

CALCUL DES PONDÉRATIONS DANS LE PANEL EUROPÉEN DE MÉNAGES

Christine Chambaz et Nadine Legendre

Introduction

Le panel européen de ménages a été lancé en 1994 dans les douze pays composant alors l'Union européenne. Son objectif principal est d'étudier la dynamique d'emploi et de revenus des *personnes*. En France, comme pour toutes les enquêtes auprès des ménages réalisées par l'Insee, l'échantillon initial est composé de logements. Ce sont d'abord les ménages qui sont contactés, même si les individus constituent la cible de l'enquête. L'enquête comporte d'ailleurs deux questionnaires, l'un adressé au ménage (sur sa composition, ses conditions de logement...), l'autre aux individus de 17 ans et plus (sur les revenus, l'emploi, la santé, les relations sociales,...).

Contrairement à ce que son nom laisserait supposer, le panel européen est donc *un panel d'individus*. Les règles d'évolution de l'échantillon ont effectivement été fixées en référence aux individus. A la base de ces règles, deux grands principes définissent deux catégories d'individus au regard de l'enquête :

- les individus composant les ménages tiré l'année 1 constituent l'échantillon de base du panel. Ils seront suivis annuellement tant que le panel durera, même s'ils déménagent. Ils forment la population des « *individus panel* ». Par définition, tous les individus des ménages répondant en vague 1 du panel sont des individus panel. Parmi eux, se trouvent des personnes ayant effectivement répondu au questionnaire individuel, mais aussi des individus n'y ayant pas répondu, soit qu'ils ont refusé, soit qu'ils étaient trop jeunes. Par convention, les enfants nés au cours des vagues suivantes de mère individu panel sont également des individus panel.

- les autres adultes des ménages dans lesquels se trouve au moins un individu panel sont interrogés, mais seulement aussi longtemps qu'ils restent avec cet individu panel (adulte ou enfant). Ils constituent la population des « *individus non-panel* ». Par définition, cette population n'existe qu'à partir de la vague 2 du panel, et devrait être de contours mouvants au fil des vagues à venir. Sa prise en compte permet, pour l'analyse transversale, de pallier en partie l'appauvrissement de l'échantillon consécutif aux sorties d'individus panel du

champ de l'enquête. Elle doit en principe assurer la représentativité instantanée de la population des ménages enquêtés¹.

Le panel européen des ménages n'est pas en fait un vrai panel, mais un suivi de cohorte. Il n'y a en effet pas de renouvellement de l'échantillon (voir Deville, 1998).

Compte tenu de l'existence de ces deux catégories de population répondant à l'enquête, deux types de pondérations sont donc calculées :

- une pondération longitudinale, propre aux individus panels ayant répondu en vague 1², et répondant en vague 2. Cette pondération doit permettre l'étude des évolutions individuelles. Elle dérive des poids de base des individus, définis comme les poids initiaux corrigés de l'évolution de l'échantillon³. Pour une période donnée ($t, t+n$), on peut calculer autant de pondérations longitudinales qu'on définit de sous-périodes de suivi. On peut par exemple, sur la période (1994, 1996), décider de suivre les individus ayant répondu à la fois en 1994, en 1995 et en 1996 ; on peut aussi préférer regarder le parcours de ceux qui ont répondu en 1994 et 1996, ou en 1995 et 1996. Au sein d'un même ménage, il peut y avoir partage des poids de base : tous les individus panel répondants reçoivent alors le même poids longitudinal. En vague 2, cela n'a aucun intérêt : les poids longitudinaux correspondent aux poids de base. En revanche, cette opération sur les poids de base pourra ultérieurement permettre d'attribuer un poids longitudinal à des individus non-panel mais ayant répondu à plusieurs enquêtes successives.

- une pondération transversale, à utiliser pour l'analyse en coupe des réponses individuelles. Cette pondération est commune à tous les adultes répondants d'un ménage, qu'ils soient individus panel ou non.

1. A l'immigration près, sauf à considérer que les immigrants récents intègrent tous des ménages préexistant à leur arrivée. Compte tenu de la faiblesse des flux migratoires, cette restriction pourra, sur le court terme, être négligée.

2. Par commodité, et pour assurer la cohérence avec les hypothèses retenues lors du redressement de la vague 1 (non réponse individuelle au sein des ménages répondants considérée comme négligeable, et donc non redressée), tous les individus panel sont réputés avoir répondu en vague 1.

3. non-réponse, mais aussi déformation de la population des ménages. Théoriquement, seule la non-réponse devrait être corrigée pour le calcul de ces poids. Nous verrons cependant que le souci d'assurer une cohérence lors de la diffusion de résultats provenant d'une part d'analyses transversale, d'autre part d'analyses longitudinales, nous a conduits à recalculer ces poids sur la structure de la population observée par ailleurs en 1995 (au moment de la réalisation de la vague 2).

Le calcul de ces pondérations a été effectué en accord avec la méthode recommandée par Eurostat, selon le schéma suivant⁴ :

- (i) identification des individus panel et de leur statut de réponse ;
- (ii) analyse de la non-réponse individuelle ;
- (iii) estimation de probabilités de non-réponse sur des groupes homogènes ;
- (iv) correction de la non-réponse : calcul des poids de base en vague t (V_t) par correction des poids de base en vague (t-1) (V_{t-1}), en fonction de probabilités de non-réponse estimées sur des groupes homogènes ;
- (v) calage de ces poids de base sur la structure par âge et sexe de la population en t, estimée à partir de l'enquête Emploi ;
- (vi) calcul des poids longitudinaux et transversaux, avec et sans étape de calage.

Plusieurs modèles de non-réponse ont été testés. L'objet de cette étude est de présenter l'incidence du choix du modèle sur la distribution des poids et sur l'estimation de quelques indicateurs.

Dans une première partie, nous présentons une analyse des caractéristiques principales des individus panel adultes (17 ans et plus) non-répondants. Nous proposons ensuite, pour la vague 2, une correction de la non-réponse par groupes homogènes (CNRGH) à partir de deux modèles. L'un utilise les seules caractéristiques socio-démographiques des individus, l'autre intègre des variables de comportement, tel le déménagement consécutif à un éclatement du ménage-source.

Nous décrivons ensuite les distributions des poids longitudinaux obtenus selon les deux modèles, et des poids transversaux dérivés, calculés selon la méthode de partage des poids. Nous comparons enfin les valeurs de quelques indicateurs estimés à l'aide de l'un ou l'autre jeu de pondérations, sur la population totale, puis sur les sous-populations les plus mobiles (jeunes, personnes seules...). Les analyses présentées ici reposent essentiellement sur le redressement de la vague 2 du panel. Des éléments sur la correction de la non-réponse en vague 3 figurent en annexe 2

4. Une description plus complète du schéma théorique pourra être trouvée dans EUROSTAT (1995).

1. Evolution de l'échantillon entre les vagues 1, 2 et 3

1.1 L'attrition diminue entre les vagues 2 et 3

En vague 1 (V1), 7 344 ménages avaient été interrogés, soit environ 76 % des ménages appartenant au champ de l'enquête. Ils comprenaient 18 9916 individus, dont 14 524 adultes susceptibles de répondre au questionnaire individuel. Parmi ces derniers, 193 avaient refusé de répondre. La non-réponse individuelle au sein des ménages répondants était donc faible (1,3 %).

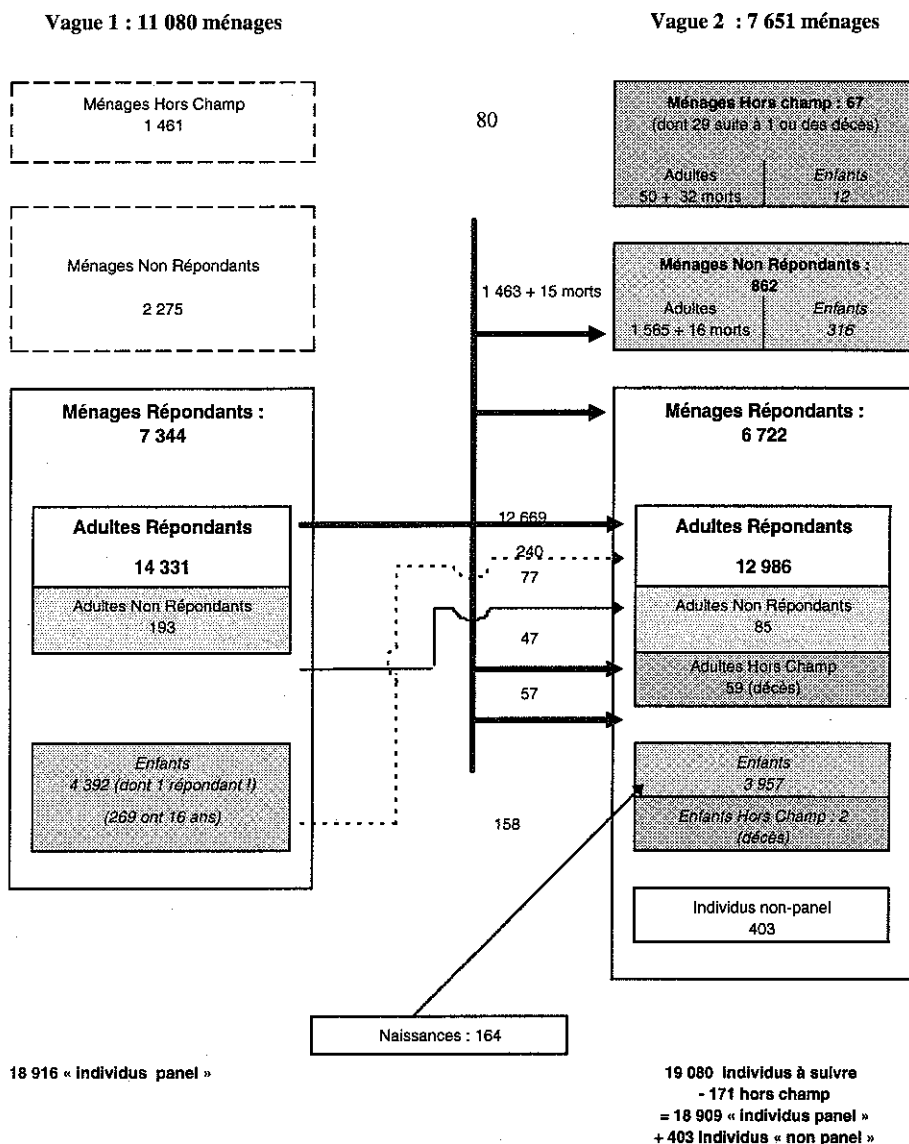
En vague 2 (V2), les individus des ménages non-répondants en V1 sont considérés hors champ. Seuls les 18 916 individus « panel » des 7 344 ménages *ayant répondu en V1*, et eux seuls, sont censés être réinterrogés. 171 sont cependant sortis du champ de l'enquête, par décès, déménagement en ménage collectif ou départ à l'étranger. A contrario, 164 nouveau-nés sont apparus. Au total, ce sont donc 18 909 individus panels à suivre, dont 14 636 adultes de 17 ans ou plus, susceptibles de répondre au questionnaire individuel. Parmi ces adultes, 268 avaient moins de 17 ans en V1.

Tous les individus à interroger ne se trouvent pas dans le ménage qui était le leur en V1. Un certain nombre de ménages ont en effet éclaté, par départ d'un de leurs membres ; il y a ainsi eu création de 307 nouveaux ménages. Le nombre de ménages à contacter pour réaliser les entretiens individuels s'est donc accru de 4,2 %, passant à 7 651 [voir schéma 1].

Les taux de réponse sont relativement bons, meilleurs qu'en V1 : 6 722 ménages ont répondu à l'enquête, soit 88,7 % des 7 584 ménages entrant dans le champ de l'enquête. Au niveau individuel, le taux d'acceptation de l'enquête est le même : 88,8 %, avec 12 986 adultes panel ayant accepté l'interview. Parmi les ménages répondants, 85 individus panel ont refusé de répondre (0,7 %, moins qu'en V1) ; 40 % d'entre eux avaient déjà refusé en V1. Les jeunes de 17 ans, nouvellement entrés dans le groupe adulte des individus panel sont nombreux à accepter l'enquête lorsque leur ménage avait répondu en V1 (89,3 %).

Le taux d'attrition, défini comme l'écart à 1 du rapport entre le nombre d'individus panel répondants en V2 et le nombre d'individus panel répondants en V1 est de 9,4 %.

Schéma 1 : Évolution de l'échantillon entre les vagues 1 et 2 - Schéma simplifié (devenir et origine des adultes répondants)

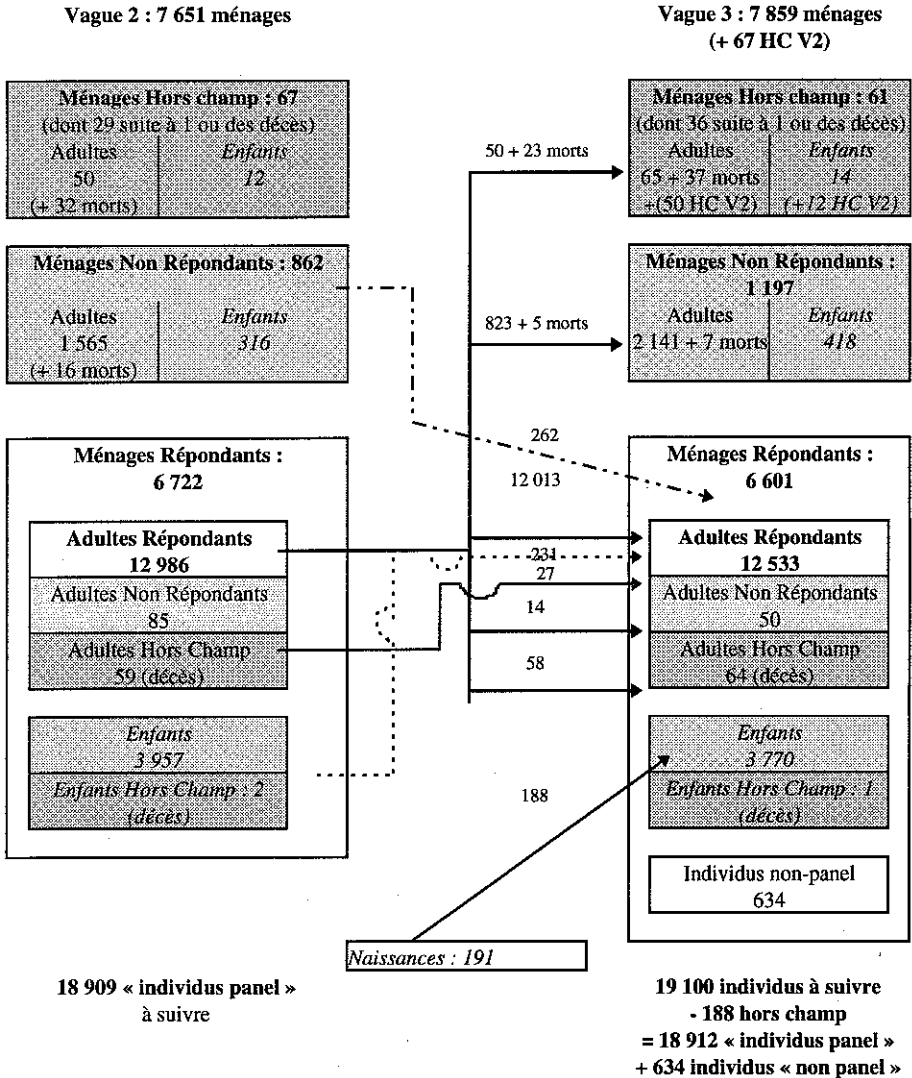


En vague 3 (V3), les 18 909 individus entrant dans le champ de V2 doivent théoriquement être suivis, qu'ils aient ou non répondu en V2. Des mouvements de sortie du champ de l'enquête se sont cependant produits entre V2 et V3 : 109 individus « panel » sont morts, 79 autres ont déménagé en ménage collectif ou à l'étranger. A *contrario*, 191 nouveau-nés de mère panel sont apparus. Au total, ce sont donc 18 912

individus panels à suivre, dont 14 724 adultes de 17 ans ou plus pour une interrogation individuelle. Parmi ces derniers, 260 avaient moins de 17 ans en V2.

Entre V2 et V3, 304 nouveaux ménages sont apparus par éclatement d'un ménage de V2. 67 ménages étant sortis du champ dès la vague 2, le nombre de ménages à contacter pour réaliser les entretiens individuels est donc désormais de $7\,651 - 67 + 275 = 7\,859$, soit 3,1 % de plus qu'en V2 [voir schéma 2].

Schéma 2 : Evolution de l'échantillon entre les vagues 2 et 3 - Schéma simplifié (devenir et origine des adultes répondants)



Les taux de réponse sont encore relativement bons, même s'ils sont plus faibles

qu'en V2⁵ : 6 601 ménages ont répondu à l'enquête, soit 84,6 % des 7 798 ménages entrant dans le champ de l'enquête. Au niveau individuel, 12 533 adultes panel (85,1 %) ont accepté l'interview. Ils représentent 96,5% des individus ayant répondu en vague 2. Parmi les ménages répondants, peu d'individus panel ont refusé de répondre (0,4 %) ; et la plupart avaient déjà refusé en V1.

Le taux d'attrition, défini comme l'écart à 1 du rapport entre le nombre d'individus panel répondants en V3 et le nombre d'individus panel répondants en V2 est de 3,5 %.

1.2. Choix d'un modèle de non-réponse individuelle

Avant de présenter le modèle retenu pour la correction de la non-réponse individuelle, il importe de bien préciser ce qu'on entend par non-réponse :

- la non-réponse ne concerne ici, rappelons-le, que les individus panel. Les individus non-panel ont par définition un poids de base longitudinal nul, le corriger n'aurait donc aucun sens...

- s'agissant des enfants, ils sont réputés avoir répondu dès lors que leur ménage a été contacté. Pour eux la correction de la non-réponse revient à corriger leur poids de base du taux de non-contact des ménages comprenant des enfants. Ce taux est de 1,29 % en V2, et de 1,75 % en V3 parmi les répondants de V2. Par convention, les nouveau-nés se voient affecter la moitié du poids de leur mère.

- au niveau des adultes (individus de 17 ans et plus), c'est l'existence d'un questionnaire individuel qui définit la réponse. Dans la suite, nous ne parlerons plus que de cette population.

Nous aurions pu estimer d'abord une probabilité de contact de l'individu à travers son ménage, puis une probabilité de non-réponse conditionnelle au contact. Nous avons préféré estimer directement la probabilité de non-réponse en introduisant dans le modèle explicatif des variables elles-mêmes fortement corrélées à la probabilité de contact : nationalité, type de ménage, catégorie socioprofessionnelle, voire déménagement du ménage, etc.

Deux modèles explicatifs de la non-réponse des individus panels adultes ont été estimés, en V3 comme en V2. Pour chacun d'entre eux, nous avons d'abord retenu comme variables exogènes des variables collectées lors de la vague d'enquête précédente, et donc disponibles à la fois pour les individus répondants et les individus non-répondants : sexe, âge en tranches, niveau d'études, type de ménage,

5. Les règles de suivi des individus, qui imposent de suivre les personnes ayant refusé de répondre en V2, engendrent mécaniquement ce taux plus élevé.

charges liées au logement, nationalité, activité et catégorie socioprofessionnelle de la personne de référence du ménage, type de commune, pauvreté monétaire du ménage. La différence entre les deux modèles porte sur l'introduction ou non d'une variable supplémentaire décrivant la mobilité géographique *depuis* la vague précédente.

Dans un premier temps, nous avons donc banni toutes les variables de comportement, telles que le déménagement ou l'éclatement du ménage. Nous considérons en effet qu'il était abusif de supposer une homogénéité de comportement entre les individus qui déménagent.

Dans un second temps, nous sommes revenus sur cette considération, et avons réintroduit les variables de mobilité géographique. Leur omission revenait en effet à supposer que les individus mobiles ne se distinguent pas des autres en terme de structure ni de comportement, une fois prises en compte quelques variables socio-démographiques de base. Cela constitue une hypothèse tout aussi forte, en particulier dans le cadre du panel dont un des principaux objectifs est justement d'étudier les corrélations entre différentes formes de mobilité (professionnelle, démographique, en terme de revenus,...).

Pour les vagues 2 et 3, les variables les plus discriminantes du point de vue de la non-réponse sont sensiblement les mêmes, soit :

- lorsqu'on estime les modèles sans variables de comportement : le type de ménage, la nationalité de la personne de référence du ménage, sa catégorie socioprofessionnelle et le type de commune, suivi par l'âge de l'individu en vague 2 ; la catégorie socioprofessionnelle de la personne de référence du ménage, sa nationalité, l'âge de l'individu, le type de ménage puis l'activité de la personne de référence et le poids des charges liées au logement... en vague 3 [tableau 1].

- lorsqu'on estime les modèles avec variables de comportement : le déménagement suite à un éclatement de ménage, le type de ménage, la nationalité de la personne de référence du ménage, sa catégorie socioprofessionnelle puis le type de commune... en vague 2 ; le déménagement suite à un éclatement de ménage, la catégorie socioprofessionnelle de la personne de référence du ménage, sa nationalité, le type de ménage, puis l'activité de la personne de référence et le poids des charges liées au logement... en vague 3 [tableau 2].

La variable de déménagement suite à un éclatement est de loin la plus discriminante. Les enquêteurs avouent d'ailleurs rencontrer beaucoup de difficultés pour obtenir l'adresse d'un enfant qui quitte le ménage, et dont les parents jugent que ce n'est pas la peine « d'aller l'embêter ». De même, les individus ne se montrent pas forcément coopératifs quand on leur demande les coordonnées d'un ex-conjoint. Elle semble toutefois moins discriminante en vague 3 qu'en vague 2.

Tableau 1 : Un premier modèle de non-réponse des adultes individus panel, sans variables de comportement

Variable	VAGUE 2		VAGUE 3			
	Paramètre	Ecart-type	Paramètre	Ecart-type		
Constante	-2 0990	0 0645	-2.4121	0.1121		
Homme	Référence					
Femme	-0,1207	0,0531	ns	ns		
17 à 24 ans	0,3138	0,0731	Référence			
25-34 ans	} Référence		0,2119	0,0968		
35-54 ans			Référence	0,1226		
55-64 ans			-0,3396			
65 ans et plus			Référence			
Niveau d'études atteint :						
CAP-BEP	} Référence	0,0761	-0,1982	0,1083		
autre, < à l'enseignement supérieur			Référence	0,1185		
Enseignement supérieur			-0,2719			
	-0,2356					
Personne seule, couple sans enfant, famille monoparentale	} Référence		Référence	0,1222		
Couple avec 2 enfants			0,1050	-0,3785	0,1154	
Couple avec 1 enfant			-0,2750	0,0738	-0,2533	0,1279
Couple avec ≥ 3 enfants			-0,7214		-0,2910	
Autre ménage			0,1892		Référence	
Charges de logt supportables	Référence		Référence			
Charges de logt trop élevées	0,2144	0,0654	0,4233	0,0839		
PR française	Référence		} Référence	0,1617		
PR européenne	0,4090	0,1296				
PR autre nationalité	0,5569	0,1285			0,6964	
Agglomération parisienne	0,3163	0,0730	0,2152	0,0989		
Autre lieu	Référence		Référence			
Ménage non pauvre	Référence		non disponible			
Ménage pauvre	0,2085	0,0785				
PR agriculteur exploitant	Référence		-0,5356	0,2313		
PR artisan, commerçant	0,5076	0,0885	0,4751	0,1474		
PR cadre	} Référence	0,0607	0,3266	0,1346		
PR prof. intermédiaire			Référence	0,1170		
PR employé			0,5155	0,1083		
PR ouvrier			0,2662	0,2390		
PR non active occupée	Référence		Référence			
PR active occupée	-0,2700	0,0574	-0,4156	0,0836		
PR a un CDD	0,2949	0,1454	ns	ns		
Propriétaire du logement	ns	ns	-0,3745	0,0783		
Autre statut			Référence			
Nombre d'observations	14 634		13 095			
-2*LogL	10 016 451		6 062 729			

PR = personne de référence du ménage

Tableau 2 : Un deuxième modèle de non-réponse des adultes individus panel intégrant des variables de comportement (déménagement, éclatement du ménage)

Variable	VAGUE 2		VAGUE 3	
	Paramètre	Ecart-type	Paramètre	Ecart-type
Constante	-2,1340	0,0561	-2,3992	0,1111
Niveau d'études atteint inférieur à l'enseignement supérieur	Référence	0,0779	Référence	0,1158
Enseignement supérieur	-0,3142		-0,2382	
Personne seule	Référence		Référence	
Couple sans enfant	Référence	0,0863	Référence	0,1250
Couple avec 1 enfant	-0,3224	0,1054	-0,4375	0,1191
Couple avec 2 enfants	Référence	0,1374	-0,3333	0,1316
Couple avec ≥ 3 enfants	-0,7498		-0,3886	0,1151
Famille monoparentale	-0,2914		Référence	
Autre ménage	Référence		-0,2976	
Charges de logt supportables	Référence		Référence	
Charges de logt trop élevées	0,2113	0,0670	0,4230	0,0845
PR française	Référence) Référence	0,1636
PR européenne	0,4198	0,1328		
PR autre nationalité	0,5291	0,1333		
Agglomération parisienne	0,3215	0,0746	0,2356	0,0994
Autre lieu	Référence		Référence	
Ménage non pauvre	Référence		non disponible	
Ménage pauvre	0,2130	0,0804		
PR agriculteur exploitant	Référence		-0,4987	0,2317
PR artisan, commerçant	0,5021	0,0901	0,4797	0,1480
PR cadre) Référence	0,0619	0,3476	0,1352
PR prof. intermédiaire		Référence		0,1175
PR employé			0,5232	0,1085
PR ouvrier	0,2423		0,2638	
PR non active occupée	Référence		Référence	
PR active occupée	-0,2845	0,0583	-0,4326	0,0825
PR a un CDD	0,2764	0,1498	ns	ns
Propriétaire du logement	ns	ns	-0,3952	0,0774
Autre statut			Référence	
55-64 ans	ns	ns	-0,3263	0,1221
Autre tranche d'âge			Référence	
Déménagement suite à un éclatement de ménage	2,1048	0,1049	1,5826	0,1542
Autre cas	Référence		Référence	
Nombre d'observations	14 634		13 095	
-2*LogL	9 692 794			

PR = personne de référence du ménage

2. Poids de base des individus panel pour la vague 2

2.1 Correction de la non-réponse par groupes homogènes et calcul des poids de base pour la vague 2

La correction de la non-réponse a été effectuée en modifiant les poids de base V1 selon la formule :

$$\text{Poids de base V2} = \frac{\text{Poids de base V1}}{1 - \text{taux de non réponse}}$$

Afin de limiter la dispersion des poids résultant de la CNRGH, les taux de non-réponse ont été calculés pour des *catégories* homogènes définies par le croisement des modalités des variables apparues les plus discriminantes à l'étape précédente.

Les catégories ont ainsi été construites en croisant :

- dans le premier cas (pas de variables de comportement), le type de ménage, la nationalité de la personne de référence, sa catégorie socioprofessionnelle, le type de commune et l'âge de l'individu ;
- dans le second cas (intégration de variables de comportement) le déménagement suite à un éclatement de ménage, le type de ménage, la nationalité de la personne de référence du ménage, sa catégorie socioprofessionnelle et le type de commune.

Lorsqu'une catégorie comptait peu d'individus, on l'a regroupée avec la catégorie la plus proche pour le calcul du taux de non-réponse. La proximité entre catégories a été appréciée à partir des coefficients estimés dans le modèle. Au total, 31 catégories sont distinguées dans le premier cas, et 30 dans le second.

Les taux de non-réponse pour chacune de ces catégories sont présentés dans l'annexe 1, tableaux A et B. Ils varient, selon les catégories, entre 4,6 % et 38,8 % dans le premier cas (pas de variables de comportement), entre 2,9 % et 58,9 % dans le second (intégration de variables de comportement). La dispersion des poids devrait donc être plus importante dans le second cas.

2.2. Distribution des poids de base longitudinaux V2 des individus panel adultes

La distribution des poids de base des individus panel adultes en vague 2 est légèrement plus dispersée lorsqu'on intègre la variable de déménagement et d'éclatement de ménage à la grille de construction des catégories homogènes.

Un individu obtient un poids particulièrement élevé dans ce cas. Lorsqu'on supprime cet individu, la dispersion reste cependant légèrement plus forte que lorsque la CNRGH est réalisée sans variable de comportement. [tableau 3].

Nous reviendrons dans la partie 5 sur les caractéristiques de la population pour laquelle le rapport entre les deux séries de poids est très différent de 1.

Tableau 3 : Poids de base (corrégés de la non-réponse) des 12 986 individus panel répondants en Vague 2

	Présence de variables de déménagement et éclatement du ménage pour la CNRGH		Rapport entre ces deux séries de poids
	NON (1)	OUI (2)	(3)=(2)/(1)
Somme	45 178 331	45 178 331	
Moyenne	3 479,00	3 479,00	1,00070
Ecart-type	1 120,06	1 164,94	0,09631
Minimum	1 977,40	1 944,17	0,79964
1%	2 189,42	2 162,95	0,92087
5%	2 385,53	2 350,90	0,94971
10% D1	2 534,79	2 504,57	0,95951
25% Q1	2 817,04	2 803,93	0,97856
50% Med	3 264,46	3 248,95	0,98877
75% Q3	3 863,33	3 869,07	1,00898
90% D9	4 507,77	4 508,92	1,01317
95%	4 979,33	4 996,23	1,02216
99%	8 784,96	9 016,83	1,53573
Maximum	15 494,51	22 477,47 (*)	2,15742
Range	13 517,12	20 533,30	
Q3-Q1	1 046,29	1 065,14	
D9/D1	1,78	1,80	

(*) un seul individu a un poids si élevé. L'observation dont le poids est juste inférieur à un poids de 15395. (plus faible que le poids (1) maximum). Le rapport (2)/(1), pour cet individu au poids maximum, est de 1,498. La forte valeur de ce poids (2) maximum n'est donc que partiellement imputable à la grille de construction des catégories retenue. L'individu concerné (un homme de 30 ans, seul en vague 2 suite à un éclatement de son couple, vivant en agglomération parisienne) avait déjà un poids élevé en vague 1 : 13 297,86, la moyenne des poids étant alors de 3 109,97. Ce poids déjà fort en vague 1 tenait à deux facteurs : un poids de sondage déjà fort (logement ayant statut de résidence secondaire lors du Recensement de Population de 1990) combiné à une forte correction pour la non-réponse.

La comparaison des poids des vagues 1 et 2 montre également une amplitude des rapports de poids accrue lorsque la CNRGH intègre la variable de déménagement. Pour les individus panels adultes, ce rapport est en effet alors en moyenne de 1,1267, et varie entre 1,0297 et 2,4317⁶, contre 1,1265 en moyenne lorsque la CNRGH n'intègre pas la variable de déménagement, le rapport étant alors compris entre 1,0486 et 1,6339.

3. Calage sur marges des poids de base de la vague 2

3.1. Population et variables de calage

Le calage sur marges a pour objet d'assurer une plus grande représentativité transversale de notre échantillon. Il a cependant été réalisé avant calcul des poids transversaux, conformément aux recommandations d'Eurostat.

Nous avons choisi de ne recalculer que la population servant effectivement au calcul de ces poids transversaux : les adultes panel répondants. Nous avons opté pour un calage minimal, basé sur le respect des structures individuelles par âge décennal x sexe et du nombre d'individus par ménage, telles qu'estimées à partir de l'enquête Emploi de mars 1995.

Notre estimation donne une population plus importante que celle de l'enquête emploi. Cela peut s'expliquer par la convention adoptée pour la définition du champ de l'enquête : en l'absence d'éléments permettant de dire si un individu appartient au champ de l'enquête (parti sans laisser d'adresse), on considère qu'il en fait partie. Cela revient à surestimer le champ de l'enquête et le taux de non-réponse.

Avant calage, et quelle que soit la série de poids de base utilisée, l'échantillon du panel présente un déficit d'individus vivant dans des ménages d'1 ou 2 personnes, en particulier de femmes âgées de 72 ans ou plus.

6. la valeur maximale n'étant pas atteinte par l'individu au poids maximal.

3.2. Distribution des poids de base après calage

Tableau 4a : Distribution des poids des adultes panel répondants en vague 2 - Pas de variables de comportement (déménagement, éclatement) pour la CNRGH

	Poids avant CNRGH vague 2	Poids de base vague 2		Rapport des poids		Rapport des poids vague 2 avant et après calage (6)=(3)/(2)
		non recalé (2)	recalé (3)	non recalé (4)=(2)/(1)	recalé (5)=(3)/(1)	
Somme	40 048 604	45 178 331	44 842 091	-	-	-
Moyenne	3083,98	3479,00	3453,11	1,1265	1,1169	0,9912
Ecart-type	960,81	1120,06	1150,75	0,0601	0,0863	0,0501
Minimum	1867,58	1977,40	1868,69	1,0486	0,9258	0,8691
1%	1981,39	2189,42	2126,77	1,0486	0,9696	0,8957
5%	2152,39	2385,53	2318,54	1,0667	0,9941	0,9224
10%	2282,99	2534,79	2472,52	1,0741	1,0117	0,9273
25% Q1	2521,40	2817,04	2782,00	1,1007	1,0528	0,9506
50% Med	2904,91	3264,46	3215,17	1,1139	1,1148	0,9816
75% Q3	3421,74	3863,33	3836,20	1,1526	1,1725	1,0360
90%	3942,78	4507,77	4542,58	1,1613	1,2128	1,0667
95%	4327,9	4979,33	5049,71	1,2335	1,2497	1,0747
99%	7767,84	8784,96	8470,29	1,3211	1,3512	1,0916
Maximum	13885,4	15494,51	16376,98	1,6339	1,8431	1,0916
Range	12017,82	13517,12	14508,29	-	-	-
Q3-Q1	900,34	1046,29	1054,20	-	-	-
Mode	3022,08	4104,47	6478,70	1,1007	1,1403	1,0173

Le calage sur adultes répondants a été réalisé à l'aide du logiciel CALMAR. La méthode retenue est le Logit tronqué, les bornes retenues étant de 0,8 et 1,2, pour les deux calages réalisés.

Sans variables de comportement pour la CNRGH, les poids recalés se situent dans un rapport maximal de 1 à 8,8. Le rapport entre le premier et le 99ème centiles est de 4,0 ; entre le 5ème et le 95ème centiles, il tombe à 2,2. Le rapport entre les poids de base V2 avant et après calage en moyenne de 0,99 ; il varie entre 0,87 et 1,09 [tableau 4a].

Lorsqu'on intègre des variables de comportement pour la CNRGH (déménagement, éclatement de ménage), les poids recalés varient dans un rapport de 1 à 12,3. Le rapport entre le premier et le 99ème centiles est de 4,2 ; entre le 5ème et le 95ème centiles, il tombe à 2,2. Le rapport entre les poids de base V2 avant et après calage en moyenne de 0,99 ; il varie entre 0,90 et 1,10. [tableau 4b]. La distribution est donc ici légèrement plus dispersée (le rapport de 1 à 12,3 provenant, rappelons-le, du poids extrême d'un seul individu).

Tableau 4b : Distribution des poids des adultes panel - Intégration de variables de comportement pour la CNRGH

	Poids avant CNRGH vague 2	Poids de base vague 2		Rapport des poids		Rapport des poids vague 2 avant et après calage
	(1)	non recalé	recalé	non recalé	recalé	(6)=(3)/(2)
		(2)	(3)	(4)=(2)/(1)	(5)=(3)/(1)	
Somme	40 048	45 178 331	44 842 091	-	-	-
Moyenne	3083,98	3479,00	3453,11	1,1267	1,1176	0,9914
Ecart-type	960,81	1164,94	1185,72	0,1197	0,1335	0,0388
Minimum	1867,58	1944,17	1896,77	1,0297	0,9360	0,8979
1%	1981,39	2162,95	2116,00	1,0310	0,9703	0,9133
5%	2152,39	2350,90	2312,19	1,0511	1,0002	0,9377
10%	2282,99	2504,57	2454,37	1,0624	1,0138	0,9456
25% Q1	2521,40	2803,93	2771,16	1,0816	1,0559	0,9591
50% Med	2904,91	3248,95	3205,86	1,1105	1,1047	0,9889
75% Q3	3421,74	3869,07	3829,46	1,1334	1,1490	1,0142
90%	3942,78	4508,92	4517,62	1,1715	1,1915	1,0456
95%	4327,9	4996,23	5075,42	1,2084	1,2373	1,0724
99%	7767,84	9016,83	8877,38	1,6939	1,7171	1,0729
Maximum	13885,4	22477,47	23260,76	2,4317	2,5957	1,1004
Range	12017,82	20533,30	21363,99	-	-	-
Q3-Q1	900,34	1065,14	1058,30	-	-	-
Mode	3022,08	4195,42	6552,73	1,1105	1,0651	1,0361

4. Calcul des poids transversaux

4.1. La méthode employée : le partage des poids

Conformément aux recommandations d'Eurostat, c'est la méthode de partage des poids qui a été retenue pour le calcul des poids transversaux⁷.

Ce principe se traduit par le partage entre tous les membres adultes du ménage de la somme des poids de base des adultes (personnes de 17 ans et plus), que ces derniers soient individus panel ou individus non-panel. Rappelons que les poids de base des individus non-panel sont nuls. Soit :

$$\text{poids transversal du ménage } i = w_i = \frac{1}{p+n} \sum_{k=1}^p u_{i,k}$$

d'adultes panel du ménage i et n le nombre d'adultes non-panel de ce où $u_{i,k}$ représente le poids de base en V2 de l'individu k dans le ménage i , p est le nombre ménage

Tous les membres adultes répondants d'un ménage reçoivent donc le même poids transversal, défini comme la moyenne de leurs poids de base.

4.2. Distributions au niveau ménage des poids transversaux V2⁸ calculés sans ou avec calage préalable

Nous avons calculé des poids transversaux avant et après calage des poids de base V2 des adultes panel, pour chacune des deux séries de poids de base.

Parmi les 6.722 ménages répondants, 2 n'auront pas de poids transversal : l'un ne compte parmi ses trois membres qu'un seul individu panel, âgé de moins de 17 ans ; l'autre ne compte qu'un seul individu, qui n'a pas répondu au questionnaire individuel. Les distributions de poids sont donc données pour les 6 720 ménages pour lesquels leur calcul est possible [tableau 5].

Dans le cas d'un redressement sans intégration de variables de comportement lors de la CNRGH, on obtient les résultats suivants :

- avant calage, les poids transversaux se situent dans un rapport maximal de 1 à 24,8. Le rapport entre le premier et le 99ème centiles est de 6,7 ; entre le 5ème et le 95ème centiles, il tombe à 2,2.

7. pour une justification de la méthode, voir par exemple Lavallée (1995)

8. Le calcul des poids transversaux de la vague 3 repose sur le partage d'une série particulière de poids de base. Afin de ne pas perdre pour l'analyse les ménages dont aucun individu n'avait répondu en vague 2, ni n'avait donc de poids de base V3 (au sens employé jusqu'à présent), ces derniers ont été calculés par correction des poids de base V1. Nous ne les avons pour l'instant pas recalés, et nous n'en présenterons pas les distributions ici.

- après calage, le rapport maximal des poids transversaux est de 26,6, les rapports entre le premier et le 99ème centiles, et entre les 5ème et 95ème centiles restent sensiblement les mêmes (6,7 et 2,4).

Le rapport entre les poids transversaux avant et après calage est en moyenne de 1,00 ; il varie entre 0,88 et 1,09.

Dans l'autre cas, où la CNRGH intègre des variables de comportement, les poids transversaux sont davantage dispersés avant calage, mais le calage les modifie moins. Aussi, la dispersion après calage (rapport interdécile, notamment) est moins forte que dans le cas précédent :

- avant calage, les poids transversaux se situent dans un rapport maximal de 1 à 36,8. Le rapport entre le premier et le 99ème centiles est de 6,1 ; entre le 5ème et le 95ème centiles, il tombe à 2,3.

- après calage, le rapport maximal des poids transversaux est plus faible (35,7), les rapport entre le premier et le 99ème centiles, et entre les 5ème et 95ème centiles restent les mêmes (6,1 et 2,3).

Le rapport entre les poids transversaux avant et après calage est en moyenne de 0,998 ; il varie entre 0,91 et 1,07.

Tableau 5 : distribution au niveau ménage des poids transversaux de la vague 2

	Pas de variables de comportement (déménagement, éclatement) pour la CNRGH			Intégration de variables de comportement (déménagement, éclatement) pour la CNRGH		
	non calé	calé	calé/non calé	non calé	calé	calé/non calé
Somme	23 009 997	23 151 040	-	23 168 631	23 165 932	-
Moyenne	3 424,11	3 445,10	1,0040	3 447,71	3 447,31	0,9981
Ecart-type	1 202,007	1 253,73	0,0489	1 242,50	1 273,98	0,0354
Minimum	623,71	614,89	0,8847	610,14	651,27	0,9112
1%	1 399,69	1 411,89	0,9180	1 507,16	1 517,80	0,9334
5%	2 241,24	2 178,26	0,9312	2 246,93	2 221,11	0,9449
10 % D1	2 441,11	2 395,54	0,9410	2 433,81	2 402,64	0,9523
25 % Q1	2 780,53	2 766,02	0,9640	2 777,34	2 753,24	0,9623
50 % Med	3 214,74	3 203,71	0,9951	3 218,69	3 193,80	0,9985
75 % Q3	3 827,15	3 861,97	1,0433	3 840,87	3 841,87	1,0238
90 % D9	4 489,01	4 588,56	1,0747	4 517,65	4 564,27	1,0443
95 %	5 012,58	5 155,74	1,0843	5 058,19	5 155,51	1,0729
99 %	9 412,26	9 390,29	1,0916	9 388,81	9 331,05	1,0729
Maximum	15 494,51	16 376,98	1,0916	22 477,47	23 260,76	1,0729
Range	14 870,80	15 762,09	-	21 867,33	22 609,950	-
Q3-Q1	1 046,61	1 095,95	-	1 063,53	1 088,62	-
D9/D1	1,84	1,92	-	1,86	1,90	-

5. Incidence de la série de poids utilisée sur la valeur de quelques indicateurs

Les analyses présentées dans cette partie reposent essentiellement sur l'utilisation des différents jeux de pondérations calculés pour la vague 2 du panel. Les séries de poids calculées pour la vague 3 ne seront que peu mobilisées ici.

Cette partie comporte d'abord une rapide description des populations dont les poids sont le plus affectés par l'introduction ou non de la variable de déménagement lors de la correction de la non-réponse. Sont ensuite comparées diverses estimations : de la taille de la population, de sa structure et, données plus au coeur de l'enquête, des niveaux de vie et de la mobilité en terme d'activité. Ces dernières estimations sont déclinées sur la population entière et sur quelques segments de population particulièrement mobiles (jeunes, personnes seules...)

5.1 Population dont le poids de base est fortement affecté par l'introduction ou non des variables de déménagement et d'éclatement du ménage

Les poids de base de l'individu sont ici considérés comme fortement affectés par l'introduction des variables de comportement (déménagement, éclatement du ménage), s'ils diffèrent de plus de 5 %. 1,8 % des individus panels adultes ont un poids de base supérieur d'au moins 5 % lorsqu'on intègre les variables de déménagement. *A contrario*, 5,7 % ont un poids au moins 5 % plus faible.

Les individus dont le poids augmente fortement sont, en octobre 1995, dans près de 8 cas sur 10 des locataires (1/3 dans la population totale des individus panel adultes). 63 % sont des actifs occupés, 12,2 % des chômeurs, et 14,8 % des étudiants. 42 % d'entre eux ont changé de situation professionnelle entre octobre 1994 et octobre 1995. Parmi les actifs occupés, 1/3 ont un emploi à durée déterminée. Ils sont en effet plus jeunes que le reste de la population : 54 % ont moins de 25 ans, et 32 % entre 25 et 35 ans. Ce sont essentiellement des personnes seules (41,3 %) et des couples sans enfant (37,1 %). 1/3 a un niveau de diplôme supérieur au baccalauréat. Ils sont sur-représentés en milieu urbain, hors agglomération parisienne [tableau 6].

La population dont le poids est plus faible lorsqu'on intègre les variables de comportement est à la fois semblable et différente : moins souvent active occupée (43,4 %), mais plus souvent au chômage (8,5 %) que notre échantillon dans son ensemble. 17 % ont changé de situation professionnelle entre les deux vagues d'enquête, moins que le groupe précédent, mais plus que notre échantillon dans son ensemble. 25 % des actifs occupés occupent un CDD. Cette population est plus âgée

que la précédente, mais plus jeune que notre échantillon dans son ensemble : 42,8 % ont moins de 25 ans, et 15,2% entre 25 et 35 ans. Ce sont essentiellement des familles monoparentales (19,2%) et des ménages complexes (33,8%). 25% sont en formation ou en cours d'études initiales. Près d'un sur trois se trouve en agglomération parisienne.

Tableau 6 : Population dont le poids varie fortement selon le modèle de non-réponse

	Individus panel adultes dont le rapport poids avec variable de comportement / poids sans variable de comportement est ...		Ensemble de l'échantillon des individus panel adultes
	< 0,95	> 1,05	
Effectif	746	236	12 986
Part dans l'échantillon	5,7	1,8	100,0 %
Locataire	43,2	78,8	32,6
Propriétaire ou accéd.	51,5	12,7	61,9
Autre cas	5,3	8,5	5,5
Actif occupé	43,4	62,9	51,0
Chômeur	8,5	12,2	6,8
Etudiant	25,2	14,8	8,6
Retraité	12,6	3,0	21,6
Autre inactif	10,3	7,1	11,9
Changement de sil. professionnelle entre les deux enquêtes	17,8	42,2	11,9
Parmi les actifs, CDD	25,4	32,0	13,4
Moins de 25 ans	42,8	54,0	15,1
25-34 ans	15,2	32,1	18,2
35-44 ans	10,6	5,1	19,4
45-54 ans	10,0	5,0	16,7
55 ans et plus	21,4	3,8	30,6
Personne seule	7,9	41,3	13,0
Couple sans enfant	13,1	37,1	25,3
Couple avec enfant(s)	26,0	8,9	44,3
Famille monoparentale	19,2	7,2	4,5
Autres ménages	33,8	5,5	12,9
Commune rurale	21,6	13,5	28,7
U.U. < 20.000 hab.	12,6	20,3	12,2
UU 20.000-100.000	10,0	19,4	13,2
UU > 100.000 hab	24,3	33,3	27,2
agglomération Paris.	31,5	13,5	13,7

5.2 Taille de la population

Le panel estime une population d'individus trop nombreuse en vague 2, mais pas assez en vague 3, et une population de ménages sensiblement identique à celle de l'enquête Emploi de mars 1995 [tableau 7]. La prise en compte de l'éclatement des ménages lors de la CNRGH conduit à une estimation du nombre de ménages plus importante. Elle revient en effet à accorder un poids plus important aux individus ayant fondé un nouveau ménage, et gonfle donc mécaniquement le nombre de ménages. Les fusions de ménages panel étant extrêmement rares, rien ne vient compenser ce mouvement de création de nouveaux ménages.

Tableau 7 : Estimation de la taille de la population

	Nombre d'individus	Nombre de ménages
Vague 1	57 913 481	22 839 615
Vague 2 (poids transversal)		
1) pas de variables de comportement pour la CNRGH		
sans calage	57 553 918	23 009 997
avec calage	56 920 563	23 151 040
2) intégration de variables de comportement pour la CNRGH		
sans calage	57 464 226	23 168 631
avec calage	56 969 036	23 165 932
Vague 3 (poids transversal)		
1) pas de variables de comportement pour la CNRGH		
sans calage	55 853 357	23 220 977
2) intégration de variables de comportement pour la CNRGH		
sans calage	55 639 616	23 352 885
Enquête Emploi mars 1995	57 000 429	23 047 168

5.3 Eléments de structure de la population

Quel que soit le jeu de pondérations transversales adopté, le panel donne un peu moins d'étudiants, un peu plus d'enfants de moins de 15 ans. Peut-être y a-t-il simplement eu des erreurs sur l'enchaînement des questions lors du remplissage des questionnaires. Les jeunes semblent légèrement sous-représentés en structure dans le panel. On n'observe pas de différence majeure selon le système de poids utilisé.

En revanche, sans calage des poids, la proportion de ménages d'une personne est davantage sous-estimée lorsqu'on utilise la série de poids *sans* variable de comportement (27,6 %) que celle *avec* (28,2 %, contre 28,9 % dans l'enquête Emploi).

Le taux de propriétaires est légèrement surestimé dans le panel, par rapport à l'enquête Emploi, et d'autant plus qu'on utilise la série de poids sans variables de comportement. Les ménages de propriétaires sont en effet moins mobiles que les autres, donc plus souvent retrouvés lors de l'enquête. Leur poids relatif dans l'échantillon est d'autant plus grand qu'on ne prend pas en compte les déménagements [tableau 8].

Tableau 8 : Estimation de la proportion de ménages propriétaires de leur logement en vague 2

Système de pondérations transversales	Proportion de ménages propriétaires (en %)
Enquête emploi mars 1995	53,5
Panel vague 1	55,0
Prise en compte de variables de comportement pour la CNRGH et	
- pas de calage	55,6
- calage	55,3
Pas de variables de comportement pour la CNRGH et	
- pas de calage	56,2
- calage	55,6

5.4 Niveaux de vie et pauvreté

Lorsqu'on s'intéresse à la population dans son ensemble, les indicateurs usuels de distribution des niveaux de vie et de pauvreté apparaissent relativement peu affectés par le système de pondération utilisé [tableau 9]. Les inégalités et la pauvreté semblent légèrement plus importantes lorsqu'on intègre la variable de déménagement lors de la correction de la non-réponse, et que l'on recalcule l'échantillon, mais les différences restent relativement faibles (0,4 % sur le niveau de vie moyen, 0,3 points sur le taux de pauvreté). Cette constatation est plutôt rassurante.

Lorsqu'on décline les indicateurs selon des tranches d'âge, ou qu'on distingue le type de ménage, les différences sont en revanche plus marquées. Selon le système de poids retenu, l'estimation du taux de pauvreté des 15-29 ans varie ainsi entre 12,7 % et 13,3 %. De la même façon, l'estimation du taux de pauvreté des personnes seules est fortement liée au jeu de pondérations utilisé. L'attrition différentielle selon les catégories peut donc avoir des répercussions non négligeables sur les analyses de disparités intercatégorielles.

Tableau 9 : Quelques indicateurs transversaux de niveau de vie* et pauvreté des individus en 1994

	Pas de variables de comportement (déménagement, éclatement) pour la CNRGH		Intégration de variables de comportement (déménagement, éclatement) pour la CNRGH	
	sans calage	avec calage	sans calage	avec calage
<i>Ensemble de la population</i>				
- moyenne	103 717	103 603	103 421	103 282
- 1er décile (D1)	45 883	45 600	45 530	45 411
- 1er quartile (Q1)	62 430	62 320	62 320	62 155
- médiane (Med)	89 000	87 949	87 881	87 759
- 3ème quartile (Q3)	124 415	124 414	124 171	124 098
- 8ème décile (D9)	175 106	175 006	174 540	174 425
- rapport interdécile (D9/D1)	3,82	3,84	3,83	3,84
- indice de Theil	0,17117	0,17151	0,17156	0,17171
- variance des logarithmes	0,35268	0,35520	0,36322	0,36421
Taux de pauvreté (RUC<Med/2)	6,8	9,0	9,0	9,1
<i>Individus</i>				
- niveau de vie moyen	92 558	92 251	91 670	91 426
Taux de pauvreté (RUC<Med/2)	12,7	12,9	13,1	13,3
<i>Personnes seules</i>				
- niveau de vie moyen	91 801	91 733	90 771	90 669
Taux de pauvreté (RUC<Med/2)	17,1	17,2	17,8	17,8

*Le niveau de vie est défini comme le rapport entre le revenu total du ménage et son nombre d'unités de consommation (UC). L'échelle utilisée pour le calcul des UC accorde un poids de 1 au premier adulte du ménage, de 0,5 aux suivants et de 0,3 aux enfants de moins de 14 ans.

Un individu est dit « pauvre » si son niveau de vie est inférieur à la moitié du niveau médian. L'estimation du seuil de pauvreté varie donc avec l'estimation de la médiane, et se fait donc conditionnellement au système de pondération.

5.5 Trajectoire d'activité des individus panels adultes

Le panel européen permet de suivre mois par mois la trajectoire d'activité des individus adultes, à travers un calendrier d'activité. Ce calendrier permet de répertorier une vingtaine de types d'occupations. Concernant l'emploi, l'enquête indique son statut (indépendant, salarié), le type de contrat qui le lie à son employeur (CDD, CDI) et son temps de travail (temps plein, temps partiel). Le calendrier ne permet donc pas de repérer un changement d'employeur, mais il indique les modifications survenues dans le statut de l'individu. A côté de l'emploi sont distinguées les périodes de chômage, de maladie, d'études ou de formation, de retraite, etc. L'exploitation de ce calendrier a conduit à définir 19 types de trajectoires sur 12 mois, puis sur 24 mois, selon le nombre de changements intervenus au cours de la période et le type d'activité principale entre deux changements. Nous ne présenterons ici que les trajectoires sur 12 mois les plus fréquentes, ou celles dont l'estimation de la fréquence est fortement influencée par le système de pondérations longitudinales retenu [tableau 10a].

Globalement, le choix du système de poids longitudinaux ne semble pas influencer sur la répartition estimée de la population entre les différentes trajectoires d'activité définies. Quel que soit le système de poids, on constate une surestimation, par rapport à la vague 1 du panel, de la proportion de personnes constamment en formation en 1993 ; la proportion estimée de retraités est au contraire plus faible. Pour les autres trajectoires d'activité, les proportions d'individus concernés coïncident très exactement d'une estimation à l'autre, quelles que soient la série de poids et la vague d'enquête utilisées.

Lorsqu'on examine ces trajectoires d'activité selon l'âge de l'individu, le diagnostic reste identique [tableau 10b].

Lorsqu'on construit les trajectoires à partir des observations relatives aux 24 mois de 1993 et 1994, on retrouve les résultats présentés ici sur des trajectoires construites à partir des seules observations de l'année 1993.

Tableau 10a : Trajectoires d'activité des individus panel au cours de l'année 1993

	Population et poids de la vague 1	Pas de variables de comportement (déménagement, éclatement) pour la CNRGH		Intégration de variables de comportement (déménagement, éclatement) pour la CNRGH	
		sans calage	avec calage	sans calage	avec calage
		Aucun changement en 1993			
- CDI temps plein	32,5	32,7	32,7	32,6	32,6
- CDI temps partiel	3,0	3,1	3,0	3,1	3,0
- CDD temps plein	1,4	1,4	1,4	1,4	1,4
- indépendant	5,3	5,1	5,1	5,1	5,1
- chômage	2,9	2,6	2,7	2,7	2,7
- retraite	23,5	22,7	23,0	22,6	23,0
- formation	19,6	20,7	20,4	20,8	20,5
Au moins 1 changement en 93					
- passage au chômage	2,1	2,1	2,1	2,1	2,1
- sortie de chômage	1,4	1,4	1,4	1,4	1,4
- autres trajectoires	7,1	6,8	6,8	6,8	6,8
<i>Ensemble</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>

Les trajectoires présentées ici font état des changements de statut des personnes sur le marché du travail. Les changements de profession ou d'employeur leur échappent. Or, on peut penser qu'un certain nombre de déménagements sont liés à de tels changements. Une question du panel permet de repérer les individus dont la situation (profession, employeur) à la date d'enquête n'est plus la même que celle relevée lors de l'enquête précédente. On note alors que l'estimation de la proportion d'individus concernés est un peu plus sensible au choix du système de poids : sans intégration de la variable de déménagement lors de la correction de la non-réponse, cette proportion est évaluée à 11,9 % ; lorsqu'on intègre cette variable, l'estimation passe à 12,2 %. Appliqué aux données de la vague 3 du panel, l'exercice conduit à

peu de choses près au même résultat, avec des proportions estimées respectivement à 11,6 % et 11,8 %. Les différences constatées en vague 2 ne sont pas amplifiées en vague 3, où l'attrition, il est vrai, est de plus faible ampleur.

Tableau 10b : Trajectoires d'activité au cours de l'année 1993 des individus panel de moins de 30 ans

	Population et poids de la vague 1	Pas de variables de comportement (déménagement, éclatement) pour la CNRGH		Intégration de variables de comportement (déménagement, éclatement) pour la CNRGH	
		sans calage	avec calage	sans calage	avec calage
		Aucun changement en 1993			
- CDI temps plein	23,4	22,7	23,3	22,9	23,2
- CDI temps partiel	1,8	1,8	1,8	1,8	1,8
- CDD temps plein	3,0	2,9	2,9	2,9	2,9
- indépendant	1,1	1,3	1,3	1,3	1,3
- chômage	2,8	2,3	2,3	2,4	2,8
- formation	45,9	48,7	47,9	48,5	48,0
Au moins 1 changement en 93					
- passage au chômage	3,8	3,6	3,7	3,6	3,6
- sortie de chômage	2,4	2,3	2,4	2,4	2,4
- autres trajectoires	15,8	14,4	14,4	14,2	14,0
<i>Ensemble</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>	<i>100,0</i>

Conclusion

Le calcul des pondérations dans le cadre d'un panel implique de suivre les individus de façon précise et de bien identifier les caractéristiques des non-répondants.

Le choix du modèle de non-réponse peut influencer fortement sur les distributions de poids. Cependant, lorsqu'on compare les résultats d'analyses effectuées avec deux séries de pondérations issues de modèles différents, il s'avère que les résultats sont globalement proches. Les poids construits dans chacun des systèmes ne divergent en effet fortement que pour des segments très particuliers, et peu nombreux, de la population.

Les statistiques produites sur ces sous-populations particulières semblent, en revanche, être affectées par ce choix. Les écarts constatés sont statistiquement non significatifs mais, en l'absence de calcul de précision des estimations effectuées, ils peuvent conduire à des interprétations contradictoires. Supposons par exemple que le taux de pauvreté en 1993 était de 8,7 %. Selon le système de poids retenu, on conclura tantôt à une stabilité (8,8 %), tantôt à un accroissement (9,1 %) de la proportion de pauvres...

Eléments de bibliographie

CASES C., « Méthodologie du panel européen de ménages : exploitation des données de la vague 1 du fichier français », *Document de Travail de la Direction des Statistiques Démographiques et Sociales*, n° F9705, Insee, 1997.

CHAMBAZ C., SAUNIER J.-M., VALDELIEVRE H., « Méthodologie du panel européen de ménages : exploitation des données de la vague 2 du fichier français », *Document de Travail de la Direction des Statistiques Démographiques et Sociales*, n° F9715, Insee, décembre 1997.

DEVILLE J.-C., « Les enquêtes par panel : en quoi diffèrent-elles des autres enquêtes ? suivi de Comment attraper une population en se servant d'une autre ? », *Journées de Méthodologie Statistique*, 1998, à paraître.

EUROSTAT, « Groupe de Travail « Panel Communautaire de Ménages », Paris 18 et 19 Septembre 1995, Pondération Longitudinale », *Doc. PAN 51/95*, Eurostat, juillet 1995.

LAVALLEE P., « Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode de partage des poids », *Techniques d'enquête*, Vol.21, n°1, Statistique Canada, juin 1995.

LEGENDRE N., « Calcul des pondérations du fichier français de la vague 3 du panel européen de ménages », *Note interne*, 1998, à paraître.

Correction de la non-réponse par catégories homogènes : taux de non-réponse par catégories

Tableau A : Taux de non-réponse par catégories retenus pour le calcul des poids de base Vague 2 - Pas de variables de comportement (déménagement, éclatement du ménage)

				Effectif	Taux de non- réponse	
Enfants individus panel				4 062	0,0129	
Adultes individus panel				14 632	0,1144	
Couples + 1 enfant	PR française	PR artisan	-	180	0,1740	
		PR ouvrier	-	< 25 ans ≥ 25 ans	111 536	0,1023 0,0797
	PR autre CS	Hors aggl. Paris	< 25 ans	161	0,0828	
			≥ 25 ans	823	0,0690	
		Agglo. Paris	-	186	0,0563	
			-	104	0,1686	
	PR	-	-	89	0,0625	
	Couples + ≥ 3 enfants	PR française	PR artisan	-	190	0,0929
			PR ouvrier	-	< 25 ans ≥ 25 ans	190 430
		PR autre CS	Hors aggl. Paris	< 25 ans	219	0,0693
≥ 25 ans				476	0,0464	
Agglo. Paris			-	105	0,0676	
			-	262	0,1164	
PR		-	-	171	0,1893	
Ménages complexes		PR française	PR artisan	-	566	0,1128
			PR ouvrier	-	< 25 ans ≥ 25 ans	126 801
		PR autre CS	Hors aggl. Paris	< 25 ans	117	0,3880
	≥ 25 ans			60	0,1233	
	Agglo. Paris		-	649	0,1389	
	Autres ménages (*)	PR française	PR artisan	Hors aggl. Paris	94	0,2430
			PR ouvrier	-	< 25 ans ≥ 25 ans	290 2 105
PR autre CS		Hors aggl. Paris	< 25 ans	132	0,1994	
			≥ 25 ans	562	0,1239	
		Agglo. Paris	< 25 ans	3 802	0,0914	
			≥ 25 ans	124	0,1279	
PR		PR art. / ouvri.	-	767	0,1136	
PR		PR autre CS	-	282	0,2132	
PR		PR autre CS	-	112	0,1912	

PR = personne de référence du ménage

(*) : personnes seules, couples sans enfants, couples avec 2 enfants, familles monoparentales

Tableau B : Taux de non-réponse par catégories retenus pour le calcul des poids de base Vague 2 - intégration de variables de comportement (déménagement, éclatement du ménage)

						Effectif	Taux de non-réponse
Enfants individus panel						4 062	0,0129
Adultes individus panel						14 632	0,1144
Pas de	Couple	PR française	PR artisan	-	-	176	0,1724
déménagement ou	+ 1 enfant		PR ouvrier	-	-	630	0,0692
alors non			PR autre CS	Hors agg Paris	≤ bac.	669	0,0751
consécutif					> bac.	288	0,0486
à un				Agglo Paris	≤ bac.	90	0,0350
éclatement					> bac.	87	0,0288
de ménage		PR étrangère	-	-	-	100	0,1096
	Couple +	PR française	PR artisan	-	-	87	0,0343
	≥ 3 enfants		PR ouvrier	-	-	588	0,0562
			PR autre CS	Hors agg Paris	≤ bac.	452	0,0588
					> bac.	215	0,0301
				Agglo Paris	-	105	0,0676
		PR étrangère	-	-	-	255	0,0935
Famille mono-parentale	-		PR artisan ou ouvrier	-	-	148	0,1513
			PR autre CS	Hors agg Paris	-	378	0,0754
				Agglo Paris	-	98	0,0629
Autre type de ménage	PR française	PR artisan	Hors agg Paris	-	-	767	0,1464
			Agglo Paris	-	-	107	0,2239
		PR ouvrier	Hors agg Paris	≤ bac.	2 590	0,1177	
				> bac.	115	0,1021	
			Agglo Paris	-	186	0,1466	
		PR autre CS	Hors agg Paris	≤ bac.	3 461	0,0995	
				> bac.	1 172	0,0814	
			Agglo Paris	≤ bac.	581	0,1329	
				> bac.	360	0,0792	
		PR étrangère	PR ouvrier ou artisan	-	-	353	0,2346
			PR autre CS	-	-	139	0,1217
déménagement suite à un éclatement de ménage	Couple + 1 ou ≥3 enfants et famille monopar.	-	-	-	-	171	0,4096
	Autre type de ménage	-	PR artisan ou ouvrier	-	-	131	0,5888
		-	PR autre CS	-	-	134	0,4084

PR = personne de référence du ménage

Calcul des poids longitudinaux en vague 3

La correction de la non-réponse a été effectuée en modifiant les poids de base V2 selon la formule :

$$\text{Poids de base V3} = \frac{\text{Poids de base V2}}{1 - \text{taux de non réponse}}$$

Les catégories homogènes par rapport à la non-réponse ont été construites en croisant :

- dans le cas où la CNRGH n'intègre pas de variables de comportement : la catégorie socioprofessionnelle de la personne de référence, sa nationalité, l'âge de l'individu et le poids des charges de logement lorsque les effectifs concernés le permettaient. Un nombre très important de catégories homogènes (47) a été défini, les taux de non-réponse variant entre 0,7% et 17,3%. Certaines, très proches, auraient pu être regroupées.

- dans le cas contraire (intégration de variables de comportement) : le déménagement suite à un éclatement de ménage, la catégorie socioprofessionnelle de la personne de référence du ménage, sa nationalité et le type de ménage. 33 catégories ont ainsi été définies, avec des taux de non-réponse plus différenciés, compris entre 1,9% et 31,5%.

Distribution des poids de base V3 des individus panel adultes

L'objectif étant de suivre des trajectoires sur 3 ans, les poids de base V3 ont été calculés pour les seuls individus panel ayant répondu à la fois en vague 1 et en vague 2. Ceux qui n'avaient pas répondu en vague 2 avaient en effet un poids de base nul à cette date. Leur poids de base V3 était donc également nul.

Ce choix correspond à une restriction du champ de l'enquête, et conduit donc à éliminer des analyses tous les ménages répondant en vague 3 mais dont aucun individu n'avait répondu en vague 2. Une deuxième série de poids de base V3 a donc été calculée par correction non plus des poids de base V2 mais des poids de base V1. Nous ne présentons pas le modèle de correction ici, ni la distribution des poids qui en découle.

Tableau C : Poids de base (corrigés de la non-réponse) des 12 244 individus panel répondants en vague 3 et ayant répondu en vague 2

	Présence de variables de déménagement et éclatement du ménage pour la CNRGH		Rapport entre ces deux séries de poids
	NON (1)	OUI (2)	(3)=(2)/(1)
Somme	45 354 696	45 314 628	
Moyenne	3 704,24	3 700,97	1,00073
Ecart-type	1 229,89	1 273,76	0,10499
Minimum	2 056,79	2 010,54	0,72250
1%	2 313,79	2 265,97	0,85901
5%	2 498,77	2 474,24	0,93200
10% D1	2 660,57	2 632,96	0,95035
25% Q1	2 983,94	2 968,24	0,97100
50% Med	3 454,61	3 438,65	0,98875
75% Q3	4 110,09	4 104,66	1,02723
90% D9	4 841,15	4 846,42	1,00865
95%	5 404,15	5 478,19	1,06207
99%	9 553,46	9 682,93	1,52679
Maximum	16 620,98	24 438,25	2,25961
Range	14 564,19	22 427,71	-
Q3-Q1	1 126,14	1 136,42	-
D9/D1	1,82	1,84	

En vague 3 comme en vague 2, la distribution des poids de base des individus panel adultes est légèrement plus dispersée lorsqu'on intègre la variable de déménagement et d'éclatement de ménage à la grille de construction des catégories homogènes.

L'individu dont le poids était particulièrement élevé dans ce cas en vague 2 conserve un poids beaucoup plus gros que l'individu qui le suit immédiatement, mais sans accroître la distance. Lorsqu'on supprime cet individu, la dispersion est toujours légèrement plus forte que lorsque la CNRGH est réalisée sans variable de comportement. [tableau C]

La comparaison des poids des vagues 2 et 3 montre là encore une plus grande déformation des poids lorsqu'on intègre les variables de déménagement et d'éclatement de ménage. Sans ces variables, le rapport des poids V3/V2 s'établit en moyenne à 1,0696, variant entre 1,0072 et 1,2090. Avec ces variables, le rapport moyen V3/V2 est légèrement plus faible (1,0693), mais sa distribution couvre un intervalle plus large ([1,0192 ; 1,4590]). Le diagnostic est le même lorsqu'on compare les poids des vagues 1 et 3.