial (1985).

Dans la mesure où les refus pourraient altérer l'échantillon cylindré 85-87 pour les analyses du revenu, le système de pondération devrait pouvoir en tenir compte, mais ceci reste à vérifier.

## 2.4 L'effet des pondérations de la troisième vague sur la répartition de l'échantillon cylindré 85-87 selon les déciles du revenu (1985)

Le premier chapitre a montré que les relations entre le système de pondération de la seconde vague d'enquêtes et les variables du revenu étaient relativement faibles. Les résultats étaient donc peu modifiés par l'introduction de poids ajustés pour 1986.

Le tableau suivant présente l'information nécessaire à une vérification similaire. Il porte cette fois sur l'échantillon cylindré 85-87.

Table 2.5: IMPACT DU SYSTEME DE PONDERATION DE LA TROISIEME VAGUE SUR LA REPARTITION DE L'ECHANTILLON 85-87 SELON LES DECILES DU REVENU (ETABLIS SUR L'ECHANTILLON 85, COMPLET)

Répartition de l'échantillon cylindré 85-87					
Déciles selon le revenu disponible	Pondérations = biais de sélection initial (1985)	Pondérations troisième vague (1986)	Déciles selon le revenu disponible par U.C	Pondérations = Biais de sélection initial 1986	Pondérations troisième vague (1986)
	----- %	----- %		----- %	----- %
1er décile	9.0	9.5	1er décile	9.8	9.3
2e décile	10.2	10.1	2e décile	10.5	10.4
3e décile	9.8	9.7	3e décile	10.3	9.9
4e décile	10.5	10.1	4e décile	9.7	9.8
5e décile	9.8	9.8	5e décile	10.5	10.3
6e décile	10.4	10.2	6e décile	9.4	9.6
7e décile	9.9	9.8	7e décile	9.7	9.9
8e décile	10.6	10.6	8e décile	10.1	10.5
9e décile	10.1	10.3	9e décile	10.2	10.3
10e décile	9.7	9.9	10e décile	9.8	10.0
ensemble*	100.0	100.0		100.0	100.0

\* n= 4511

D'une façon générale le système de pondération développé pour la troisième vague tend encore à uniformiser davantage la répartition de l'échantillon cylindré 85-87 selon les déciles du revenu. Le seul exemple de modification notoire concerne le premier décile selon le R.D/U.C (9.8% -> 9.3%).

A l'instar du bilan 85-86, nous pouvons donc conclure au fait que le système de pondération développé pour la troisième vague

ne perturbe pas la répartition de l'échantillon selon les déciles du revenu.

L'usure de l'échantillon au bout de trois années ne semble pas de nature à remettre en cause les possibilités d'analyse diachronique des revenus au niveau d'un échantillon cylindré "85-87".

Deuxième partie

L'effet de l'attrition sur les  
variables "emploi"

## Introduction

La démarche adoptée pour cette analyse sera en tout point identique à celle qui a été suivie au cours de la première partie de ce rapport. Il s'agit, cette fois, d'examiner l'impact de l'usure de l'échantillon sur les analyses relatives à l'emploi.

Les différentes facettes de ce thème, qui sont traitées dans le questionnaire, ne seront pas abordées dans cette étude. La présente démarche visant surtout à vérifier si le phénomène d'attrition risque de compromettre la qualité de certains types d'analyses, il semble préférable que les premières vérifications s'appliquent surtout aux caractéristiques les plus fondamentales du thème envisagé.

Dans cet esprit, les deux variables suivantes ont été retenues:

- *emploi*
- *et position par rapport au marché du travail*

La première variable distingue simplement les personnes qui ont un emploi de celles qui n'en disposent pas; elle permettra de dresser un bilan général des effets de l'attrition après la deuxième et après la troisième vague d'enquêtes.

La seconde variable fournit des indications plus nuancées à propos de la position des personnes par rapport au marché du travail:

1. personnes non concernées (trop jeunes ou aux études)
2. personnes actives (disposant d'un emploi)
3. personnes à la recherche d'un emploi
4. personnes pensionnées (retraite ou invalidité)
5. autres cas (ex. tient le ménage)



## Chapitre 3

### L'effet de l'attrition 85-86 sur les variables "emploi"

#### 3.1 Taux global de survie selon les variables "emploi", dans la seconde vague.

##### 3.1.1

Le tableau suivant permet de constater que l'évolution de l'échantillon entre la première vague et la seconde vague n'a pas d'effet sur la proportion des personnes qui ont un emploi (test du  $X^2$  non significatif et coefficient de contingence = 0.02). La répartition de l'échantillon cylindré 85-86 selon la variable "emploi" est identique (à 0.5% près) à la distribution observée dans l'ensemble de l'échantillon initial.

Les "sans emploi" sont légèrement surreprésentés dans le groupe des personnes qui ne répondent pas en 1986 mais cet écart n'est pas significatif sur le plan statistique.

Table 3.1: EMPLOI SELON LE DIAGNOSTIC SIMPLIFIE 85-86

DIAGNOSTIC SIMPLIFIE 85-86			
EMPLOI	Non réponses en 1986 %	Personnes présentes en 1985 et 1986 %	TOTAL %
OUI	35.4	38.6	38.1
NON	64.6	61.4	61.9
ENSEMBLE ECH. POND. WGT 85	100.0	100.0	100.0 (n=6110)

chi carré = n.s à p = 0.05; coeff.cont. = 0.02

### 3.1.2

L'évolution de l'échantillon entre 1985 et 1986 n'est pas sans effet sur la répartition des personnes selon leur position par rapport au marché du travail. L'association est de très faible intensité (coefficient de contingence = 0.08).

L'échantillon cylindré 85-86 et l'échantillon initial présentent des répartitions quasi similaires.

Les non-réponses enregistrées en 1986, présentent quelques particularités:

- les personnes non-concernées et les personnes actives sont légèrement sous-représentées;
- les autres positions sont sur-représentées.

Table 3.2: POSITION SUR LE MARCHÉ DU TRAVAIL SELON LE DIAGNOSTIC SIMPLIFIÉ 85-86

DIAGNOSTIC SIMPLIFIÉ 85 - 86			
Position marché du travail	Non - réponses en 1986	Personnes présentes en 1985 et 1986	TOTAL
	%	%	%
non concernés	22.4	27.2	26.5
actifs	35.4	38.6	38.1
recherche emploi	2.6	1.2	1.4
pensionnés	15.2	10.6	11.3
autres cas	24.3	22.4	22.7
Ensemble éch.pond.WGT85	100.0	100.0	100.0 n=6110

chi carré = signif. à  $p = 0.01$ ; coeff.cont. = 0.08

## 3.2 Différenciation des "non-réponses en 1986"

### 3.2.1

Dans la section précédente, il est apparu que la variable "emploi" était indépendante de l'évolution globale de l'échantillon entre 1985 et 1986. Qu'en est-il si l'on précise l'origine des "non-réponses"?

- Le tableau suivant montre que les "non-réponses" consécutives à des refus se répartissent aléatoirement sur la variable "emploi".
- Il n'en va pas de même pour les "départs naturels": ceux-ci se trouvent plus souvent localisés au niveau de la modalité "sans emploi".

Table 3.3: EMPLOI SELON LE DIAGNOSTIC 85 -86

EMPLOI	DIAGNOSTIC 85 - 86			TOTAL
	Présents 85 - 86 %	Refus %	Départs naturels %	
OUI	38.6	37.3	22.7	38.1
NON	61.4	62.7	77.3	61.9
ENSEMBLE éch.pond. WGT85	100.0	100.0	100.0	100.0 (n= 6110)

Chi carré sign. à p= 0.01; coeff. cont.= 0.04

### 3.2.2

Le caractère aléatoire de la distribution des refus se confirme à l'examen de la position des individus par rapport au marché du travail.

L'effectif réduit des départs naturels n'autorise pas de conclusions. Cette forme d'usure de l'échantillon touche trois fois plus souvent les "pensionnés".

L'élimination des départs naturels réduit l'intensité de l'association entre l'évolution de l'échantillon et la position des personnes par rapport au marché du travail (coeff.cont. = 0.06). Il s'agit d'une première indication. L'analyse effectuée sur un nombre de vagues d'enquêtes plus élevé fournira des renseignements plus stables et plus fiables dans la mesure où le nombre des départs naturels s'élèvera inévitablement.

Table 3.4: POSITION SUR LE MARCHÉ DU TRAVAIL SELON LE DIAGNOSTIC 85 - 86

DIAGNOSTIC SIMPLIFIÉ 85 - 86				
Position marché du travail	Présents 85 et 86	Refus	Départs naturels	TOTAL
	%	%	%	%
non concernés	27.2	22.3	23.3	26.5
actifs	38.6	37.3	22.7	38.1
recherche emploi	1.2	2.5	3.7*	1.4
pensionnés	10.6	13.0	30.9	11.3
autres cas	22.4	25.0	19.5	22.7
Ensemble éch.pond.WGT85	100.0	100.0	100.0	100.0 n=6110

chi carré signif. à  $p = 0.01$ ; coeff.cont.= 0.11

(\*) effectif inférieur à 20 unités

### 3.3 Vérification de l'effet du système de pondération de la seconde vague.

Les sections précédentes ont montré que l'échantillon cylindré 85-86 s'inscrit dans la continuité de l'échantillon initial. Etant donné que les "refus" enregistrés en 1986 sont distribués de façon aléatoire sur les deux variables relatives à l'emploi, le système de pondération en vigueur pour la seconde vague ne devrait pas modifier les proportions observées dans l'échantillon des personnes ayant participé aux deux vagues d'enquêtes [corrigé du biais de sélection initial (WGT85)].

Table 3.5: POSITION SUR LE MARCHÉ DU TRAVAIL : DISTRIBUTION DES FREQUENCES DANS L'ECHANTILLON CYLINDRE 85-86. COMPARAISON SELON LE MODE DE PONDERATION.

Position marché du travail	Fréquences (%) échantillon cylindré 85 -86		
	Pondérations 1985	Pondérations 1986	Echantillon initial complet
non concernés	27.2	26.5	26.5
actifs	38.6	38.4	38.1
recherche emploi	1.2	1.2	1.4
pensionnés	10.6	11.1	11.3
autres cas	22.4	22.8	22.7
Ensemble éch.pond.WGT85	100.0 n=5167	100.0 n=5167	100.0 n=6110

La comparaison porte sur le passage du système de pondération en vigueur pour la première vague d'enquête (WGT85) au système de pondération appliqué à la seconde vague (WGT86).

La position des personnes par rapport au marché du travail garde le même profil sous les deux systèmes de pondération.

Il faut signaler que la variable "emploi" est prise en compte dans l'analyse qui conduit à la construction du système de pondération "1986"; il est donc logique que le facteur de pondération (WGT86) soit corrélié avec la position des personnes par rapport au marché du travail (coeff. de cont. = 0.52).



## Chapitre 4

### L'effet de l'attrition 85 - 87 sur les variables "Emploi"

#### 4.1 Taux global de survie selon les variables "emploi" dans la troisième vague

##### 4.1.1

Sur quatre personnes interrogées en 1985, trois sont interrogées une seconde fois en 1987.

En trois années, l'échantillon initial s'est réduit de 25%.

Cette usure de l'échantillon a-t-elle des répercussions sur l'étude de l'emploi?

- A première vue, il semble bien que l'on puisse apporter une réponse négative à cette question.
- Dans le tableau ci-après, on observe, certes, quelques écarts entre les pourcentages observés et attendus; mais, ces écarts sont très faibles et non significatifs sur le plan statistique.

Nous pouvons donc conclure ce premier examen de la manière suivante:

- d'un point de vue général, l'évolution de l'échantillon après trois vagues n'a aucun effet sur la variable "emploi".

Table 4.1: EMPLOI SELON LE DIAGNOSTIC SIMPLIFIE 85 - 87

DIAGNOSTIC SIMPLIFIE 85-86			
EMPLOI	Non réponses totalisées en 1987 %	Personnes présentes en 1985 et 1987 %	TOTAL %
OUI	36.5	38.7	38.1
NON	63.5	61.3	61.9
ENSEMBLE ECH.POND. WGT 85	100.0	100.0	100.0 (n=6110)

chi carré = N.S à p = 0.13

#### 4.1.2

Par contre, l'évolution de l'échantillon après la troisième vague d'enquête influence très légèrement le profil de la répartition des personnes par rapport au marché du travail. Bien qu'elle soit très faible, cette association est significative.

- L'échantillon des "survivants" est peu différent de l'échantillon initial.
- Les "départs naturels" sur-représentent les pensionnés et les personnes à la recherche d'un emploi: cette observation avait déjà été signalée après la seconde vague d'enquêtes.

Table 4.2: POSITION SUR LE MARCHÉ DU TRAVAIL SELON LE DIAGNOSTIC SIMPLIFIÉ 85 - 87

DIAGNOSTIC SIMPLIFIÉ 85 - 87			
Position marché du travail	Non - réponses totalisées en 1987	Personnes présentes en 1985 et 1987	TOTAL
	%	%	%
non concernés	22.9	27.7	26.5
actifs	36.5	38.7	38.1
recherche emploi	2.4	1.1	1.4
pensionnés	15.0	10.0	11.3
autres cas	23.1	22.6	22.7
Ensemble éch.pond.WGT85	100.0	100.0	100.0 n=6110

chi carré = signif. à p = 0.001 - coeff. cont.= 0.09

## 4.2 Différenciation des non-réponses en 1987

### 4.2.1

Lorsque les non-réponses sont prises dans leur ensemble, elles n'ont pas d'effet significatif sur le taux des personnes ayant un emploi. Ceci a été montré dans la section précédente. Lorsque les non-réponses sont différenciées, l'évolution de l'échantillon semble bien influencer ce taux de personnes ayant un emploi.

Les départs naturels, à caractère démographique, sont particulièrement le fait de personnes n'ayant pas d'emploi.

Les refus de répondre se distribuent de manière aléatoire entre les personnes ayant un emploi et les personnes sans emploi.

Table 4.3: EMPLOI SELON LE DIAGNOSTIC 85 - 87.

EMPLOI	DIAGNOSTIC 85 - 87			TOTAL
	Présents 85 - 87 %	Refus %	Départs naturels %	%
OUI	38.7	38.1	24.0	38.1
NON	61.3	61.9	76.0	61.9
ENSEMBLE éch.pond. WGT85	100.0	100.0	100.0	100.0 (n= 6110)

Chi carré sign. à  $p = 0.001$  - coeff.cont. = 0.05

### 4.2.2

La différenciation des non-réponses provoque un effet identique sur la répartition des personnes en fonction de leur position par rapport au marché du travail. L'association déjà observée dans la section précédente, se renforce.

L'origine de cette association est toujours la même: les départs naturels sont particulièrement le fait des pensionnés ou inversement, les départs naturels sont sous-représentés parmi les actifs et les personnes non-concernées.

En matière de refus, les quelques écarts relevés en faveur des chômeurs et des pensionnés sont minimes et peuvent encore être considérés comme non-significatifs.

Table 4.4: POSITION SUR LE MARCHE DU TRAVAIL SELON LE DIAGNOSTIC 85 - 87

DIAGNOSTIC 85 - 87				
Position marché du travail	Présents 85 et 87	Refus	Départs naturels	TOTAL
	%	%	%	%
non concernés	27.7	23.6	17.5	26.5
actifs	38.7	38.1	24.0	38.1
recherche emploi	1.1	2.3	3.6	1.4
pensionnés	10.0	12.6	33.3	11.3
autres cas	22.6	23.3	21.6	22.7
Ensemble éch.pond.WGT85	100.0	100.0	100.0	100.0 n=6110

Chi carré = sign. à  $p = 0.001$  - coeff. cont. = 0.14

### 4.3 Vérification de l'effet du système de pondération de la troisième vague.

Jusqu'à présent, les examens convergent vers les mêmes conclusions:

- le cylindrage de l'échantillon 85 - 87 n'a aucun effet sur la variable "EMPLOI" et très peu sur la variable "POSITION M.T";
- les refus se répartissent aléatoirement selon deux variables relatives à l'emploi.

Comment cette situation évolue-t-elle, lorsque les pondérations '1987' contrôlent l'échantillon cylindré et rétablissent l'égalité des probabilités de sélection?

Le tableau ci-dessous montre que toutes les corrections induites par le système de pondération de la troisième vague, tendent à rapprocher la répartition de la variable "POSITION M.T" dans l'échantillon cylindré de celle observée dans l'échantillon initial complet.

Ce redressement de la répartition est toutefois moins "parfait" que celui qui s'était opéré lors de la seconde vague (par l'intervention des poids 86). Mais, il faut bien noter que ces ajustements conduisent à des écarts qui n'excèdent pas finalement 0.6%.

Table 4.5: POSITION SUR LE MARCHÉ DU TRAVAIL: DISTRIBUTION DE FREQUENCES DANS L'ECHANTILLON CYLINDRE 85 - 87. COMPARAISON SELON LE MODE DE PONDERATION.

Position marché du travail	Fréquences (%) échantillon cylindré 85 -87		
	Pondérations 1985	Pondérations 1987	Echantillon initial complet
			%
non concernés	27.7	26.7	26.5
actifs	38.7	38.6	38.1
recherche emploi	1.1	1.1	1.4
pensionnés	10.0	10.7	11.3
autres cas	22.6	22.9	22.7
Ensemble éch.pond.WGT85	100.0 n=4511	100.0 n=4511	100.0 n=6110

## Conclusion

Dans une étude longitudinale de type "panel", l'usure de l'échantillon est inexorable.

Toute l'organisation du travail sur le terrain vise à réduire l'importance de ce phénomène. Des travaux méthodologiques permettent d'analyser les déformations éventuelles de l'échantillon et d'en localiser les origines. Ce type d'examen permet de construire le système de pondération. Complexe et souple à la fois, ce système doit maintenir la "représentativité" de l'échantillon au fil des vagues successives.

Cependant, ces différentes précautions ne sont pas suffisantes pour garantir la validité des analyses de toutes les données collectées dans le panel. Des contrôles supplémentaires sont donc nécessaires pour vérifier que toutes les conditions requises pour l'exécution d'un programme de recherche sont satisfaites.

La démarche proposée dans ce rapport est une illustration de ce type de contrôle.

Il faut encore mentionner un type de risque qui n'a pas été envisagé dans cette étude; l'insuffisance des effectifs peut hypothéquer toute possibilité d'assurer le suivi longitudinal de tel ou tel groupe. L'examen qui a été effectué porte sur une période encore très limitée (trois vagues). Ceci réduit d'autant l'apparition d'un tel risque. Mais le cas des "personnes à la recherche d'un emploi" est significatif. Elles ne représentent pas plus de 1.4% de l'échantillon de départ. Ce taux correspond à ce qui est enregistré au niveau de la population totale du pays. Bien que la taille de ce groupe soit déjà très réduite dans la première vague, les effectifs restent suffisants après trois ans pour réaliser une étude longitudinale. Mais qu'advient-il après quatre ou cinq années?

Cet exemple illustre bien le type de problème que l'usure de l'échantillon ne manquera pas de poser.

Les risques de déformation de l'échantillon ne dépendent pas exclusivement de l'attrition globale de l'échantillon:

des taux de non-réponses, même s'ils sont faibles, peuvent produire des dommages importants, pour peu que ces non-réponses ne soient pas réparties aléatoirement dans l'échantillon et qu'aucun système correctif ne soit mis en oeuvre pour neutraliser ces tendances.

C'est dans cette perspective que fut présenté l'examen des revenus et de l'emploi. Les conclusions sont satisfaisantes puisque l'évolution de l'échantillon sur des périodes de deux et de trois années n'entraînent aucune conséquence en ce qui concerne la qualité des analyses relatives au revenu et à l'emploi. Les vérifications nécessaires ont été faites à différents niveaux de telle sorte que l'on puisse -dans chaque

cas- surveiller l'évolution des refus, des départs naturels et de l'échantillon cylindré ainsi que les effets du système de pondération sur cet échantillon cylindré.

Un échantillon cylindré sur les deux ou trois premières vagues (85 - 86 - 87) offre toute garantie pour les analyses portant sur l'évolution des situations de revenus et d'emploi. Il s'inscrit bien dans la continuité de l'échantillon initial et complet.

## Table des Matières

Chapitre 1	Examen de l'effet de l'attrition sur les deux mesures du revenu	9
1.1	Organisation de la démarche . . . . .	9
1.2	Bilan de l'attrition 85-86 . . . . .	10
1.3	Examen du taux de survie [85-86] dans les déciles du revenu . . . . .	11
1.4	Différenciation des non-réponses en 1986 .	14
1.5	L'effet des pondérations de la seconde vague sur les déciles du revenu . . . . .	17
Chapitre 2	L'attrition 1985/1987 et les mesures du revenu	21
2.1	Bilan de l'attrition 85, 86, 87. . . . .	23
2.2	Examen du taux de survie (85-87) dans les déciles du revenu . . . . .	27
2.3	La différenciation des non-réponses en 1987. . . . .	28
2.4	L'effet des pondérations de la troisième vague sur la répartition de l'échantillon cylindré 85-87 selon les déciles du revenu (1985) .	30
Chapitre 3	L'effet de l'attrition 85-86 sur les variables "emploi"	35
3.1	Taux global de survie selon les variables "emploi", dans la seconde vague. . . . .	35
3.1.1	. . . . .	35
3.1.2	. . . . .	36
3.2	Différenciation des "non-réponses en 1986" . . . . .	37
3.2.1	. . . . .	37
3.2.2	. . . . .	38
3.3	Vérification de l'effet du système de pondération de la seconde vague. . . . .	39
Chapitre 4	L'effet de l'attrition 85 - 87 sur les variables "Emploi"	41
4.1	Taux global de survie selon les variables "emploi" dans la troisième vague . . . . .	41
4.1.1	. . . . .	41
4.1.2	. . . . .	43
4.2	Différenciation des non-réponses en 1987 .	44
4.2.1	. . . . .	44
4.2.2	. . . . .	44
4.3	Vérification de l'effet du système de pondération de la troisième vague. . . . .	45



## Tableaux

Table 1.1:	BILAN DES NON-REPONSES LORS DE LA SECONDE VAGUE : ETAT DE L'ECHANTILLON INTERROGE EN 1985 . . . . .	.10
Table 1.2:	TAUX DE SURVIE SELON LES DECILES DU REVENU DISPONIBLE ET DU REVENU DISPONIBLE PAR UNITE DE CONSOMMATION. . . . .	.11
Table 1.3:	REPARTITION DES CATEGORIES DE NON-REPONSES OBSERVEES EN 1986, SELON LES DECILES DU REVENU (ECH.PONDERE) . . . . .	.14
Table 1.4:	COMPARAISON <sup>5</sup> DES EFFETS DE CHAQUE MODE DE PONDERATION SUR LES DECILES DU REVENU <sup>6</sup> . . . . .	.18
Table 2.1:	BILAN DES NON-REPONSES EN 1986 ET 1987: ETAT DE L'ECHANTILLON INTERROGE EN 1985. . . . .	.24
Table 2.2:	EVOLUTION DE L'ECHANTILLON INITIAL EN 1987, SELON LE BILAN ETABLI EN 1986. . . . .	.25
Table 2.3:	TAUX DE SURVIE (85-87) SELON LES DECILES DU R.D ET R.D/U.C ETABLIS EN 1985. . . . .	.27
Table 2.4:	REPARTITION DES CATEGORIES DE NON-REPONSES OBSERVEES EN 1987, SELON LES DECILES DU REVENU (ECH.PONDERE) . . . . .	.28
Table 2.5:	IMPACT DU SYSTEME DE PONDERATION DE LA TROISIEME VAGUE SUR LA REPARTITION DE L'ECHANTILLON 85-87 SELON LES DECILES DU REVENU (ETABLIS SUR L'ECHANTILLON 85, COMPLET) . . . . .	.30
Table 3.1:	EMPLOI SELON LE DIAGNOSTIC SIMPLIFIE 85-86 . . . . .	.35
Table 3.2:	POSITION SUR LE MARCHE DU TRAVAIL SELON LE DIAGNOSTIC SIMPLIFIE 85-86 . . . . .	.36
Table 3.3:	EMPLOI SELON LE DIAGNOSTIC 85 -86 . . . . .	.37
Table 3.4:	POSITION SUR LE MARCHE DU TRAVAIL SELON LE DIAGNOSTIC 85 - 86 . . . . .	.38
Table 3.5:	POSITION SUR LE MARCHE DU TRAVAIL : DISTRIBUTION DES FREQUENCES DANS L'ECHANTILLON CYLINDRE 85-86. COMPARAISON SELON LE MODE DE PONDERATION. . . . .	.39
Table 4.1:	EMPLOI SELON LE DIAGNOSTIC SIMPLIFIE 85 - 87 . . . . .	.42
Table 4.2:	POSITION SUR LE MARCHE DU TRAVAIL SELON LE DIAGNOSTIC SIMPLIFIE 85 - 87 . . . . .	.43
Table 4.3:	EMPLOI SELON LE DIAGNOSTIC 85 - 87. . . . .	.44
Table 4.4:	POSITION SUR LE MARCHE DU TRAVAIL SELON LE DIAGNOSTIC 85 - 87 . . . . .	.45
Table 4.5:	POSITION SUR LE MARCHE DU TRAVAIL: DISTRIBUTION DE FREQUENCES DANS	

5. cette comparaison est réalisée sur l'échantillon des personnes présentes en 1985 et 1986

6. ces déciles ont été établis sur l'ensemble de l'échantillon observé en 1985

L'ECHANTILLON CYLINDRE 85 - 87. COMPARAISON  
SELON LE MODE DE PONDERATION. . . . .46