

**Rapport intermédiaire sur la qualité
des données françaises EU-SILC 2007**

TABLE DES MATIÈRES

Table des matiÈres	2
1 Indicateurs transversaux communs de l'Union européenne	4
1.1 Indicateurs communautaires transversaux communs fondés sur la dimension transversale des statistiques EU-SILC	4
1.1.1 Taux de pauvreté après transferts sociaux	4
1.1.2 Inégalité de la distribution des revenus Q80/Q20, ratio interquintile	9
1.1.3 Intensité de la pauvreté	9
1.1.4 Dispersion autour du seuil de pauvreté	9
1.1.5 Taux de pauvreté ancré à une date fixée (2005).....	10
1.1.6 Taux de pauvreté avant transferts	10
1.1.7 Inégalité de la distribution des revenus : coefficient de Gini	11
1.2 Autres indicateurs	11
1.2.1 Niveau de vie moyen	11
1.2.2 Écart de rémunération entre les sexes non ajusté	11
2 Précision	12
2.1 Plan d'échantillonnage	12
2.1.1 Type de plan d'échantillonnage	12
2.1.2 Unités d'échantillonnage.....	12
2.1.3 Critères de stratification et de sous-stratification	12
2.1.4 Taille de l'échantillon et critères d'attribution	13
2.1.5 Systèmes de sélection de l'échantillon	13
2.1.6 Répartition de l'échantillon au cours du temps et renouvellement de l'échantillon : groupes de rotation	13
2.1.7 Pondération.....	13
2.1.8 Substitutions	19
2.2 Erreurs d'échantillonnage	20
2.3 Erreurs non dues à l'échantillonnage.....	23
2.3.1 Erreurs dans la base de sondage et la couverture	23
2.3.2 Erreurs de mesure et de traitements	24
2.3.3 Erreurs de non-réponse	28
2.4 Mode de collecte des données	35
2.5 Durée de l'entretien	36
2.6 Les imputations	36
3 Comparabilité	42
3.1 Concepts et définitions de base	42
3.2 Composantes du revenu	43

3.2.1	Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC	43
3.2.2	La source ou la procédure utilisée pour la collecte des variables du revenu.....	46
3.2.3	La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes.....	46
3.2.4	La méthode utilisée pour obtenir les variables cibles du revenu sous la forme requise (c'est-à dire en tant que valeur brute)	46
4	CohÉrence	48
4.1	Comparaison des variables cibles avec des sources extérieures	48
4.1.1	Salaire et revenu assimilé (PY010N).....	48
4.1.2	Les allocations chômage (PY090N)	49
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension au survivant (PY110N)	50
4.1.4	Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante	50
4.1.5	Revenus sociaux (HY050G, HY060G et HY070G),	51
4.1.6	Prestations familiales (HY050N),.....	51
4.1.7	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation).	51
4.1.8	Loyers	52
4.1.9	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	52
4.1.10	Revenu disponible du ménage et niveau de vie des individus.....	53

1 INDICATEURS TRANSVERSAUX COMMUNS DE L'UNION EUROPEENNE

Les indicateurs ont été estimés grâce aux programmes fournis par Eurostat.

1.1 Indicateurs communautaires transversaux communs fondés sur la dimension transversale des statistiques EU-SILC

1.1.1 Taux de pauvreté après transferts sociaux

Le taux de pauvreté après transferts sociaux est défini comme la proportion de personnes ayant un niveau de vie inférieur au seuil de pauvreté. Tous les individus d'un ménage ont le même niveau de vie, qui est égal au revenu disponible divisé par le nombre d'unités de consommation du ménage. Le seuil de pauvreté est fixé à 60 % de la médiane du niveau de vie des individus.

1.1.1.1 Taux de pauvreté selon l'âge et le sexe

	Total			0-17	0 - 64			18+			18 - 64		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	13,1	13,8	12,4	15,7	13,1	13,6	12,5	12,4	13,1	11,6	12,1	12,8	11,5
Non pauvres	86,9	86,2	87,6	84,3	86,9	86,4	87,5	87,6	86,9	88,4	87,9	87,2	88,5

	18 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	20,8	20,7	20,9	10,6	11,8	9,4	10,9	11,0	10,8	13,4	14,3	12,2
Non pauvres	79,2	79,3	79,1	89,4	88,2	90,6	89,1	89,0	89,2	86,6	85,7	87,8

Effectifs non pondérés

	Total			0 - 17	0 - 64			18+		
	Total	F	H	Total	Total	F	H	Total	F	H
Total	25 907	13 374	12 533	6 314	22 019	11 218	10 801	19 593	10 273	9 320
Pauvres	3 299	1 773	1 526	968	2 806	1 474	1 332	2 331	1 307	1 024
Non pauvres	22 608	11 601	11 007	5 346	19 213	9 744	9 469	17 262	8 966	8 296

	18 - 64			18 - 24			25 - 49			50 - 64		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	15 705	8 117	7 588	2 356	1 192	1 164	8 405	4 367	4 038	4 944	15 705	8 117
Pauvres	1 838	1 008	830	436	228	208	874	509	365	528	1 838	1 008
Non pauvres	13 867	7 109	6 758	1 920	964	956	7 531	3 858	3 673	4 416	13 867	7 109

	65+		
	Total	F	H
Total	3 888	2 156	1 732
Pauvres	493	299	194
Non pauvres	3 395	1 857	1 538

Répartition de la population totale selon le sexe

Total	F	H
100,0	51,4	48,6

Répartition de la population totale selon l'âge et le sexe

	0 - 17	0 - 64	18+	18 - 64	18 - 24	25 - 49	50 - 64	65+	Total
Total	21,8	83,5	78,2	61,7	8,8	34,5	18,4	16,5	100,0
F	20,6	81,7	79,4	61,1	8,7	33,5	18,9	18,3	100,0
H	23,1	85,5	76,9	62,3	9,0	35,5	17,9	14,6	100,0

Répartition de la population pauvre selon le sexe

Total	F	H
100,0	53,9	46,1

Répartition de la population pauvre selon l'âge et le sexe

	0 - 17	0 - 64	18+	18 - 64	18 - 24	25 - 49	50 - 64	65+	Total
Total	26,1	83,1	73,9	57,0	14,0	27,8	15,2	16,9	100,0
F	24,2	80,9	75,8	56,7	13,1	28,6	15,0	19,1	100,0
H	28,2	85,7	71,8	57,5	15,1	26,9	15,4	14,3	100,0

1.1.1.2 Taux de pauvreté selon la situation professionnelle vis à vis de l'emploi la plus fréquente et le sexe

	Total			En emploi			Sans emploi : total			Sans emploi : chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	12,1	12,9	11,2	6,5	6,1	6,8	18,2	18,9	17,2	32,6	28,3	36,9
Non pauvres	87,9	87,1	88,8	93,5	93,9	93,2	81,8	81,1	82,8	67,4	71,7	63,1

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Pauvres	11,1	11,6	10,4	26,0	26,0	26,0
Non pauvres	88,9	88,4	89,6	74,0	74,0	74,0

Sont exclus du champ les personnes de moins de 18 ans au 1^{er} janvier de l'année de l'enquête et les individus qui n'ont pas gardé le même statut d'occupation pendant plus de la moitié de l'année.

Effectifs non pondérés

	Total			En emploi			Sans emploi : total			Sans emploi : chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	19 187	10 058	9 129	10 190	4 841	5 349	8 997	5 217	3 780	1 035	542	493
Pauvres	2 216	1 255	961	608	268	340	1 608	987	621	338	157	181
Non pauvres	16 971	8 803	8 168	9 582	4 573	5 009	7 389	4 230	3 159	697	385	312

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Total	4 801	2 410	2 391	3 161	2 265	896
Pauvres	490	251	239	780	579	201
Non pauvres	4 311	2 159	2 152	2 381	1 686	695

Répartition de la population totale selon la situation professionnelle la plus fréquente et le sexe

	Total	En emploi	Sans emploi : total	Sans emploi : chômeurs	Sans emploi : retraités	Sans emploi : autres inactifs
F	100,0	45,6	54,4	5,4	26,3	22,7
H	100,0	56,0	44,0	5,7	26,8	11,5
Total	100,0	50,6	49,4	5,6	26,6	17,3

Répartition de la population pauvre selon le statut professionnel le plus fréquent et le sexe

	Total	En emploi	Sans emploi : total	Sans emploi : chômeurs	Sans emploi : retraités	Sans emploi : autres inactifs
F	100,0	21,3	78,7	11,7	23,3	43,6
H	100,0	33,4	66,6	18,7	24,7	23,2
Total	100,0	26,7	73,3	14,8	23,9	34,6

1.1.1.3 Taux de pauvreté selon le type de ménage

	Sans enfant : total	Personnes seules (total)	2 adultes, moins de 65 ans	2 adultes, dont au moins un de 65 ans ou +	Sans enfant : autres
Pauvres	11,4	17,4	7,7	9,2	9,8
Non pauvres	88,6	82,6	92,3	90,8	90,2

	Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Pauvres	14,6	27,2	8,0	10,3	17,8	22,9
Non pauvres	85,4	72,8	92	89,7	82,2	77,1

Effectifs non pondérés

	Sans enfant : total	Personnes seules (total)	2 adultes, moins de 65 ans	2 adultes, dont au moins un de 65 ans ou +	Sans enfant : autres
Total	10 838	2 875	3 642	2 766	1 555
Pauvres	1 173	485	284	246	158
Non pauvres	9 665	2 390	3 358	2 520	1 397

Total	Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Total	15 069	1 531	3 099	5 160	3 645	1 634
Pauvres	2 126	390	276	420	681	359
Non pauvres	12 943	1 141	2 823	4 740	2 964	1 275

Taux de pauvreté des personnes seules selon l'âge ou le sexe

	F	H	< 65	65+
Pauvres	17,7	16,9	16,9	18,0
Non pauvres	82,3	83,1	83,1	82,0

Effectifs non pondérés

	F	H	< 65	65+
Total	1 782	1 093	1 641	1 234
Pauvres	312	173	263	222
Non pauvres	1 470	920	1 378	1 012

Répartition de la population totale selon le type de ménage

Sans enfant : total	Personnes seules (total)	2 adultes, moins de 65 ans	2 adultes, dont au moins un de 65 ans ou +	Sans enfant : autres
47,4	15,1	15,6	10,7	6,0

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
52,6	5,4	11,8	18,4	11,1	5,8

Répartition des personnes seules selon l'âge ou le sexe

F	H	< 65	65+
58,6	41,4	59,3	40,7

Répartition de la population pauvre selon le type de ménage

Sans enfant : total	Personnes seules (total)	2 adultes, moins de 65 ans	2 adultes, dont au moins un de 65 ans ou +	Sans enfant : autres
41,3	20,1	9,2	7,6	4,5

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
58,7	11,3	7,3	14,6	15,1	10,3

Répartition des personnes seules pauvres selon l'âge ou le sexe

F	H	< 65	65+
59,8	40,2	57,8	42,2

1.1.1.4 Taux de pauvreté selon le statut d'occupation du logement

Les propriétaires ont un risque de pauvreté plus faible que les locataires même quand le taux de pauvreté est estimé comme ici sans prendre en compte les loyers imputés.

	Total	Locataires	Propriétaires ou logés à titre gratuit
Pauvres	13,1	21,1	9,0
Non pauvres	86,9	78,9	91,0

Effectifs non pondérés

	Total	Locataires	Propriétaires ou logés à titre gratuit
Total	25 907	8 027	17 880
Pauvres	3 299	1 772	1 527
Non pauvres	22 608	6 255	16 353

Répartition de la population totale selon le statut d'occupation du logement

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	31,0	69,0

Répartition de la population pauvre selon le statut d'occupation du logement

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	53,7	46,3

1.1.1.5 Taux de pauvreté selon le type de ménage et l'intensité d'emploi

L'intensité d'emploi dans le ménage (W) mesure la part du temps passé en emploi pour les personnes du ménage en âge de travailler (entre 18 et 64 ans) sur une période donnée. Il s'agit précisément du rapport entre le nombre de mois travaillés et le nombre de mois travaillables (mois travaillés, au chômage, en retraite ou en inactivité) pour l'ensemble des personnes du ménage. L'intensité d'emploi est donc comprise entre 0 et 1. Une valeur de 0 correspond à l'absence d'emploi au sein du ménage sur la période, une valeur de 1 à un nombre de mois travaillés égal au nombre de mois travaillables (emploi saturé).

	Sans enfant W=0	Sans enfant 0<W<1	Sans enfant W=1	Avec enfants W=0	Avec enfants 0<W<0,5	Avec enfants 0,5<W<1	Avec enfants W=1
Pauvres	18,7	9,4	5,0	70,9	43,8	15,8	5,4
Non pauvres	81,3	90,6	95,0	29,1	56,2	84,2	94,6

Sont exclus du champ les ménages composés seulement de personnes de moins de 18 ans, de plus de 65 ans ou n'ayant pas travaillé au moins un mois au cours de l'année de référence ou d'étudiants.

Effectifs non pondérés

	Sans enfant W=0	Sans enfant 0<W<1	Sans enfant W=1	Avec enfants W=0	Avec enfants 0<W<0,5	Avec enfants 0,5<W<1	Avec enfants W=1
Total	1 830	2 618	3 148	842	851	4 959	8 406
Pauvres	328	235	147	602	375	754	387
Non pauvres	1 502	2 383	3 001	240	476	4 205	8 019

Répartition de la population totale selon le type de ménage et l'intensité d'emploi

Total	Sans enfant W=0	Sans enfant 0<W<1	Sans enfant W=1	Avec enfants W=0	Avec enfants 0<W<0,5	Avec enfants 0,5<W<1	Avec enfants W=1
100,0	8,1	11,6	13,9	3,7	3,8	21,9	37,1

Répartition de la population pauvre selon le type de ménage et l'intensité d'emploi

Total	Sans enfant W=0	Sans enfant 0<W<1	Sans enfant W=1	Avec enfants W=0	Avec enfants 0<W<0,5	Avec enfants 0,5<W<1	Avec enfants W=1
100,0	11,6	8,3	5,2	21,3	13,3	26,7	13,7

1.1.1.6 Seuil de pauvreté (valeurs indicatives)

En 2006, le seuil de pauvreté s'élève à 9 940 euros par an. Une personne seule est pauvre si elle a un revenu disponible inférieur à 828 euros par mois et un couple avec deux enfants de moins de 14 ans, s'il a un revenu disponible inférieur à 20 874 euros (soit 1 740 euros par mois).

1.1.2 Inégalité de la distribution des revenus Q80/Q20, ratio interquintile

Le rapport interquintile est le rapport entre la somme des niveaux de vie des 20% de personnes ayant les niveaux de vie les plus élevés et celle des 20% de personnes ayant les niveaux de vie les plus faibles.

Le cinquième de la population le plus riche possède 3,8 fois plus de niveaux de vie, en masse, que le cinquième de la population le plus pauvre.

1.1.3 Intensité de la pauvreté

L'intensité de la pauvreté est estimée comme l'écart de la médiane des niveaux de vie de la population pauvre au seuil de pauvreté. Si les pauvres n'avaient aucun revenu, cet écart serait de 100 %. S'ils avaient tous un niveau de vie très proche du seuil de pauvreté, l'écart serait presque nul.

Total			0 - 17		18+		18 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
16,8	16,5	17,2	14,8	17,8	17,2	18,6	17,4	16,9	18,1	18,9	18,9	19,0

Effectifs non pondérés

Total			0 - 17		18+		18 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
3 299	1 773	1 526	968	2 331	1 307	1 024	1 838	1 008	830	493	299	194

1.1.4 Dispersion autour du seuil de pauvreté

Le seuil de pauvreté pourrait aussi être fixé à 40, 50 et 70 % de la médiane du niveau de vie, L'étude des taux de pauvreté ainsi obtenus permet d'apprécier la concentration des niveaux de vie autour du seuil de pauvreté.

	Total	F	H
<i>40 % de la médiane</i>			
Pauvres	3,0	3,1	2,8
Non pauvres	97,0	96,9	97,2
<i>50 % de la médiane</i>			
Pauvres	6,6	6,8	6,3
Non pauvres	93,4	93,2	93,7
<i>70 % de la médiane</i>			
Pauvres	21,0	21,9	19,9

	Total	F	H
Non pauvres	79,0	78,1	80,1

Effectifs non pondérés

	total	F	H
Total	25 907	13 374	12 533
40 % de la médiane			
Pauvres	744	400	344
Non pauvres	25 163	12 974	12 189
50 % de la médiane			
Pauvres	1 662	892	770
Non pauvres	24 245	12 482	11 763
70 % de la médiane			
Pauvres	5 285	2 854	2 431
Non pauvres	20 622	10 520	10 102

1.1.5 Taux de pauvreté ancré à une date fixée (2005)

Il s'agit du taux de pauvreté calculé en 2007, en prenant comme seuil celui de 2005, corrigé de l'inflation.

	Total			0-17	0 - 64			18+			18 - 64		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	13,0	13,6	12,3	15,6	12,9	13,5	12,3	12,3	13,0	11,5	12,0	12,6	11,3
Non pauvres	87,0	86,4	87,7	84,4	87,1	86,5	87,7	87,7	87,0	88,5	88,0	87,4	88,7

	18 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	20,6	20,4	20,7	10,5	11,7	9,2	10,7	10,7	10,7	13,3	14,0	12,2
Non pauvres	79,4	79,6	79,3	89,5	88,3	90,8	89,3	89,3	89,3	86,7	86,0	87,8

Effectifs non pondérés

	Total			0-17	0 - 64			18+			18 - 64		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	3 263	1 752	1 511	959	2 775	1 457	1 318	2 304	1 291	1 013	1 816	996	820
Non pauvres	22 644	11 622	11 022	5 355	19 244	9 761	9 483	17 289	8 982	8 307	13 889	7 121	6 768

	18 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	430	225	205	863	504	359	523	267	256	488	295	193
Non pauvres	1 926	967	959	7 542	3 863	3 679	4 421	2 291	2 130	3 400	1 861	1 539

1.1.6 Taux de pauvreté avant transferts

Les transferts considérés sont les prestations sociales, y compris les revenus de remplacement (chômage, retraites). Le seuil de pauvreté est celui estimé à partir du revenu disponible après transferts.

1.1.6.1 Taux de pauvreté avant tout transfert social selon l'âge et le sexe

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	46,2	48,8	43,3	36,9	48,7	51,7	45,4	36,2	38,4	33,9	95,3	95,7	94,8
Non pauvres	53,8	51,2	56,7	63,1	51,3	48,3	54,6	63,8	61,6	66,1	4,7	4,3	5,2

Effectifs non pondérés

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Total	25 907	13 374	12 533	6 314	19 593	10 273	9 320	15 705	8 117	7 588	3 888	2 156	1 732
Pauvres	11 377	6 151	5 226	2 321	9 056	5 005	4 051	5 379	2 954	2 425	3 677	2 051	1 626
Non pauvres	14 530	7 223	7 307	3 993	10 537	5 268	5 269	10 326	5 163	5 163	211	105	106

1.1.6.2 Taux de pauvreté avant transferts sociaux autres que retraites selon l'âge et le sexe

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	25,8	26,6	25,0	35,5	23,1	24,1	22,0	24,5	25,6	23,3	18,0	19,1	16,6
Non pauvres	74,2	73,4	75,0	64,5	76,9	75,9	78,0	75,5	74,4	76,7	82,0	80,9	83,4

Effectifs non pondérés

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Total	25 907	13 374	12 533	6 314	19 593	10 273	9 320	15 705	8 117	7 588	3 888	2 156	1 732
Pauvres	6 610	3 523	3 087	2 229	4 381	2 420	1 961	3 745	2 032	1 713	636	388	248
Non pauvres	19 297	9 851	9 446	4 085	15 212	7 853	7 359	11 960	6 085	5 875	3 252	1 768	1 484

1.1.7 Inégalité de la distribution des revenus : coefficient de Gini

Le coefficient de Gini permet de mesurer l'écart de la distribution des revenus à une distribution purement égalitaire. Il s'élève à 26,4 % en 2006.

1.2 Autres indicateurs

1.2.1 Niveau de vie moyen

En 2006, le niveau de vie des individus s'élève en moyenne à 18488 euros par unité de consommation, soit 1 541 euros par mois.

1.2.2 Écart de rémunération entre les sexes non ajusté

Ils sont fournis par l'enquête Emploi qui dispose d'un plus grand échantillon et de données plus précises sur le temps de travail.

2 PRÉCISION

2.1 Plan d'échantillonnage

2.1.1 Type de plan d'échantillonnage

Le système d'échantillonnage de l'enquête SILC en France repose sur un échantillon-maître. Ce dernier constitue par définition une réserve de logements regroupant, d'une part des logements recensés en mars 1999, d'autre part des logements achevés après cette date, appelés « logements neufs ».

L'unité d'échantillonnage gérée dans ces bases de sondage est le logement. Ce qui se passe au sein du logement relève du terrain : en l'occurrence, dans cette enquête, lorsque l'échantillon est entrant, on interroge l'ensemble des individus de chaque ménage présent dans le logement. L'échantillonnage de logements est stratifié, à plusieurs degrés : selon les strates, il y a deux ou trois degrés de tirage. Dans tout échantillon entrant, puisqu'on est exhaustif au dernier degré de sondage, on peut dire qu'il s'agit d'un sondage d'individus en grappes.

2.1.2 Unités d'échantillonnage

Au premier degré, l'unité d'échantillonnage est un groupe de communes contiguës. En zone rurale, ce groupe correspond à un découpage ad-hoc. En revanche, en zone urbaine l'unité primaire est, soit un ensemble d'unités urbaines, soit une unité urbaine (l'unité urbaine est une entité regroupant des communes, définie à partir des données du recensement). Au second degré, l'unité secondaire est, soit un groupe ad-hoc de communes lorsqu'on se situe en zone rurale, soit un logement si on est en zone urbaine. Le troisième degré n'existe qu'en zone rurale, et les unités de tirage correspondantes sont des logements.

2.1.3 Critères de stratification et de sous-stratification

Le partage rural/urbain constitue une première stratification préalable au premier degré de tirage. Elle est enrichie de deux manières. D'une part, la strate urbaine distingue en fait trois niveaux :

- unités urbaines de moins de 20 000 habitants (« petit urbain »),
- unités urbaines comprises entre 20 000 et 100 000 habitants (« moyen urbain »),
- unités urbaines de plus de 100 000 habitants (« grand urbain »),

D'autre part, chaque strate ainsi définie -y compris la strate rurale- est elle-même ventilée en sous-strates selon la région administrative. Comme il y a 22 régions administratives, la stratification préalable comporte 88 strates (dont 2 sont vides).

En zone rurale, il n'y a pas de stratification préalable au second degré. En revanche, dans les strates urbaines, il y a une sous-stratification constituée par des groupes de communes. Il n'y a pas de critères sociodémographiques utilisés pour définir ces groupes de communes, on peut seulement dire qu'il s'agit de communes contiguës regroupées afin d'atteindre une certaine taille - mais cela permet au moins de distinguer ville-centre et banlieue.

Enfin, l'ultime stratification est préalable au tirage des logements dans les groupes de communes tirés. Elle distingue quatre types de logements :

- les logements recensés de type « résidence principale » ;
- les logements recensés de type « résidence secondaire » ;
- les logements recensés de type « résidence vacante » ;
- les logements neufs,

2.1.4 Taille de l'échantillon et critères d'attribution

L'échantillon 2004 comprenait 16 000 logements (ventilés en 9 sous-échantillons), et les sous-échantillons entrant en 2005 et 2006 en comprenait chacun 3 000. Les sous-échantillons entrants des années à venir devraient être également calibrés à 3 000 logements.

2.1.5 Systèmes de sélection de l'échantillon

L'échantillon de logements de SILC est tiré à probabilités égales (taux f) à l'exception des logements recensés secondaires qui sont tirés avec la probabilité $f/4$ et des logements recensés vacants en strate rurale qui sont tirés au taux $f/2$.

Cas de la strate rurale

Le tirage des 128 unités primaires de l'échantillon-maître parmi les 227 recensées a été effectué proportionnellement à la taille définie en nombre de résidences principales, selon un algorithme de tirage équilibré (l'équilibrage a porté sur des structures par âge et sur le revenu déclaré aux services fiscaux). Le tirage de 719 unités secondaires de l'échantillon de EU-SILC sur les 3 217 unités de l'échantillon-maître a également été proportionnel à la taille en nombre de résidences principales. Le tirage des logements est effectué au sein du groupe de communes par un algorithme systématique sur fichier trié.

Le choix des tailles des échantillons d'unités primaires et secondaires est dicté par des considérations de coût de déplacement et de charge moyenne des enquêteurs, sachant que la règle générale consiste à affecter un enquêteur à une unité primaire.

Cas des strates urbaines jusqu'à 100 000 habitants

Dans le petit urbain, l'échantillon-maître comprend 75 unités primaires parmi les 975 recensées et dans le moyen urbain, 93 unités primaires parmi 180.

L'échantillonnage des unités primaires s'effectue selon les mêmes modalités qu'en strate rurale. Même chose pour le tirage des logements au sein des groupes de communes. Les tailles d'échantillons sont déterminées par les mêmes règles qu'en strate rurale.

Cas du « grand urbain »

Les 53 unités urbaines de plus de 100 000 habitants sont retenues. Dans chaque groupe de communes de l'unité urbaine, le tirage des logements est effectué par un algorithme systématique sur fichier trié. La taille de l'échantillon découle directement de la probabilité de sélection des logements.

2.1.6 Répartition de l'échantillon au cours du temps

L'échantillon est de type rotatif : chaque année, on interroge neuf sous-échantillons, tirés chacun selon la méthode exposée ci-dessus. Chaque sous-échantillon est donc un panel interrogé neuf années de suite. Les tirages des sous-échantillons sont indépendants d'une année sur l'autre mais, par construction, ils sont toujours effectués dans les mêmes unités primaires puisque celles-ci constituent un échantillon-maître.

2.1.7 Renouvellement de l'échantillon : groupes de rotation

En 2007, l'échantillon entrant était constitué de 3 000 logements tirés de la façon décrite plus haut. L'échantillon des réinterrogés était constitué des logements où habitaient en 2006 au moins une personne répondante et n'appartenant pas à l'échantillon sortant. Il comprenait 9 934 logements.

2.1.8 Pondération

La pondération se fait au niveau logement.

En 2007, nous avons estimé pour la troisième fois des poids pour un échantillon comprenant à la fois des entrants et des réinterrogés. Nous avons suivi la méthodologie présentée par Pascal Ardilly et Pierre Lavallée (Symposium 2003, Stat Canada).

2.1.8.1 Effet du plan d'échantillonnage (entrants)

Les poids sont définis au niveau logement et sont égaux à l'inverse de la probabilité d'inclusion qui découle du plan de sondage. Les ménages et les individus ont la même pondération que le logement auquel ils appartiennent. Chaque année, les pondérations initiales, notées WI, varient dans un rapport de 1 à 4 puisque les logements recensés secondaires ont quatre fois moins de chance d'être tirés que les logements recensés principaux ou que les logements neufs (cf. 2.1.5).

Si l'on regarde l'ensemble des sous-échantillon, le poids WI varie dans un rapport de 1 à 7 en raison de la différence de taille entre l'échantillon de 2004 et les échantillons suivants.

2.1.8.2 Ajustements pour non-réponse

Pour les réinterrogés, il y a deux types de non-réponse totale : la non-réponse des individus au moment du tirage « initial », c'est-à-dire lorsqu'on échantillonne le panel entrant, et la non-réponse à la date courante. Les modèles de non-réponse ont été estimés pour les deux types de non-réponse.

La probabilité de non-réponse initiale est notée PI.

La correction de la non-réponse initiale

La correction de la non-réponse initiale pour les entrants en 2007

Les modèles de non-réponse ont été estimés sur les logements de l'échantillon qui faisaient partie du champ en 2007 : les enquêteurs les ont reconnus comme résidence habituelle d'au moins une personne. Nous avons différencié trois types de modèles de non-réponse totale selon les informations disponibles dans les bases de sondage ou recueillies par l'enquêteur en 2007 :

- les résidences principales au dernier recensement de la population (RP) et en 2007,
- les résidences recensées secondaires, occasionnelles ou vacantes au RP,
- les logements neufs (construits après mars 1999).

Toutes catégories de logements confondues, les motifs principaux de non-réponse sont liés à :

- la région,
- le fait d'habiter une maison ou un appartement en 2007,
- le nombre de pièces du logement en 1999,
- l'âge de la personne de référence (plus ou moins de 60 ans) en 1999,

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les 1 824 ménages répondants dans 11 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité de réponse initiale d'un ménage dans une sous-population donnée vaut donc :

$$PI = \frac{\text{Nombre de ménages répondants dans la sous-population}}{\text{Nombre de ménages échantillonnés dans la sous population}}$$

La correction de la non-réponse initiale pour les réinterrogés

La probabilité de non-réponse initiale a été estimée de la même façon sur les échantillons entrants en 2004, 2005 et 2006 : le nombre de groupe et les motifs principaux de non réponse variant à la marge. Par ailleurs en 2004, compte tenu de la taille d'échantillon plus grande (tous les échantillons étaient entrants, ce sont cinq modèles de non-réponses qui ont été estimés et non trois (cf. Rapport final sur la qualité des données françaises EU-SILC 2006).

Pour l'ensemble des ménages, PI est comprise entre 0,5 et 0,9. Le poids corrigé de la non-réponse initiale (WI/PI) varie dans un rapport de 1 à 10.

La correction de la non-réponse en réinterrogation

Les modèles de non-réponse sont estimés au niveau individuel. Une probabilité de réponse est attribuée à chaque individu panel qui est encore dans le champ en 2007. Est considéré comme répondant tout individu qui appartient à un ménage répondant (DB135='1').

On estime trois (2007-2004) modèles sur tous les individus qui répondent en même interrogation depuis 2004, soit :

- un modèle pour les individus interrogés en 2ème vague
 - o sur les entrants 2004 encore dans le champ 2005
 - o sur les entrants 2005 encore dans le champ 2006
 - o sur les entrants 2006 encore dans le champ 2007

- un modèle pour les individus interrogés en 3ème vague :
 - o sur les entrants 2004 encore dans le champ 2006
 - o sur les entrants 2005 encore dans le champ 2007

- un modèle pour les individus interrogés en 4ème vague :
 - o sur les entrants 2004 réinterrogés en 2007

Cela suppose une stabilité des comportements de non réponse dans le temps.

Probabilité de répondre en 2ème vague sachant qu'on a répondu à la première

Un modèle de non-réponse a été estimé sur les adultes réinterrogés en 2^{ème} vague. L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir déménagé avec l'ensemble de son ménage depuis la dernière vague,
- le fait d'habiter dans une maison ou non
- la situation familiale (seul/non),
- avoir un contrat à durée indéterminée ou non
- la localisation géographique (agglomération parisienne/autre),
- le quartile de niveau de vie (1er, 2d ou au-dessus de la médiane) du ménage.

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, les adultes réinterrogés en 2ème vague ont été réparti en neuf groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(2ème vague) vaut donc :

$$\frac{\text{Nombre d'individus répondants dans la sous-population}}{\text{Nombre d'individus panel dans la sous population}}$$

La probabilité $PR(2^{\text{ème}} \text{ vague}/1^{\text{ère}} \text{ vague})$ est comprise entre 0,5 et 0,9.

Probabilité de répondre à la 3ème vague sachant qu'on a répondu à la 1ère vague

Un modèle de non-réponse a été estimé sur les adultes panel répondants en 2005 et réinterrogés en 2007 ou répondants en 2004 et réinterrogés en 2006. L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir déménagé avec l'ensemble de son ménage depuis la dernière vague,
- le fait d'habiter dans une maison ou non
- le quartile de niveau de vie (1^{er} ou plus) du ménage ,
- la localisation géographique (agglomération parisienne/autre),
- la situation familiale (seule/non),
- avoir un contrat à durée indéterminée ou non

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les individus dans 10 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité $PR(3^{\text{ème}} \text{ vague}/1^{\text{ère}} \text{ vague})$ vaut donc :

$$\frac{\text{Nombre d'individus répondants dans la sous-population}}{\text{Nombre d'individus panel dans la sous population}}$$

Pour les enfants, pour lesquels nous n'avons que peu d'information, nous appliquons un taux de réponse homogène.

La probabilité $PR(3^{\text{ème}} \text{ vague}/1^{\text{ère}} \text{ vague})$ est comprise entre 0,5 et 0,9.

Probabilité de répondre en 4ème vague sachant qu'on a répondu à la 1ère vague

Un modèle de non-réponse a été estimé sur les adultes panel interrogés en 4ème vague et répondants en 1ère vague, soit les entrants en 2004 encore dans le champ en 2007. L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir déménagé depuis 2004,
- la nationalité (magrébine/autre),
- la situation familiale (seule/non),
- avoir en contrat à durée indéterminée ou non
- le quartile de niveau de vie (1^{er} ou plus) du ménage,

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les individus dans 7 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité $PR(4^{\text{ème}} \text{ vague}/1^{\text{ère}} \text{ vague})$ vaut donc :

$$\frac{\text{Nombre d'individus répondants dans la sous-population}}{\text{Nombre d'individus panel dans la sous population}}$$

La probabilité $PR(4^{\text{ème}} \text{ vague}/1^{\text{ère}} \text{ vague})$ est comprise entre 0,5 et 0,9.

Pour les enfants (moins de 16 ans en 2004), pour lesquels nous n'avons que peu d'information, nous appliquons un taux de réponse homogène.

Partage des poids

Soit :

- $a_{2007,k}$ le sous-échantillon panel à enquêter l'année t en $k^{\text{ième}}$ interrogation
- $W_i(2007,k)$ le poids de sondage « brut » de l'individu i : il s'agit du poids de sondage du logement dans lequel se trouve i à la date de son tirage en tant qu'individu-panel, soit en 2007- $k+1$.
- $\Phi_l(2007,k)$ la probabilité de répondre en première interrogation des ménages du logement l , c'est-à-dire en $t - k + 1$.
- $\theta_i(2007,k)$ la probabilité de répondre en 2007 sachant qu'on appartient à l'échantillon $a_{2007,k}$ et qu'on a répondu en $2007 - k + 1$.
- $Npan_l(2007)$ est le nombre d'individus panel du logement l en 2007.
- $Nchamp_l^{t,t'}(2007)$ est le nombre d'individus habitant le logement l en 2007 qui étaient dans le champ en t' .
- $Nbchamp_l^{t,t'}(2007)$ est la somme des années passés dans le champ de t à t' par les individus appartenant au logement l en 2007.
- $Npers07$ est le nombre d'individus dans le logement en 2007.

En 2007, on attribue à chaque individu panel répondant le poids suivant :

$$W_i^{*2007} = \frac{W_i(2007,k)}{9} \times \frac{1}{\Phi_l(2007,k)} \times \frac{1}{Npan_l(2007)} \sum_{\substack{j \in l \\ j \in r_{t,k}}} \frac{1}{\theta_j(2007,k)} \\ \times \frac{9 \times Npan_l(2007)}{6 \times Nchamp_l^{2004}(2007) + Nchamp_l^{2005,2006}(2007) + Npers_l(2007)}$$

Tous les individus répondant d'un même logement ont donc le même poids qu'on peut décomposer en le poids de tirage corrigé de la non-réponse initiale puis de la non-réponse en réinterrogation des individus panel présents en 2007 et du partage des poids.

Où la correction relative au partage des poids correspond à :

$$\frac{9 \times Npan_l(2007)}{6 \times Nchamp_l^{2004}(2007) + Nchamp_l^{2005,2006}(2007) + Npers_l(2007)}$$

Le poids après le partage des poids varie dans un rapport de 1 à 52.

2.1.8.3 Ajustements aux données extérieures (niveau, variables utilisées et sources) et « troncature »

L'enquête, réalisée en mai 2007, est, comme toutes les enquêtes après des ménages de l'Insee, calée sur les marges issues de l'enquête Emploi de 2006. En effet, toutes les enquêtes ménages

d'une année civile N sont calées sur la dernière enquête Emploi disponible au début de l'année N. Toutes les enquêtes dont la collecte a lieu pendant l'année civile N sont donc calées sur les marges constituées à partir des 4 enquêtes Emploi trimestrielles de l'année N-1.

Contrairement aux autres enquêtes ménages de l'Insee, l'enquête Emploi ignore la notion de budget séparé, et un logement n'est occupé que par un seul ménage. Il a donc fallu, dans un premier temps, reconstituer cette notion de ménage-logement et déterminer la personne de référence afin de pouvoir réaliser le calage.

Le calage et la correction de la non-réponse de seconde interrogation ont été effectués avec le logiciel Calmar 2. Les poids en entrée sont ceux issus du partage des poids. Nous avons utilisé la fonction de distance « logit ».

Les variables et les modalités introduites dans le calage sont les suivantes. Il s'agit du :

- nombre de ménages par tranche d'âge de la personne de référence (5 modalités, des moins de 31 ans aux 76 ans et plus). L'âge est mesuré au 31/12/2006.
- nombre de ménages par tranche de densité d'habitat : rural, unité urbaine de moins de 20 000 habitants, unité urbaine de 20 000 à 100 000 habitants, unité urbaine de plus de 100 000 habitants, Région parisienne.
- nombre de ménages par type :
 - personne seule,
 - couple sans enfant,
 - couple avec 1 enfant,
 - couple avec 2 enfants ou plus,
 - famille monoparentale,
 - autre configuration ;
- nombre d'hommes par tranche d'âge (6 modalités, des moins de 15 ans aux 76 ans et plus) ;
- nombre de femmes par tranche d'âge (6 modalités) ;
- nombre de ménages selon le diplôme de la personne de référence :
 - sans diplôme, non déclaré,
 - diplôme inférieur baccalauréat (CAP, BEPC),
 - Baccalauréat, bac+2,
 - diplôme supérieur ;
- nombre de ménages selon la catégorie socioprofessionnelle de la personne de référence (activité actuelle ou ancienne activité) :
 - agriculteurs (retraités ou non),
 - indépendants et professions libérales (retraités ou non),
 - professeurs et instituteurs actifs,
 - professeurs et instituteurs retraités,
 - employés et ouvriers actifs,
 - employés et ouvriers retraités,
 - autres ;
- nombre de ménages par région :
 - Île-de-France et bassin parisien,
 - Nord, Est et Ouest,
 - Sud-Ouest, Centre-Est et Méditerranée.

Les rapports de poids avant/après calage évoluent entre 0,5 et 2,5. Les pondérations des logements issues de ce premier calage varient quant à elles dans un rapport allant de 1 à 80.

2.1.8.4 Pondération transversale finale

Les pondérations (au niveau logement) obtenues à l'issue du calage sont affectées à tous les ménages répondants qui occupent ce logement et à tous les individus qui appartiennent à un ménage répondant.

Pour les répondants, le poids est modifié pour corriger de la non-réponse individuelle. Les pondérations des individus répondants sont corrigées de façon à garder la même structure par âge et par sexe (16 modalités : croisement du sexe par l'âge en 8 modalités, des moins de 16 ans aux 76 ans et plus) sur l'ensemble de la population. Elle ne respecte donc pas la structure par âge et par sexe au sein d'un ménage. Comme le taux de non-réponse individuel n'est que de 0,3 %, l'impact de cette correction est très limité.

Pour les enfants âgés de 2 à 12 ans, le poids est modifié de façon à avoir la même structure par année de naissance que dans l'enquête Emploi.

Il existe donc trois systèmes de poids pour les individus : le premier (RB050) est parfaitement cohérent avec le système de poids des ménages, les autres permettent de limiter les biais des estimateurs calculés seulement sur certains individus (PB040 pour les questions du questionnaire individuel et RL070 pour les questions relatives aux modes de garde).

2.1.9 Substitutions

Sans objet dans le cas de la France.

2.2 Erreurs d'échantillonnage

Les écarts-type calculés estiment l'écart-type causé par l'échantillonnage corrigé de la non-réponse et du calage. Ils sont linéarisés puis estimés en utilisant le logiciel Poulpe développé à l'INSEE après linéarisation.

La variance du taux de pauvreté obtenue est supérieure de 5 % à celle obtenue par un tirage aléatoire simple de même taille. Cet effet de sondage s'explique par l'utilisation d'un tirage à plusieurs degrés.

En raison du trop grand nombre de valeurs nulles, nous ne sommes pas parvenus à linéariser le taux de pauvreté avant tout transfert et nous n'avons estimé ni écart-type ni effet de sondage pour cet indicateur.

Indicateurs	Valeur	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Taux de pauvreté après prestations sociales (en pourcentage)						
Total	13,1	0,4	3,1	1,05	25 907	24 693
Hommes total	12,4	0,4	3,2	1,07	12 533	11 755
Femmes total	13,8	0,4	2,9	1,05	13 374	12 749
0-17 ans	15,7	0,8	4,8	1,03	6 314	6 111
18-24 ans	20,8	1,0	4,6	1,08	2 356	2 180
25-49 ans	10,6	0,4	3,6	1,04	8 405	8 064
50-64 ans	10,9	0,5	4,8	1,05	4 944	4 694
65+ ans	13,4	0,6	4,8	1,24	3 888	3 136
18+ ans	12,4	0,3	2,6	1,07	19 593	18 343
18-64 ans	12,1	0,3	2,9	1,04	15 705	15 065
0-64 ans	13,1	0,4	3,1	1,03	22 019	21 344
Hommes 18-24 ans	20,9	1,2	5,8	1,08	1 164	1 073
Hommes 25-49 ans	9,4	0,4	4,4	1,10	4 038	3 674
Hommes 50-64 ans	10,8	0,6	5,9	1,10	8 117	7 350
Hommes 65+ ans	12,2	0,8	6,5	1,16	1 732	1 487
Hommes 18+ ans	11,6	0,4	3,0	1,11	9 320	8 428
Hommes 18-64 ans	11,5	0,4	3,3	1,08	7 588	7 020
Hommes 0-64 ans	12,5	0,4	3,5	1,03	10 801	10 464
Femmes 18-24 ans	20,7	1,2	5,8	1,10	1 192	1 083
Femmes 25-49 ans	11,8	0,5	3,9	1,03	4 367	4 220
Femmes 50-64 ans	11	0,6	5,6	1,03	15 705	15 245
Femmes 65+ ans	14,3	0,7	5,1	1,24	2 156	1 743
Femmes 18+ ans	13,1	0,4	2,7	1,07	10 273	9 591
Femmes 18-64 ans	12,8	0,4	3,1	1,04	8 117	7 826
Femmes 0-64 ans	13,6	0,4	3,3	1,05	11 218	10 721

Indicateurs	Valeur	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Taux de pauvreté après prestations sociales (en pourcentage)						
Actifs occupés	6,5	0,3	4,8	1,05	10 190	9 671
Chômeurs	32,6	1,7	5,3	1,03	1 035	1 006
Retraités	11,1	0,5	4,3	1,13	4 801	4 252
Autres inactifs	26,0	1,1	4,0	1,05	3 161	3 015
Hommes, actifs occupés	6,8	0,4	5,3	1,09	5 349	4 893
Hommes, chômeurs	36,9	2,1	5,7	1,02	493	485
Hommes, retraités	10,4	0,6	5,5	1,13	2 391	2 123
Hommes, autres inactifs	26,0	1,4	5,3	1,06	896	847
Femmes, actives occupées	6,1	0,4	6,0	1,09	4 841	4 454
Femmes, chômeuses	28,3	2,0	7,2	1,09	542	497
Femmes, retraitées	11,6	0,6	5,0	1,11	2 410	2 171
Femmes, autres inactives	26,0	1,1	4,3	1,02	2 265	2 217
Personnes seules, < 65 ans	16,9	0,7	4,0	1,23	1 641	1 336
Personnes seules, 65+ ans	18,0	1,0	5,5	1,09	1 234	1 133
Personnes seules, hommes	16,9	0,8	4,8	0,99	1 093	1 100
Personnes seules, femmes	17,7	0,8	4,4	0,97	1 782	1 836
Personnes seules, total	17,4	0,6	3,4	1,00	2 875	2 886
2 adultes, sans enfant, tous les deux < 65	7,7	0,5	6,7	1,01	3 642	3 597
2 adultes, sans enfant, au moins 1 65+	9,2	0,8	9,2	0,94	2 766	2 942
Autres ménages sans enfant	9,8	1,2	12,6	1,07	1 530	1 427
Familles monoparentales	27,2	1,9	7,1	1,04	1 531	1 474
2 adultes, 1 enfant	8,0	0,9	10,8	1,12	3 099	2 760
2 adultes, 2 enfants	10,3	0,9	9,1	1,24	5 160	4 168
2 adultes, 3+ enfants	17,8	2,1	11,7	1,15	3 645	3 164
Autres ménages avec enfants	22,9	2,9	12,5	1,16	1 590	1 368
Ménages sans enfant	11,4	0,4	3,2	1,10	10 838	9 870
Ménages avec enfants	14,6	0,6	4,4	1,09	15 069	13 835
Propriétaires	9,0	0,4	4,0	1,02	17 880	17 591
Locataires	21,1	0,8	3,9	1,04	8 027	7 718
Ménages sans enfant, w = 0	18,7	1,0	5,4	1,04	1 830	1 753
Ménages sans enfant, 0 < w < 1	9,4	0,7	7,3	1,10	2 618	2 390
Ménages sans enfant, w = 1	5,0	0,4	8,5	1,26	2 148	1 708
Ménages avec enfants, w = 0	70,9	3,5	5,0	0,93	842	902
Ménages avec enfants, 0 < w < 0,5	43,8	3,3	7,6	0,97	851	878
Ménages avec enfants, 0,5 < w < 1	15,8	1,2	7,5	0,96	4 959	5 181
Ménages avec enfants, w = 1	5,4	0,5	9,5	1,11	8 406	7 540
Médiane des niveaux de vie (en euros par unité de consommation)	16 567	53,9	0,3	1,13	25 907	23 015
Seuil de pauvreté - personnes seules (en euros)	9 940	32,1	0,3	1,16	25 907	22 283
Seuil de pauvreté - 2 adultes, 2 enfants (en euros)	20 874	50,5	0,2	1,00	25 907	26 005

Indicateurs	Valeur	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Inégalité de la distribution des revenus : rapport Q80/Q20	3,8	0,1	1,6	1,03	25 907	25 225
Écart de risque de pauvreté (en pourcentage)						
Total	16,8	0,7	4,0	1,04	3 299	3 172
Hommes total	17,2	0,7	4,3	1,04	1 526	1 473
Femmes total	16,5	0,7	4,4	1,03	1 773	1 720
0-17 ans	14,8	0,9	6,1	1,00	968	966
18-64 ans	17,4	0,7	4,2	1,05	1 838	1 757
65+ ans	18,9	1,2	6,4	1,26	493	390
18+ ans	17,8	0,7	3,7	1,08	2 331	2 162
Hommes, 18-64 ans	18,1	0,9	5,2	1,08	830	772
Hommes, 65+ ans	19,0	1,3	7,0	1,24	194	156
Hommes, 18+ ans	18,6	0,8	4,3	1,11	1 024	924
Femmes, 18-64 ans	16,9	0,7	4,4	1,06	1 008	952
Femmes, 65+ ans	18,9	1,5	8,1	1,33	299	225
Femmes, 18+ ans	17,2	0,7	3,9	1,08	1 307	1 207
Taux de pauvreté selon le seuil de pauvreté (en pourcentage)						
40 % de la médiane	3,0	0,2	6,2	0,93	25 907	27 763
50 % de la médiane	6,6	0,3	4,1	1,04	25 907	24 817
70 % de la médiane	21,0	0,4	1,9	1,06	25 907	24 349
Taux de pauvreté avant prestations sociales autres que retraites (en pourcentage)						
Total	25,8	0,4	1,5	1,02	25 907	25 369
Homes total	25,0	0,4	1,7	1,02	12 533	12 322
Femmes total	26,6	0,4	1,6	1,05	13 374	12 781
0-17 ans	35,5	0,9	2,4	1,00	6 314	6 325
18-64 ans	24,5	0,4	1,5	1,05	15 705	14 900
65+ ans	18,0	0,6	3,4	1,26	3 888	3 074
18+ ans	23,1	0,3	1,4	1,06	19 593	18 572
Hommes, 18-64 ans	23,3	0,4	1,8	1,08	7 588	7 058
Hommes, 65+ ans	16,6	0,7	4,5	1,16	1 732	1 493
Hommes, 18+ ans	22,0	0,4	1,7	1,07	9 320	8 734
Femmes, 18-64 ans	25,6	0,4	1,6	1,07	8 117	7 617
Femmes, 65+ ans	19,1	0,7	3,7	1,27	2 156	1 702
Femmes, 18+ ans	24,1	0,4	1,5	1,09	10 273	9 422
Coefficient de Gini (en pourcentage)	26,4	0,3	1,3	1,00	25 907	26 011
Niveau de vie moyen (en euros par unité de consommation)	18 488				25 907	

Composantes du revenu disponible	Moyenne par ménage (en euros)	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Revenu disponible total (HY020)	28 797	162	0,6	1,09	10 498	9 667
Revenus tirés de la location d'une propriété ou de terres (HY040N)	508	34	6,7	1,00	10 498	10 529
Allocation famille/enfant (HY050N)	886	13	1,5	1,03	10 498	10 176
Exclusion sociale non classée ailleurs (HY060N)	165	10	6,0	1,01	10 498	10 415
Aides au logement (HY070N)	502	10	2,0	1,10	10 498	9 524
Transferts interménages perçus régulièrement en espèces (HY080N)	193	11	5,5	1,05	10 498	9 965
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	820	17	2,1	1,21	10 498	8 650
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	0	11,0	1,01	10 498	10 390
Impôts réguliers sur la fortune (HY120N)	70	18	25,5	1,00	10 498	10 458
Transferts interménages versés régulièrement en espèces (HY130N)	289	14	4,9	1,24	10 498	8 498
Remboursements consécutifs à des ajustements fiscaux (HY145N)	2 108	63	3,0	0,94	10 498	11 138
Salaire ou revenu assimilé (PY010N)	17 375	183	1,1	1,05	10 498	10 021
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	1 899	102	5,3	1,04	10 498	10 096
Allocations chômage(PY090N)	889	38	4,3	1,10	10 498	9 544
Retraites (PY100N)	7 188	75	1,0	1,04	10 498	10 088
Allocations au survivant (PY110N)	157	13	8,3	1,18	10 498	8 909
Indemnités de maladie (PY120N)	224	16	7,2	0,97	10 498	10 846
Pensions d'invalidité (PY130N)	404	23	5,7	1,11	10 498	9 443
Allocations d'éducation (PY140N)	49	4	8,2	1,41	10 498	7 425

2.3 Erreurs non dues à l'échantillonnage

2.3.1 Erreurs dans la base de sondage et la couverture

Deux bases de sondages ont été utilisées conjointement :

- Le recensement général de la population de mars 1999, qui est reconnu comme étant de très bonne qualité, même s'il y a quelques omissions de logements (la qualité du recensement est certes moins bonne pour les personnes physiques - mais la base utilisée ici est une base de logements). L'ancienneté du recensement explique qu'un certain nombre de logements de la base sont hors champ et que nous sommes obligés de prendre dans l'échantillon des résidences vacantes ou secondaires au moment du recensement alors que notre champ se limite aux résidences principales.
- Une base qui a mis à jour la précédente fin 2005, dite « base de sondage de logements neufs » (BSLN). Pour la constituer, on part de l'ensemble des logements appartenant aux permis de construire délivrés par les mairies. Ces permis sont gérés dans un fichier administratif appelé SITADEL, qui comprend environ 300 000 logements chaque année. Un extrait de SITADEL est obtenu par sondage, ce qui donne lieu à un échantillon de logements initialement « fictifs » dont la construction est suivie sur le terrain par des enquêteurs jusqu'à ce qu'il y ait achèvement du logement. À la constatation de l'achèvement, le logement entre dans la BSLN. La qualité de SITADEL et du processus de suivi sur le terrain n'est pas

finement quantifiable, mais on considère que la BSLN est tout à fait satisfaisante en matière de couverture de la construction neuve.

2.3.2 Erreurs de mesure et de traitements

Description des différentes causes d'erreurs de mesure susceptibles d'être détectées dans l'enquête

L'introduction des pièces et billets en euros en France étant assez récente (janvier 2002), de nombreux ménages continuent à « répondre en francs », même si l'enquêteur demande une réponse en euros. Un certain nombre de personnes âgées raisonnent toujours en anciens francs¹. Une calculatrice intégrée permet la conversion par l'enquêteur de montants exceptionnellement fournis en francs et les cartes code et le fascicule revenus donnent les montants en euros et en francs.

Par ailleurs, la multiplicité des périodes de référence dans l'enquête (période courante par exemple pour les conditions de logement, année N-1 pour les revenus, année N-2 - dans la stratégie française - pour les impôts) peut également être source de confusion. Cependant, toutes les questions font référence à la période d'intérêt dans leurs libellés mêmes et les questions sont depuis 2005 regroupées par période de référence.

En plus du problème de la période de référence, un certain nombre de confusions ont été constatées entre mois et année. Ce type de problème est courant s'agissant d'enquêtes sur le thème des revenus. Des erreurs correspondant à un facteur de 12 peuvent ainsi être trouvées sur certains montants. Pour d'autres variables, il y a pu y avoir des confusions entre trimestre et année. Ceci concerne en particulier les personnes âgées et les retraites, certaines d'entre elles étant versées trimestriellement.

Un autre type d'erreur fréquent dans ce type de collecte concerne le nombre de 0 associé à un montant. Les montants mentionnés correspondent en général à des chiffres élevés, parfois des centaines de milliers. Dans le cadre d'une collecte par questionnaire CAPI, c'est l'enquêteur qui saisit le chiffre, en même temps qu'il pose les questions. Il est donc possible que parfois un zéro en trop ou en moins ait pu être saisi dans un montant. Des séparateurs de milliers ont été ajoutés en 2005 afin de limiter ces erreurs.

Les ménages sont invités, dès l'envoi de la « lettre avis », à préparer pour le jour de l'entretien les documents nécessaires pour la bonne conduite de celui-ci, en particulier la déclaration de revenus. Quatre individus sur cinq (parmi ceux qui répondent) utilisent cette déclaration de revenus. Dans le cas contraire, on demande aux salariés d'utiliser les récapitulatifs annuels de salaires que leur employeur (unique la plupart du temps). Si l'enquêté n'utilise toujours pas ce dernier document, on lui demande tout simplement son salaire mensuel, éventuellement en tranches s'il refuse de le donner en clair. Enfin, une fraction des ménages ne souhaite répondre à aucune question relative aux revenus.

L'ensemble de ces facteurs (confusion dans les devises et dans les périodes de référence, multiplicité des sources) conduit à une certaine hétérogénéité des résultats de collecte qui nuit à la qualité des données.

Description de la façon dont le questionnaire a été élaboré, utilisation éventuelle d'un laboratoire cognitif, test sur le terrain du questionnaire, incidence de sa conception, de son contenu et de sa formulation.

Structure du questionnaire

L'enquête est composée d'un questionnaire Ménage (destiné à l'ensemble du ménage) et d'un questionnaire individuel posé à toutes les personnes du ménage âgées de 16 ans ou plus (au 1^{er} janvier de l'année d'enquête). Le questionnaire ménage est précédé du Tronc commun des ménages, qui constitue le socle de l'ensemble des enquêtes auprès des ménages conduites par l'Insee.

Un certain nombre de composantes du revenu des ménages est perçu au niveau individuel. Il s'agit :

- des revenus d'activité ou de remplacement (salaires, allocations de chômage ou de préretraite, pensions, retraites ou rentes...),

¹ Le 1^{er} janvier 1960, le nouveau franc a été introduit, 1 nouveau franc correspondant à 100 anciens francs.

- de certaines prestations et minima sociaux (prestations maladie, invalidité ou dépendance, bourses d'études, revenu minimum d'insertion, minimum vieillesse, etc).

En revanche, certaines composantes du revenu des ménages sont difficiles, voire impossibles, à individualiser. Il s'agit :

- des aides au logement et du loyer imputé, des prestations familiales, de l'aide sociale ;
- des revenus de la propriété, des transferts entre ménages, des revenus des enfants, des impôts et des remboursements d'intérêts hypothécaires,

Ces composantes sont donc en général collectées au niveau du ménage, Par exception, les prestations familiales et l'aide sociale, attribuées à un allocataire, sont collectées auprès de lui au niveau du questionnaire individuel.

L'architecture du questionnaire est la suivante :

- Tronc commun des enquêtes ménages (TCM) :
 - Identification du logement, contact ;
 - Tableau des habitants du logement et contour des unités de vie (THL),
- Questionnaire ménage (renseigné par un adulte quelconque du ménage « unité de vie ») :
 - Ressources et charges en période courante ;
 - Changements récents et jeunes enfants ;
 - Conditions de logement - résidence principale ;
 - Revenus non individualisables sur la période de référence ;
 - Endettement et confort financier,
- Questionnaire individuel (renseigné par chaque adulte du ménage, le recours à un proxy étant toléré) :
 - Biographie et ressources culturelles ;
 - Activité, emploi, profession ;
 - Revenus individualisables sur la période de référence ;
 - Santé ;
 - Module secondaire (participation sociale).

Pour limiter les erreurs de collecte, l'équipe de conception a d'une part introduit dans CAPI des filtres et des contrôles, d'autre part autorisé une collecte en francs pour de nombreux montants, en particulier lorsque les documents sont anciens (tableau d'amortissement pour la collecte des intérêts d'emprunt, prix de vente du logement).

Exhaustivité et absence de doubles comptes dans la collecte des revenus

Arguant du fait que l'enquêteur n'est pas capable a priori de connaître toutes les sources de revenus du ménage et que ce dernier peut oublier d'en mentionner une, on met en œuvre une stratégie générale de « balayage » des revenus. Cette stratégie consiste à :

- recenser les différents types de revenus perçus par le ménage au cours de l'année de référence (année civile 2006), avant de collecter les montants correspondants,
- vérifier que le montant n'a pas été inclus antérieurement dans un autre revenu pour éviter les double-comptes,
- essayer systématiquement, lorsque l'enquête n'a pas pu ou voulu renseigner un montant, d'obtenir une information en tranches.

Utilisation des documents administratifs et fiscaux

Le questionnement est fonction des documents utilisés. Parmi ces documents, la ou les déclarations des revenus de l'année de référence 2006 (pré-remplie par l'administration fiscale, reçue en mai 2006) sont privilégiées pour les revenus d'activité et de remplacement des adultes, qui sont en quasi-totalité imposables. La déclaration de revenus présente l'avantage d'un mode de recueil des revenus annuels à la fois plus facile et plus fiable. D'une part, les récapitulatifs annuels sont déjà faits. D'autre part, le concept mesuré (revenu déclaré) permet une étude de qualité par comparaison de la distribution des revenus avec celle de l'enquête Revenus fiscaux (enquête de référence à l'Insee pour la distribution des revenus).

Dans le questionnaire individuel, l'enquêté indique combien de déclarations de revenus 2006 il a effectué, et s'il consent à s'y reporter. Pour les revenus d'activité et de remplacement, le questionnement prend deux formes différentes selon la réponse :

- une boucle (dans CAPI) indexée sur le nombre de déclarations de revenus en cas de consentement,
- sinon, une boucle indexée sur le nombre de sources de revenus (nombre d'employeurs pour les salaires, d'entreprises pour les revenus d'indépendants, de caisses ou organismes de versement pour les retraites ou pensions...).

Dans le premier cas, on obtient un montant annuel de revenu déclaré, qui doit être complété par une interrogation synthétique sur l'existence et le montant de revenus complémentaires non déclarés.

Dans le second cas, on reconstitue un montant annuel de revenu perçu cumulé par source de manière plus analytique : durée de perception, régularité du montant mensuel, et montant mensuel en cas de versements réguliers ou montant annuel en cas de versements irréguliers.

Codification de la profession

L'intégration du logiciel de codification SICORE dans le questionnaire CAPI permet de reconnaître les libellés de profession : SICORE (Système Informatique de COdage des Réponses aux Enquêtes) est un système de chiffrement automatique développé par l'Insee. Les enquêteurs sont avertis lorsque le libellé renseigné est inconnu de la base des professions. Ils peuvent alors saisir d'autres libellés qui sont tous enregistrés. Ce ci permet de corriger les erreurs de saisie et de préciser les libellés.

Questionnement longitudinal

Le questionnement longitudinal a commencé en 2005. Les personnes réinterrogées ont répondu à un questionnaire un peu différent de celui des entrants si elles avaient accepté lors de la dernière vague que les informations qu'elles avaient fournies leur soient restituées l'année suivante.

Dans ce cas, le questionnement tient compte des réponses données l'année précédente dans la formulation des questions, voire en ne reposant pas certaines questions, ce qui allège l'interview.

Des contrôles de cohérence peuvent être effectués d'une vague à l'autre, ce qui réduit le bruit des données. Ils portent :

- sur les dates de certains événements (calendrier d'activité par exemple),
- sur la non-perception d'un type de revenu perçu l'année précédente,
- sur l'évolution de montants détaillés de composantes du revenu ou du coût du logement (fourchette entre -20 % et + 30 %),
- sur l'évolution de montants agrégés de revenus (impôts, salaires, retraites).

Tests du questionnaire

Le questionnaire transversal de 2007 a bénéficié des bilans des collectes 2004 à 2006. Le bilan de la collecte 2006 a été effectué grâce aux questionnaires papiers remplis par l'ensemble des enquêteurs, des synthèses rédigées par les directions régionales et des comptes-rendus des réunions de bilans organisées en Auvergne, Centre et Rhône-Alpes.

Le questionnaire de 2007 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2006 dans les directions régionales de Franche-Comté et de PACA. Il a été mené auprès de 116 ménages

(répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2006 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2007. Ce type de test permet de se placer dans des configurations particulières qui n'ont pas été rencontrées lors des tests sur le terrain.

À l'issue de chaque test et suite aux bilans de collecte et aux différentes remarques des enquêteurs, des modifications du questionnaire (structure, formulation) sont proposées au maître d'ouvrage par l'équipe de conception. Ces modifications du questionnaire sont validées à l'occasion d'un comité de pilotage du projet.

Les tests servent aussi à mesurer les durées de collecte qui permettent de calculer la rémunération des enquêteurs pour l'enquête. L'équipe de conception fournit aux enquêteurs une feuille permettant le relevé manuel des temps d'interview, ces relevés venant compléter les temps mesurés dans CAPI.

Informations sur l'intensité et l'efficacité de la formation des enquêteurs: nombre de jours de formation, test des compétences avant d'entreprendre le travail sur le terrain (taux de succès, etc).

L'Insee dispose d'un réseau d'enquêteurs stable. L'Institut a recours dans la mesure du possible aux mêmes enquêteurs d'une vague sur l'autre, ce qui présente au moins deux avantages : les enquêteurs connaissent bien l'enquête, et les enquêtés sont plus facilement fidélisés par un enquêteur qu'ils connaissent déjà.

Formation des enquêteurs

Pour les tests, l'équipe de conception assure directement la formation des enquêteurs.

En revanche, pour la collecte en grandeur réelle (mai 2007), l'équipe de conception ne forme pas directement les enquêteurs mais les gestionnaires responsables de l'enquête dans les directions régionales. Les gestionnaires des directions régionales forment ensuite les enquêteurs SILC. Cependant, la formation des enquêteurs reproduit à l'identique la formation des gestionnaires. L'équipe de conception SILC fournit aux gestionnaires des directions régionales (en plus des documents de collecte) les transparents et les exercices CAPI. La formation dispensée aux gestionnaires et aux enquêteurs a duré trois jours pour ceux qui ne connaissaient pas encore l'enquête et 1,5 jours pour ceux qui avaient déjà participé à la collecte précédente.

Lors de la collecte, chaque enquêteur est accompagné au moins une fois par un agent de l'Insee (gestionnaire de l'enquête, concepteur...).

Documents de collecte

En vue de la collecte, l'équipe de conception prépare les documents de collecte fournis aux enquêteurs. Ceux-ci sont constitués :

- d'un argumentaire (utilisé pour convaincre les ménages récalcitrants),
- d'une lettre avis envoyée aux enquêtés pour les prévenir de la venue d'un enquêteur,
- d'une plaquette de présentation du dispositif envoyée avec la lettre avis
- d'une instruction aux enquêteurs,
- d'un cahier des cartes (les cartes sont utilisées dans le cas où l'énumération des modalités d'une question par l'enquêteur serait trop longue ou fastidieuse),
- d'un fascicule « revenus », compléments d'information pour les enquêteurs sur certaines aides et prestations existantes et sur les conditions à remplir pour en bénéficier, réalisé afin d'aider les enquêteurs à retrouver un type de revenus lorsque le ménage ne sait pas exactement ce qu'il perçoit,
- d'une architecture du questionnaire,
- de fiches de prise de contact et de suivi. Les fiches de suivi comportent quelques informations (numéros de téléphone fixe ou portable, adresse internet, personne-relais) pour anticiper les déménagements ;

- d'un document présentant les résultats de l'enquête de l'année précédente, donné à tous les ménages répondant.

2.3.3 Erreurs de non-réponse

2.3.3.1 Taille de l'échantillon obtenu

Les deux tableaux ci-dessous fournissent le nombre de ménages, puis le nombre d'individus correspondants, pour lesquels l'entretien a été accepté suite à la collecte 2007. Ces nombres sont répartis selon les 9 sous-échantillons, la France ayant opté pour un échantillon rotatif par neuvième (9 interrogations pour un individu panel). La variable « Nombre d'années dans le panel » correspond au nombre d'interrogations annuelles, pour le sous-échantillon concerné, restant à mener juste avant la collecte 2007. Ainsi le sous-échantillon pour lequel le « Nombre d'années dans le panel » vaut 1 correspond aux ménages qui ont été interrogés pour la dernière fois en 2007. L'échantillon pour lequel cette variable vaut 9 correspond à l'échantillon entrant en 2007.

Nombre de ménages pour lesquels un entretien est accepté pour la base de données	
Ventilation par groupe de rotation	
Nombre d'années dans le panel	Nombre de ménages
1	927
2	938
3	887
4	918
5	876
6	903
7	1 557
8	1 668
9	1 824
Total	10 498

Nombre de personnes âgées de 16 ans et plus, membres des ménages dont l'entretien est accepté pour la base de données, pour lesquels un entretien personnel est achevé,	
Ventilation par groupe de rotation	
Nombre d'années dans le panel	Nombre de personnes
1	1 795
2	1 818
3	1 717
4	1 765
5	1 719
6	1 752
7	3 044
8	3 181
9	3 452
Total	20 243

2.3.3.2 Non-réponse des unités

Le taux de non-réponse des ménages est donné de manière détaillée dans le tableau ci-dessous. On obtient un taux de non-réponse des ménages de 14,9 %. Il est de 30,9 % pour les entrants et de 10,7 % pour les réinterrogés.

Adresses contactées avec succès DB120=11	12 268
Total DB120	12 735
Adresse inexistante DB120=23	399
Adresses validées sélectionnées	12 336
Taux de contact à l'adresse Ra	99,4%
Entretiens de ménages achevés et acceptés pour la base de données DB135=1	10 498
Ménages éligibles aux adresses contactées	12 268
Total DB130	
Proportion d'entretiens de ménages achevés et acceptés pour la base de données Rh	85,6%
Taux de non-réponse des ménages NRh	14,9%

Le taux de non-réponse individuelle, ainsi que le taux global, et le détail des calculs sont donnés ci-dessous.

Nombre entretiens individuels achevés RB250 = 11	20 243
Nombre d'individus éligibles dans les ménages dont les entretiens ont été achevés et acceptés pour la base de données RB245 = 1	20 357
Rp	99,4%
Taux de non-réponse individuelle NRp	0,6%
Taux global de non-réponse individuelle *NRp	15,4%

On obtient ainsi un taux global de non-réponse individuelle de 15,4% (31,5% pour les entrants et 11,3 % pour les réinterrogés), très proche du taux de non-réponse des ménages à 14,9 %.

2.3.3.3 Ventilation des ménages selon l'« enregistrement du contact à l'adresse », le « résultat du questionnaire ménage », et l'« acceptation de l'entretien » pour chaque groupe de rotation et le total

Ventilation des ménages selon l' « enregistrement du contact à l'adresse » (DB120)

DB075	Enregistrement du contact à l'adresse			Total :
Nombre d'années dans le panel	11 : Adresse contactée	21 : Impossible de localiser l'adresse	23 : Adresse inexistante	
1	1 018 99,3%	7 0,7%	0 0,0%	1 025
2	1 038 99,6%	4 0,4%	0 0,0%	1 042
3	988 99,6%	4 0,4%	0 0,0%	9 92
4	1 026 99,8%	2 0,2%	0 0,0%	1 028
5	963 99,9%	1 0,1%	0 0,0%	964
6	1 004 99,9%	1 0,1%	0 0,0%	1 005
7	1 761 99,4%	11 0,6%	0 0,0%	1 772
8	1 879 99,7%	5 0,3%	0 0,0%	1 884
9	2 591 85,7%	33 1,1%	399 13,2%	3 023
Total	12268 96,4%	68 0,5%	399 3,1%	12 735

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130)

DB130		Résultat du questionnaire "ménage"				Total
Nombre d'années dans le panel	11 : Questionnaire rempli	21 : Refus de coopérer	22 : Ensemble du ménage absent	23 : Ménage incapable de répondre	24 : Autres raisons	
1	929 91,3%	37 3,6%	17 1,7%	14 1,4%	21 2,1%	1 018
2	939 90,5%	28 2,7%	30 2,9%	17 1,6%	24 2,3%	1 038
3	887 89,8%	35 3,5%	26 2,6%	16 1,6%	24 2,4%	988
4	921 89,8%	38 3,7%	17 1,7%	22 2,1%	28 2,7%	1 026
5	877 91,1%	39 4,1%	16 1,7%	13 1,4%	18 1,9%	963
6	906 90,2%	39 3,9%	28 2,8%	14 1,4%	17 1,7%	1 004
7	1 557 88,4%	70 4,0%	45 2,6%	42 2,4%	47 2,7%	1 761
8	1 669 88,8%	90 4,8%	39 2,1%	32 1,7%	49 2,6%	1 879
9	1 832 70,7%	358 13,8%	217 8,4%	120 4,6%	64 2,5%	2 591
Total	10 517 85,7%	734 6,0%	435 3,6%	290 2,4%	292 2,4%	12 268

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) :

DB135	Acceptation de l'entretien		Total
Nombre d'années dans le panel	1 : Entretien accepté pour la base de données	2 : Entretien rejeté pour la base de données	
1	927 99,8%	2 0,2%	929
2	938 99,9%	1 0,1%	939
3	887 100,0%	0 0,0%	887
4	918 99,7%	3 0,3%	921
5	876 99,9%	1 0,1%	877
6	903 99,7%	3 0,3%	906
7	1 557 100,0%	0 0,0%	1 557
8	1 668 99,9%	1 0,1%	1 669
9	1 824 99,6%	8 0,4%	1 832
Total	10 498 99,8%	19 0,2%	10 517

2.3.3.4 Ventilation des unités remplacées

Sans objet dans le cas de la France.

2.3.3.5 Non-réponse à certaines questions

Les tableaux suivants fournissent par composante du revenu net disponible la part d'unités percevant cette composante, le pourcentage de valeur manquante et le taux d'information partielle. Ce dernier s'élève à 52,9 % pour le revenu disponible total. Les prestations sociales hors retraites sont les variables qui sont les plus sous-déclarées ou mal renseignées par les ménages.

Ventilation des non-réponses à certaines questions

	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Ménages</i>			
Revenu disponible total (HY020)	100,0%	1,8%	52,9%
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et des pensions au survivant (HY022)	99,2%	2,3%	37,0%
Revenu disponible des ménages avant transferts sociaux (HY023)	97,0%	4,7%	22,9%
Revenus tirés de la location d'une propriété ou de terres (HY040N)	8,4%	-	-
Allocation famille/enfant (HY050N)	28,2%	5,4%	10,7%
Exclusion sociale non classée ailleurs (HY060N)	4,8%	1,4%	0,5%
Aides au logement (HY070N)	23,1%	5,7%	2,4%
Transferts interménages perçus régulièrement en espèces (HY080N)	5,8%	-	-
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	77,8%	-	-
Revenus perçus par les personnes de moins de 16 ans (HY110N)	2,0%	-	-
Impôts réguliers sur la fortune (HY120N)	1,6%	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	8,9%	-	-
Remboursements consécutifs à des ajustements fiscaux (HY145N)	87,8%	5,9%	15,1%

	% de personnes de plus de 16 ans...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Individus</i>			
Salaire ou revenu assimilé (PY010N)	53,7%	3,9%	0,7%
Avantages en nature (PY020N)	0,8%	-	-
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,3%	-	0,3%
Allocations chômage(PY090N)	7,4%	-	0,7%
Retraites (PY100N)	25,0%	-	7,7%
Allocations au survivant (PY110N)	1,0%	-	0,1%
Indemnités de maladie (PY120N)	4,2%	-	0,0%
Pensions d'invalidité (PY130N)	3,0%	-	0,1%
Allocations d'éducation (PY140N)	2,1%	-	-

2.3.3.6 Nombre total des non-réponses

La non-réponse se décompose en non-réponse partielle (non-réponse à une question) et en non-réponse totale. La non-réponse au niveau ménage (cf. partie 2.3.3.2) concerne 1 838 ménages, soit 14,9 % des ménages.

Taille d'échantillon et non-réponse totale par type d'indicateur

	Taille d'échantillon obtenu (nb de personnes prises en compte dans le calcul)	Non réponse partielle	Hors champ	Non-réponse totale des individus	Taux de non réponse individuelle
Taux de pauvreté après transferts selon l'âge et le sexe	25 907	0	0	0	0,0%
Taux de pauvreté après transferts selon le statut professionnel	19 187	0	6 606 ¹	114 ²	0,6%
Taux de pauvreté après transferts selon le type de ménage	25 907	0	0	0 ³	0,0%
Taux de pauvreté après transferts selon le statut d'occupation du logement	25 907	0	0	0	0,0%
Taux de pauvreté après transferts selon l'intensité d'emploi	22 654	0	3 234 ⁴	19 ⁵	0,1%
Médiane du revenu disponible par unité de consommation - après transferts sociaux	25 907	0	0	0	0,0%
Seuil de pauvreté - après transferts sociaux - personnes seules	25 907	0	0	0	0,0%
Seuil de pauvreté - après transferts sociaux - 2 adultes, 2 enfants	25 907	0	0	0	0,0%
Inégalité de la distribution de revenu - Rapport du partage de revenus S80/S20	25 907	0	0	0	0,0%
Écart moyen relatif du taux de pauvreté	3 299	0	22 608 ⁶	0	0,0%
Dispersion autour du seuil de pauvreté	25 907	0	0	0	0,0%
Taux de pauvreté avant transferts sociaux autres que retraites	25 907	0	0	0	0,0%
Taux de pauvreté avant tout transfert social	25 907	0	0	0	0,0%
Coefficient de Gini	25 907	0	0	0	0,0%
Revenu disponible par uc moyen	25 907	0	0	0	0,0%

¹ dont 5 550 enfants de moins de 16 ans, 757 entre 16 et 18 ans, et 299 individus qui n'ont pas gardé le même statut professionnel plus de la moitié de l'année.

² 114 individus n'ont pas rempli de questionnaire individuel, leur statut d'occupation est donc inconnu.

³ Nous n'avons pu déterminer le type de ménage pour 17 individus à cause de la non-réponse individuelle de 19 personnes âgées de 16 à 24 ans.

⁴ dont 68 personnes dans un ménage d'étudiants, 3 166 dans un ménage avec seulement des enfants dépendants ou des personnes âgées de plus de 65 ans.

⁵ 19 individus appartenant à un ménage répondant dont un individu n'a pas rempli de questionnaire individuel.

⁶ 22 608 non-pauvres.

Nous n'avons aucun cas d'indicateurs reposant sur moins de 50 observations (cf. partie 1).

2.4 Mode de collecte des données

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,4 %. En effet, les ménages répondants ont tous rempli au moins un questionnaire individuel et les proxys sont autorisés.

Ventilation des membres de ménages répondants âgés du plus de 16 ans selon le statut des données

RB250		Statut des données		
Nombre d'années dans le panel		11 Information collectée par enquête	14 Pas de contact	Total
1	Nombre Pourcentage	1 795 99,2%	14 0,8%	1 809
2	Nombre Pourcentage	1 818 99,5%	10 0,5%	1 828
3	Nombre Pourcentage	1 717 99,4%	10 0,6%	1 727
4	Nombre Pourcentage	1 765 99,7%	6 0,3%	1 771
5	Nombre Pourcentage	1 719 99,6%	7 0,4%	1 726
6	Nombre Pourcentage	1 752 99,5%	9 0,5%	1 761
7	Nombre Pourcentage	3 044 99,5%	14 0,5%	3 058
8	Nombre Pourcentage	3 181 99,5%	16 0,5%	3 197
9	Nombre Pourcentage	3 452 99,2%	28 0,6%	3 480
Total	Nombre Pourcentage	20 243 99,4%	114 0,6%	20 357

Ventilation des membres de ménages répondants âgés de plus de 16 ans selon le type d'entretien

Seuls les membres du ménage ont le droit de répondre à la place d'un individu. Le tableau suivant détaille le taux de recours au proxy, qui s'élève globalement à 27,9%.

RB260		Type d'entretien	
Nombre d'années dans le panel		2 Face à face CAPI	5 Proxy
1	Nombre Pourcentage	1 320 73,5%	475 26,5%
2	Nombre Pourcentage	1 317 72,4%	501 27,6%
3	Nombre Pourcentage	1 227 71,5%	490 28,5%
4	Nombre Pourcentage	1 260 71,4%	505 28,6%
5	Nombre Pourcentage	1 225 71,3%	494 28,7%
6	Nombre Pourcentage	1 225 69,9%	527 30,1%
7	Nombre Pourcentage	2 152 70,7%	892 29,3%
8	Nombre Pourcentage	2 300 72,3%	881 27,7%
9	Nombre Pourcentage	2 565 74,3%	887 25,7%
Total	Nombre Pourcentage	14 591 72,1%	5 652 27,9%

2.5 Durée de l'entretien

La durée moyenne de l'entretien pour la collecte 2007 est de 53 minutes (58 minutes pour les entrants et 52 pour les réinterrogés).

2.6 Les imputations

L'imputation est d'abord nécessaire parce qu'il existe des données manquantes ou en tranches. Par ailleurs des données sont erronées et un estimateur de sondage suppose que l'on dispose des mesures sur tout l'échantillon tiré au hasard et que les valeurs sont les vraies valeurs. Ainsi il faut corriger ces erreurs qui peuvent engendrer des biais (par exemple si l'erreur est toujours dans le même sens : francs à la place d'euros). Cela peut se faire sur dire d'expert et en cas de doute par imputation (simulation dans une loi où les paramètres sont des paramètres estimés éventuellement conditionnellement à une information auxiliaire de type tranche). La méthode générale de l'imputation est présentée en annexe.

Avant toute chose, il convient de présenter certaines difficultés rencontrées. Les revenus dans l'EU-SILC 2007 sont relatifs à plusieurs dates. L'impôt payé a pour assiette les revenus imposables perçus au cours de l'année 2005. Les revenus collectés sont relatifs à l'année 2006. Enfin certains revenus concernent la période courante de l'enquête, soit mai 2007, ou bien une moyenne estimée sur une période relativement récente. Nous devons faire l'hypothèse d'une certaine continuité, pour pouvoir comparer ces différents revenus et les contrôler.

Certains revenus tirés de la collecte 2007 ont été modifiés. En effet, des comparaisons faites avec des fichiers externes tels l'enquête Revenus fiscaux (ERF), l'enquête Logement, ou les statistiques de la Caisse nationale des allocations familiales (Cnaf) font apparaître des écarts. De nombreuses raisons expliquent ces écarts. Nous avons décelé des erreurs francs/euros, des revenus déclarés comme étant annuels alors qu'ils sont mensuels, trimestriels, voire quadrimestriels. Pour les seules retraites,

une soixantaine de cas ont été trouvés, expliquant pourquoi le montant est erroné, Nous décrivons ici de façon succincte les imputations réalisées afin de compenser ces erreurs de collecte.

Les travaux ont porté sur plusieurs types de revenus. Une attention particulière a été apportée aux composantes principales du revenu disponible des ménages. Nous n'avons pas été en mesure de contrôler des revenus secondaires tels les bourses pour les élèves ou les revenus fonciers, faute d'informations internes et externes suffisamment fiables.

Nous décrivons maintenant les opérations concernant les revenus principaux.

Les deux méthodes retenues pour les imputations

L'imputation des revenus individuels est menée de deux façons différentes, selon que le ménage est enquêté pour la première fois ou non. Dans le premier cas, l'imputation est transversale : une équation du revenu est estimée sur les répondants et permet d'imputer le revenu des non-répondants. Dans le second cas, nous souhaitons utiliser le revenu donné par l'individu à une des deux dates pour estimer le revenu manquant perçu à l'autre date. Pour ce faire, nous estimons une équation du ratio entre les revenus des deux années sur les répondants aux deux vagues. Ce ratio est ensuite estimé pour les individus n'ayant répondu qu'à une enquête afin d'attribuer le revenu manquant. Cette méthode est appliquée pour les imputations des salaires et des retraites.

Salaire ou revenu assimilé (PY010N)

Il est nécessaire dans un premier temps de définir sur quelles données l'imputation va porter. Nous avons d'abord comparé le salaire déclaré avec les maxima observés dans l'ERF en tenant compte du sexe du salarié et de sa catégorie socioprofessionnelle (sur une position). Dans ce cas, la valeur collectée est remplacée par la valeur imputée. Nous attribuons aussi un salaire aux individus déclarant en percevoir et qui n'ont donné aucun renseignement sur le montant, ainsi qu'aux individus ayant répondu par un montant en tranches.

L'imputation est menée par strates. Huit strates sont créées à partir du sexe, de l'emploi qualifié ou non, et du secteur d'emploi, privé ou public. Nous sélectionnons différentes variables pouvant expliquer le salaire dans chaque strate. Un tronc commun de variables explicatives est formé par l'ancienneté dans la profession et son carré, l'emploi atypique ou non, et le diplôme du salarié. Pour les salariés du privé nous y ajoutons le type de contrat, le fait d'avoir un emploi en Île-de-France ou pas, la proportion de femmes dans le secteur et le fait d'être cadre ou pas. Enfin pour les salariés du public, en plus des variables du tronc commun, nous complétons avec le fait d'être enseignant ou pas, fonctionnaire d'État ou pas et le grade.

Le salaire mensuel ou le ratio entre les salaires des deux années consécutives est imputé. Nous tenons ensuite compte du nombre de mois d'activité déclaré à l'enquête pour estimer le salaire annuel. Un travail particulier est nécessaire pour les salariés à temps partiel.

Les salaires des non-salariés

Nous traitons également les salaires d'individus dont l'activité principale n'est pas salariée et pour lesquels nous ne disposons donc pas d'information sur l'activité rémunérée par un salaire. Le petit nombre de cas et le manque d'information, nous ont amenés à imputer pour ces individus des salaires moyens de personnes ayant les mêmes caractéristiques.

Préretraites

Un petit nombre de préretraites sont à imputer : l'individu a déclaré percevoir ce revenu et n'en a pas donné le montant. Nous nous sommes contentés de lui attribuer le montant moyen de préretraites d'individus ayant des caractéristiques similaires.

Allocations de vieillesse (PY100N) ou pension au survivant (PY110N)

Les retraites sont traitées différemment des autres revenus. L'hypothèse majeure justifiant ce choix est que les retraites sont quasi stables. Comme nous ne disposons pas d'une retraite courante (relative au mois de l'enquête), nous utilisons le revenu courant comme élément de contrôle et nous le comparons à la somme des revenus de l'année de référence du ménage. Pour les ménages concernés, la retraite est un élément prépondérant du revenu total, ce qui justifie la comparaison.

Nous ajoutons un autre contrôle, cette fois entre les années 2006, 2005, 2004 et 2003, toujours sous la même hypothèse nous comparons les impôts sur le revenu déclarés en 2006, 2005, 2004 et 2003 que nous calculons. Deux contrôles sont donc possibles avant de prendre une décision.

À l'aide d'un logiciel approprié nous pouvons observer, pour le ménage, l'ensemble des revenus cumulés, le nombre de mois de retraite déclaré, le nombre d'individus et leur description : âge, catégorie socioprofessionnelle, etc. L'expertise a fait ressortir en 2004 des cas types d'erreurs, ce qui permet de corriger automatiquement une grande partie des retraites erronées. Les erreurs les plus fréquentes portent sur les confusions francs/euros, ainsi que sur les périodes de perception.

Lorsqu'il n'a pas été possible de déterminer l'origine de l'erreur, le montant de la retraite est imputé. Deux strates sont utilisées, suivant que le conjoint de la personne retraitée est vivant ou pas. Pour les personnes dont le conjoint n'est pas décédé les variables explicatives du montant de la retraite perçue sont le sexe, le secteur d'activité (privé ou public), la qualification, le diplôme, l'âge et son carré, et l'ancienneté dans la profession. Pour les retraités dont le conjoint est décédé, ces variables sont complétées par le secteur d'activité de l'ex-conjoint ainsi que sa qualification.

Selon le rang d'interrogation de l'individu, la retraite ou le ratio des retraites des deux années consécutives est estimé afin d'imputer un montant.

Suite à ce travail, la distribution des montants des retraites perçues est proche de celle de l'ERF.

Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante

Ces revenus sont collectés sous deux formes : d'une part la forme fiscale, comprenant les amortissements et autres abattements, et d'autre part la forme privée, correspondant au revenu net déterminé par le ménage. Le revenu fiscal est jugé peu réaliste, c'est pourquoi le revenu privé est privilégié. Ainsi si les deux revenus sont renseignés dans l'enquête et si le revenu privé est vraisemblable, le revenu privé est le seul pris en compte. À défaut de revenu privé, le revenu fiscal est retenu. Si les deux types de revenus sont manquants ou peu crédibles, l'imputation se fait par hot-deck.

Prestations familiales

Les prestations familiales sont calculées sur barème pour l'ensemble des ménages. La principale difficulté est la période de référence des revenus pour les prestations sous condition de ressources. Pour les Caisses d'allocations familiales (Caf), jusque juillet d'une année N, les revenus retenus pour le calcul des aides est celui de l'année N-2 ; à partir de juillet N, les revenus retenus sont ceux de N-1, Nous utilisons uniquement les revenus de l'année N pour imputer les prestations de l'année N.

Les imputations des prestations familiales n'ont pas posé de problèmes majeurs, les résultats sont conformes aux données des CAF. Les valeurs calculées sont imputées si l'individu n'a pas donné de montant ou si le ratio valeur calculée/valeur collectée n'est pas dans la fourchette 0,2--5.

Aides au logement

Les aides au logement sont calculées sur barème pour tous les locataires et les accédants à la propriété. Nous utilisons le loyer déclaré à l'enquête, et à défaut, un loyer imputé. La masse des allocations collectées est inférieure de 10 % aux données de la Cnaf corrigées de la différence de champ. Des aides sont donc attribuées à certains ménages, de façon aléatoire, afin de disposer du bon nombre de bénéficiaires. Le calcul des aides pour les locataires ne pose pas de problème majeur même si, comme pour les prestations familiales, la période des revenus n'est pas exactement celle retenue par les CAF. Le calcul des aides aux accédants à la propriété diffère néanmoins du calcul des mêmes aides dans l'ERF. Nous avons en effet choisi d'appliquer le barème locatif aux accédants, en nous aidant du loyer fictif imputé.

Minima sociaux

Trois minima sociaux sont imputés dans SILC : le revenu minimum d'insertion (RMI), le minimum vieillesse et l'allocation de parent isolé (API). Si un ménage déclare avoir perçu un minimum, celui-ci est comparé au barème ; il n'est changé qu'en cas de différence majeure. C'est un moyen de rectifier les erreurs d'unité monétaire. Selon les données brutes, SILC comprend 50 % des bénéficiaires du

RMI, 15 % des bénéficiaires du minimum vieillesse et 50 % des bénéficiaires de l'aide au parent isolé (API).

Les méthodes utilisées sont proches des méthodes d'imputation utilisées dans l'ERF. Elles présentent une limite. Le revenu retenu pour le calcul du RMI est un revenu trimestriel que nous ne connaissons pas. Le RMI imputé est égal à la différence entre le plafond du RMI et les revenus de l'année 2004. Ce plafond dépend du type de famille et du nombre de personnes à charge. Ainsi calculé, aucun minimum ne peut être imputé à une famille dont les revenus annuels sont supérieurs au plafond. Cette famille a pourtant pu être éligible, si les revenus d'un trimestre se sont avérés insuffisants. C'est la raison pour laquelle nous n'avons pas corrigé les montants déclarés à l'enquête, pour les ménages dont le revenu annuel est supérieur au plafond.

Nous n'avons pas traité les dispositifs d'intéressement à la prise ou à la reprise d'un emploi.

Aucun calage n'a été fait, le nombre de bénéficiaires du RMI après imputation étant suffisamment proche des données connues.

Loyer imputé

Un loyer fictif est imputé aux propriétaires, aux accédants, aux usufruitiers, aux logés gratuitement et aux locataires payant un loyer inférieur au prix du marché.

La méthode retenue comprend quatre étapes, dont deux régressions :

- 1) Estimation d'une équation de loyer à partir des données de l'enquête logement 2002. L'estimation porte sur les logements locatifs du parc privé hors loi de 1948. Les variables explicatives sont les caractéristiques du logement (surface, confort, sanitaires, équipement, état du logement ...) et de localisation (tranche de taille d'agglomération, zone climatique, typologie socio-économique de Nicole Tabard ...). Deux équations distinctes sont estimées, l'une pour les appartements (variable expliquée : le logarithme du loyer au m²) et l'autre pour les maisons (variable expliquée : le logarithme du loyer, la surface figurant parmi les variables explicatives).
- 2) Les équations précédemment estimées sont utilisées pour imputer un loyer fictif aux propriétaires occupants ainsi qu'aux ménages logés gratuitement et un loyer de marché aux locataires du parc social ou en Loi de 1948 de l'enquête logement. On a rajouté à la valeur issue de l'équation un résidu tiré selon une procédure de hot-deck stratifié.
- 3) Ce loyer imputé est régressé sur deux types de variables : des variables du tronc commun des enquêtes ménages de l'Insee d'une part, et des variables géographiques d'autre part. À caractéristiques sociodémographiques et de localisation identiques, les logements occupés par les accédants à la propriété sont d'une qualité moyenne supérieure à ceux des propriétaires sans charge de remboursement, qui sont eux-mêmes de meilleure qualité que ceux du parc social. Estimer une seule équation aurait pu biaiser les estimations. Aussi huit régressions distinctes ont-elles été estimées sur des segments relativement homogènes du parc :
 - appartements, propriétaires sans charge de remboursement et ménages logés gratuitement
 - appartements, accédants à la propriété ;
 - appartements, locataires du parc social ou loi de 1948 ;
 - appartements, locataires du parc libre louant vide ;
 - maisons, propriétaires sans charge de remboursement et ménages logés gratuitement ;
 - maisons, accédants à la propriété ;
 - maisons, locataires du parc social ou loi de 1948 ;
 - maisons, locataires du parc libre louant vide.

² La crise du logement consécutive à la Seconde Guerre mondiale a obligé les autorités à prendre des mesures contre la flambée des loyers en les fixant par décret. C'est l'objet de la loi du 1^{er} septembre 1948. Elle est applicable aux immeubles construits antérieurement à cette date principalement dans des communes de plus de 4 000 habitants.

- 4) Les huit équations estimées sont exportées vers l'enquête SILC pour y imputer soit un loyer fictif soit un loyer manquant. Lors de l'imputation on rajoute à la valeur prédite un résidu tiré selon une procédure de hot-deck stratifié.
- 5) Les résultats avant prise en compte de l'augmentation des loyers sont présentés ici.

Loyers mensuels réels ou fictifs en appartement

	Propriétaires		Accédants		Locataires parc social		Locataires parc libre	
	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC
Effectif	2 441 483	2 253 695	1 167 597	1 146 581	4 238 130	3 923 845	3 316 355	3 213 289
Moyenne	703	666	674	623	313	488	479	489
Minimum	57	77	115	73	14	61	32	33
Q5	259	251	281	279	164	215	230	189
Q10	314	295	332	308	191	255	269	247
Q25	425	415	432	450	236	327	345	327
Q50	579	565	588	568	290	436	430	452
Q75	855	815	808	720	361	579	552	591
Q90	1 183	1 215	1 081	971	448	767	691	782
Q95	1 437	1 440	1 376	1 247	521	928	827	928
Maximum	10 438	4 456	3 968	2 262	3 637	3 655	3 951	2 063

Loyers mensuels réels ou fictifs en maison individuelle

	Propriétaires		Accédants		Locataires parc social		Locataires parc libre	
	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC	Enquête Logement 2006	EU-SILC
Effectif	8 428 662	8 439 957	3 972 366	4 441 335	820 883	1 044 488	1 505 235	1 698 715
Moyenne	482	466	643	631	349	464	500	479
Minimum	30	32	49	98	37	67	36	58
Q5	157	151	259	270	159	179	189	180
Q10	201	191	329	329	197	211	260	255
Q25	293	277	452	450	267	308	352	355
Q50	414	400	590	585	344	420	462	451
Q75	580	568	767	762	418	576	605	583
Q90	799	767	997	947	487	755	760	720
Q95	986	941	1 733	1 081	540	897	875	838
Maximum	7 383	6 046	5 699	5 551	2 251	2 239	3 160	2 355

3 COMPARABILITÉ

3.1 Concepts et définitions de base

Population de référence

La population de référence de l'enquête est constituée par l'ensemble des ménages ordinaires (hors institutions) dont la résidence principale se situe en France métropolitaine. En 1999, 2,2 % de la population de la France métropolitaine vivait en collectivité, en institution ou était sans domicile fixe³. Les habitants des DOM-TOM représentaient alors 2,8 % des personnes vivant en ménage ordinaire.

Définition du ménage privé

Un ménage est dans l'enquête française SILC « une personne vivant seule ou un groupe de personnes vivant ensemble qui partagent les dépenses et participent à une économie ménagère commune ». On considère que c'est en contribuant aux dépenses de l'unité de vie que les membres du ménage mettent en commun leurs ressources.

1% des logements dans l'enquête SILC contiennent plusieurs ménages qui constituent des unités de vie indépendantes.

Appartenance au ménage

Font partie d'un même ménage les individus qui résident dans une même résidence habituelle en faisant budget commun. Un ménage n'est enquêté que dans sa résidence principale. En première vague, nous n'interrogeons que les unités de vie ayant pour résidence principale un logement de l'échantillon.

Période(s) de référence du revenu utilisée(s), périodes utilisées pour les impôts sur le revenu et les cotisations sociales

L'enquête est conduite en mai-juin 2007. L'année de référence est 2006. On collecte au niveau individuel et ménage les revenus perçus au cours de l'année 2006. On collecte au niveau du ménage les impôts payés en 2006 au titre des revenus perçus au cours de l'année 2005. Les cotisations sociales sont relatives aux revenus perçus en 2006.

Période de référence pour les impôts sur la fortune

On considère le montant de l'impôt de solidarité sur la fortune payé en 2006. Il est relatif au patrimoine détenu au 1^{er} janvier 2006.

Écart entre la période de référence du revenu et les variables actuelles

Les variables actuelles (par exemple les variables relatives au coût du logement) portent sur la période de collecte, c'est-à-dire les mois de mai et juin 2007, ce qui rajoute une troisième période de référence après l'année 2006 (revenus déclarés) et l'année 2005 (assiette des impôts sur le revenu).

Durée totale de la collecte de données pour l'échantillon.

La collecte SILC s'est déroulée du 14 mai au 30 juin 2007.

Informations générales sur l'activité professionnelle durant la période de référence.

La catégorie socioprofessionnelle des actifs ou des anciens actifs est codée automatiquement après la collecte par le système SICORE. 95 % des professions sont codées automatiquement, les autres professions étant « reprises » à la main par une équipe spécialisée à l'Insee.

³ Une enquête a été menée en 2001 auprès des sans domiciles fixes. Les résultats de cette enquête ont été publiés dans le n°391-392 de la revue *Économie et Statistique*.

Le codage en ISCO se fait à partir du code des professions et catégories socioprofessionnelles (PCS) à quatre chiffres, de la nomenclature d'activité française en 712 postes et du nombre de salariés de l'établissement. Dans 95 % des cas, la matrice de passage élaborée pour l'enquête Emploi permet de coder la profession de l'enquêté en un code ISCO à deux positions. Sinon, la codification en ISCO a dû se faire par imputation statistique. Le code ISCO est alors affecté aléatoirement en fonction de la répartition du code ISCO par PCS à deux chiffres dans l'enquête Emploi du deuxième trimestre de 2007.

Dans 1,6 % des cas, il y a imputation en raison d'un manque d'information dans l'enquête (PCS sur 2 positions au lieu de 4, activité codée seulement en 62 postes). Dans 3,8 % des cas, il s'agit d'un défaut de la matrice de passage entre les PCS et ISCO. Dans ces cas, l'enquête Emploi fournit un codage en une position, ce que le règlement SILC n'autorise pas alors que l'information collectée sur la profession est plus précise dans l'enquête Emploi que dans SILC.

3.2 Composantes du revenu

3.2.1 Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC

Les fichiers français respectent les définitions des statistiques EU-SILC.

Loyers imputés (HY030N), avantages en nature (PY020N), et valeur des biens produits pour la consommation propre (PY070N)

Ces trois composantes du revenu sont obligatoires à partir de 2007. Il n'est pas possible d'isoler des salaires l'avantage en nature que représente la voiture de fonction. La variable PY021 n'est donc pas renseignée.

Revenus de la propriété (HY040N et HY090N)

Pas de classification particulière.

Allocations famille/enfants (HY050)

Elles incluent l'ensemble des allocations versées par les Caisses d'allocations familiales hormis les allocations logement et le RMI. La plupart sont sans conteste des prestations en espèces versées pour faire face aux dépenses liées à l'éducation des enfants ou à la perte de revenu consécutive à l'arrêt d'activité d'un des parents pour élever ses enfants :

- l'allocation d'adoption (ADA) versée pendant 21 mois après l'adoption sous condition de ressources ;
- l'allocation parentale d'éducation (APE) versée aux parents qui arrêtent ou ont arrêté totalement ou partiellement de travailler pour élever leurs enfants. L'APE de rang 2 (2 enfants à charge) est versée à condition d'avoir travaillé deux ans (consécutifs ou non) dans les 5 ans précédant l'arrivée du dernier enfant. L'APE de rang 3 (3 enfants ou plus à charge) est versée à condition d'avoir travaillé deux ans au cours des dix ans précédant l'arrivée du dernier-né ;
- les allocations familiales qui sont versées à toutes les familles d'au moins deux enfants à charge ;
- le complément familial versé aux familles de trois enfants ou plus dont le benjamin est âgé d'au moins trois ans. Cette allocation est versée sous condition de ressources ;
- l'allocation pour jeune enfant (APJE) versée aux familles ayant un jeune enfant (de moins de trois ans), elle est aussi versée sous condition de ressources ;
- l'allocation parent isolé (API) est un minimum social versé aux personnes seules élevant seules un enfant. Cette allocation est versée durant douze mois ou bien jusqu'au mois précédant le troisième anniversaire du dernier-né ;
- l'allocation de soutien familial (ASF) est versée quand un parent ne concourt pas aux besoins de l'enfant ;
- l'allocation de rentrée scolaire (ARS) est versée en début d'année scolaire sous condition de ressources ;

- l'allocation d'éducation spéciale (AES) pour enfants handicapés ;
- l'allocation de présence parentale (APP) pour enfants handicapés ou malades en cas d'arrêt total ou partiel de travail des parents.

L'inclusion de l'aide à la famille pour l'emploi d'une assistante agréée (AFEAMA) et de l'allocation de garde d'enfant à domicile (AGED) dans les allocations familiales est plus discutable. Ces prestations pourraient en effet être classées en prestations en nature puisqu'elles couvrent une partie des dépenses liées à la garde des enfants. Nous avons cependant choisi de les inclure dans les prestations familiales. En effet, ces prestations ainsi que l'APJE et l'APE sont progressivement remplacé par la prestation d'accueil du jeune enfant (PAJE) et peuvent donc plus être distinguées des autres prestations familiales. La mise en place de la PAJE se passe de la façon suivante : les ménages qui ont un nouveau-né en 2004 passent entièrement sous le régime de la PAJE, les autres continuent de bénéficier des anciennes allocations.

En outre, les indemnités pour congé maternité ne sont pas incluses dans les allocations familiales, mais dans les indemnités maladie.

Dépenses liées à l'exclusion sociale non classées ailleurs (HY060)

Elles incluent le RMI et les aides financières sociales versées par les collectivités locales.

Aides au logement (HY070)

Pas de classification particulière.

Transferts interménages perçus/versés régulièrement en espèces (HY080 et HY130)

Ces transferts excluent les versements exceptionnels mais incluent le paiement de loyer par un tiers. Ce paiement a en effet été considéré comme un paiement en espèces.

Revenus perçus par les personnes de moins de 16 ans (HY110N)

Les enfants ne travaillant pas avant 16 ans, ils ne reçoivent donc pas de revenus d'activité. En outre, cet item doit exclure les transferts entre ménages et donc les pensions alimentaires. Cette composante est ainsi égale aux bourses d'école reçues par les élèves de famille à faibles revenus.

Impôts réguliers sur la fortune (HY120)

Cet item est égal à l'impôt de solidarité sur la fortune.

Impôts sur le revenu et cotisations sociales (HY140G) et remboursements/encaissements liés à des ajustements d'impôt (HY145N)

L'agrégat HY140G comprend le HY145N, les contributions et cotisations sociales. Celles-ci contributions sont assises sur les salaires, les pensions, les revenus d'indépendants, les allocations chômage, les retraites, les prestations familiales et les allocations logement.

L'agrégat HY145N comprend l'impôt sur le revenu des personnes physiques (IRPP), la taxe d'habitation et les impôts payés à l'étranger. La prime pour l'emploi est comptée négativement. L'IRPP assis sur les revenus de l'année N est payé l'année N+1 sous forme d'acomptes et d'un solde. Il n'est donc pas prélevé à la source et porte sur plusieurs types de revenus tous perçus l'année précédente, c'est pourquoi nous l'avons intégralement enregistré comme ajustement d'impôt. La taxe d'habitation est payée par les personnes occupant un logement au premier janvier, elle est assise sur la valeur locative du logement mais son montant dépend aussi des revenus du ménage. La prime pour l'emploi est versée aux actifs aux revenus faibles, elle est déduite de l'IRPP.

Salaire ou revenu assimilé (PY010N)

Les salaires sont nets des cotisations sociales salariées. Ils comprennent les avantages liés à la voiture de fonction.

Cotisations sociales des employeurs (PY030G)

Elles comprennent les cotisations sociales versées au régime général, mais aussi certaines taxes assises sur les salaires (versement transport, Fond national d'aide au logement, taxe d'apprentissage, formation professionnelle).

Bénéfices en espèces ou pertes de trésorerie en rapport avec une activité indépendante (y compris honoraires) (PY050N)

Nous avons privilégié les prélèvements pour usage propre. Le revenu fiscal n'est utilisé que comme approximation en cas de non-disponibilité de la première information.

Valeur des biens produits pour la consommation propre (PY070N)

Il s'agit de la production agricole du ménage au prix du marché : le montant est estimé par le ménage.

Allocations de chômage (PY090N)

Elles contiennent les allocations chômage et les préretraites pour motif économique.

Allocations de vieillesse (PY100N)

Ce sont les retraites, les préretraites (à l'exclusion de celles incluses dans les allocations chômage et les pensions d'invalidité) et le minimum vieillesse. Le minimum vieillesse est un minimum social versé sous condition de ressources aux personnes ayant atteint 65 ans.

Cet agrégat ne comprend pas les pensions de réversion et d'invalidité versées aux personnes ayant dépassé l'âge de la retraite.

Pensions au survivant (PY110N)

Ce sont les pensions de réversion, c'est-à-dire les pensions reçues par les survivants quel que soit leur âge.

Indemnités de maladie (PY120N)

Elles contiennent les indemnités journalières pour congés maladie, accidents du travail, congés maternité et les pensions d'accidents du travail.

Pensions d'invalidité (PY130N)

Les allocations suivantes sont comptées dans cette composante du revenu :

- l'allocation adulte handicapé (AAH) qui est versée aux handicapés de plus de 20 ans sous condition de ressources,
- les pensions militaires d'invalidité ou de victime de guerre,
- les pensions d'invalidité,
- les préretraites pour cause de diminution de la capacité de travail,

Sont aussi comptabilisées dans cet agrégat des aides permettant de financer la garde de personnes invalides ou dépendantes :

- l'aide personnalisée à l'autonomie (APA) qui est versée aux personnes âgées dépendantes qui ont recours à une aide à domicile,
- la prestation spécifique dépendance (PSD) qui remplit le même rôle que l'APA,
- l'allocation compensatrice pour tierce personne (ACTP), qui est versée aux handicapés de moins de 60 ans qui ont besoin d'une aide à domicile.

Les pensions versées aux personnes ayant atteint l'âge de la retraite sont incluses dans cet agrégat.

Allocations d'études (PY140N)

Il s'agit de l'ensemble des bourses versées aux plus de 16 ans.

3.2.2 *La source ou la procédure utilisée pour la collecte des variables du revenu*

Les revenus sont collectés par enquête (cf. la description de la procédure utilisée pour recueillir les revenus, partie 2.3.2)

3.2.3 *La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes*

Les variables de revenu ont été collectées nettes des cotisations sociales. En cas de recours à la déclaration de revenus, certains montants étaient y compris contributions sociales non déductibles. En l'absence d'abattement, il a été facile de réestimer les montants nets.

3.2.4 *La méthode utilisée pour obtenir les variables cibles du revenu sous la forme requise (c'est-à-dire en tant que valeur brute)*

Les impôts sont collectés. Les cotisations sociales ont été imputées sur barème.

Des valeurs brutes et nettes ont été estimées pour les agrégats suivants.

Salaire ou revenu assimilé (PY010 et PY030G)

Le salaire collecté est dans la plupart des cas le salaire déclaré au fisc.

À partir de cette valeur nous calculons le salaire brut correspondant. Pour cela trois groupes de salariés sont constitués : les agents de l'État, les salariés du privé non cadres et les cadres salariés du privé. En fonction du montant du salaire nous calculons sur barème le salaire brut correspondant au salaire déclaré. Nous cumulons les cotisations dues pour la partie du salaire inférieure à un plafond de la sécurité sociale, puis celles pour la fraction comprise entre un et trois plafonds de la sécurité sociale, puis entre trois et quatre plafonds, entre quatre et huit plafonds et enfin au-delà de huit plafonds. Les taux de certaines cotisations sont en effet différents selon le montant du salaire. Nous avons ainsi reconstitué le salaire brut à partir duquel les diverses cotisations sociales salariales et employeurs sont calculées à l'aide du barème.

Sont ainsi estimés la contribution sociale de solidarité (CSG) et la contribution de remboursement de la dette sociale (CRDS) non déductibles, la CSG déductible, les cotisations du régime général (maladie, vieillesse, veuvage, chômage et famille), les cotisations aux retraites complémentaires salariées (reçues par l'Agirc) et cadres (reçues par l'Arrco), le versement transport, la cotisation au Fond national d'aide au logement, la taxe d'apprentissage, la participation des employeurs à la formation professionnelle. Afin de contrôler les cotisations obtenues, des cas types ont été testés sur le logiciel de l'Union de recouvrement des cotisations de sécurité sociale et d'allocations familiales (URSSAF) donnant des résultats satisfaisants.

Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050)

Les cotisations relatives aux revenus des indépendants sont estimées en répartissant les revenus en trois groupes : bénéfices agricoles, bénéfices industriels et commerciaux, bénéfices non commerciaux. Les régimes sociaux des indépendants sont très divers ; des paramètres législatifs moyens ont donc été utilisés.

Retraites de régimes privés, allocations chômage (PY090), de vieillesse (PY100) et pension au survivant (PY110)

La valeur brute est égale à la valeur nette à laquelle on ajoute la CSG et la CRDS. Ces contributions sont simples à estimer parce que leur taux ne dépend que du type de revenus et non du montant de revenu.

Loyers imputés (HY030)

La valeur nette est égale à la valeur brute, qui est imputée, diminuée de la taxe foncière payée sur la résidence principale, qui est collectée.

Prestations familiales (HY050) et allocations logement (HY070)

La valeur brute est égale à la valeur nette, collectée, à laquelle s'ajoute la CRDS, imputée.

Cotisations à un régime de retraite privé (PY035), valeur des biens produits pour la consommation propre (PY070), indemnités maladie (PY130), transferts entre ménage (HY080 et HY130), remboursement de prêts hypothécaires (HY100)

La valeur brute est égale à la valeur nette.

4 COHÉRENCE

4.1 Comparaison des variables cibles avec des sources extérieures

Les données externes ayant servi à contrôler les éléments constitutifs du revenu sont diverses. Nous avons utilisé l'enquête Revenus fiscaux et sociaux et l'enquête Logement. L'enquête Revenus fiscaux et sociaux est considérée comme la source de référence sur la distribution des revenus et des niveaux de vie en France. Elle consiste en un appariement entre l'enquête Emploi et les déclarations fiscales. Les prestations sociales y étaient imputées sur barèmes jusqu'en 2005 puis collectées auprès des organismes sociaux depuis 2006.

Nous avons aussi mobilisé les statistiques de la Caisse nationale des Allocations familiales. Enfin certains dossiers sont venus compléter cette aide extérieure, entre autres le dossier sur les revenus sociaux 2007 publié par la Direction de la recherche, des études, de l'évaluation et des statistiques.

4.1.1 Salaire et revenu assimilé (PY010N)

	EU-SILC		ERFS	
Année de revenus	2005	2006	2005	2006
Salaires (milliards d'€)	449	448	489	509
Bénéficiaires (millions)	25,1	25,4	26,2	26,4
Moyenne	17 861	17 622	18 617	19 258
Minimum	10	10	1	1
D10	3 744	3 869	3 151	3 487
D20	8 350	7937	7 694	8 041
D30	11 911	11 607	11 882	12 505
D40	13 996	13 929	14 430	14 927
D50	15 888	15 925	16 492	17 034
D60	18 011	18 000	18 665	19 276
D70	20 467	20 649	21 497	22 058
D80	24 156	24 021	25 377	26 103
D90	30 974	30 840	32 640	33 761
Maximum	338 928	1 590 389	847 620	765 766

Les écarts entre les deux enquêtes sont de l'ordre de 13 % en masses et de 4 % en nombre de bénéficiaires. Ces variations trouvent en partie leur origine par des différences de champ entre les deux enquêtes. Le concept fiscal est plus large que celui de l'enquête SILC. C'est ainsi que les indemnités maladie sont incluses dans le salaire fiscal et isolées dans celui du panel. Les bourses d'étudiants sont traitées de la même façon. C'est aussi le cas des indemnités de départ en retraite ou en préretraite. C'est pourquoi les retraites sont sensiblement supérieures dans le panel. L'effet mémoire n'est pas non plus à écarter, il peut arriver dans le panel qu'un salarié oublie de mentionner certaines primes.

4.1.2 Les allocations chômage (PY090N)

	EU-SILC		ERFS	
Revenus	2005	2006	2005	2006
Chômage (milliards d'€)	26,0	24,8	27,38	25,11
Bénéficiaires (millions)	3,82	3,75	4,75	4,46
Moyenne	6 827	6 612	5764	5 629
Minimum	27	28	1	8
D10	833	800	680	621
D20	1 522	1 576	1 344	1 325
D30	2 480	2 544	2 157	2 144
D40	3 648	3 575	3 141	3 059
D50	4 800	4 788	4 226	4 078
D60	5 889	5 600	5 255	5 346
D70	7 599	7 460	6 902	6 510
D80	9 600	9 600	9 128	8 833
D90	13 488	12 653	12 227	11 739
Maximum	153 643	157 154	61 737	74 702

Les écarts sur l'estimation du nombre de bénéficiaires sont plus importants sur ce poste. La différence observée s'explique par des différences de définition. SILC ne prend en compte que les allocations chômage et les préretraites consécutives à un licenciement, tandis que ERF cumule dans ce poste les allocations chômage, toutes les allocations de préretraites du privé ou les allocations de congés de fin d'activité du secteur public, les indemnités de départ à la retraite, les primes de départ en retraite, les indemnités de membre de divers conseils, les indemnités des élus et des parlementaires.

4.1.3 Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension au survivant (PY110N)

	EU-SILC		ERF	
Année de revenus	2005	2006	2005	2006
Retraites (milliards d'€)	192	202	191	201
Bénéficiaires (millions)	12,7	13,6	14,4	14,6
Moyenne	15 039	14 850	13 274	13 732
Minimum	47	31	17	1
D10	4 463	4 340	2 503	2 696
D20	7 368	7 313	4 571	5 090
D30	9 357	9 245	7 392	7 778
D40	11 234	11 282	9 504	9 847
D50	13 200	13 327	11 691	12 139
D60	15 184	15 189	13 922	14 484
D70	17 379	17 405	16 323	16 922
D80	20 736	20 707	19 566	20 307
D90	26 499	26 073	25 033	25 839
Maximum	241 363	241 363	191 999	230 839

Certains revenus sont inclus dans les salaires (indemnités de départ en retraite ou préretraite) ou les minima sociaux (le minimum vieillesse) pour l'ERF et dans les retraites pour le panel, ceci peut expliquer le léger écart en masse concernant les retraites.

4.1.4 Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante

	EU-SILC		ERF	
Année de revenus	2005	2006	2005	2006
Revenus d'indépendants (milliards d'€)	53,0	51,7	54,5	54,2
Bénéficiaires (millions)	2,21	2,16	2,05	2,02
Moyenne	24 048	23 947	26 578	26 758
Minimum	57	100	6	1
D10	2 286	3 000	1 468	1 486
D20	6 000	7 000	3 907	3 662
D30	9 000	10 000	6 773	6 675
D40	12 000	13 000	9 883	9 847
D50	15 000	17 000	14 250	13 200
D60	20 000	20 603	19 992	17 816
D70	24 000	25 200	26 660	23 915
D80	32 355	33 060	39 853	36 355
D90	52 000	48 837	61 543	59 449
Maximum	500 000	548 816	742 739	2 410 096

Les revenus des indépendants sont abordés de façon différente dans l'ERF et le panel, aussi les comparaisons ne sont pas possibles. Le panel privilégie le revenu privé, à savoir les sommes que l'indépendant prélève sur les ressources de l'entreprise à des fins privées, pour consommer ou

épargner. L'ERF enregistre un résultat fiscal ayant pour base un forfait, un bénéfice réel ou un régime micro-entreprise, résultant d'abattements, d'amortissements et de déductions.

4.1.5 Revenus sociaux (HY050G, HY060G et HY070G),

	EU-SILC		ERF	
Année de revenus	2005	2006	2005	2006
Sociaux (milliards d'€)	37,1	41,25	40,2	41,99
Bénéficiaires (millions)	10,13	10,34	9,88	10,76

Les deux enquêtes sont proches sur ces revenus, ici les bénéficiaires se comptent en ménages.

4.1.6 Prestations familiales (HY050N),

	EU-SILC		Cnaf	
Année de revenus	2005	2006	2005	2006
Prestations familiales (milliards d'€)	20,4	22,00	23,8	25,3
Bénéficiaires (millions)	6,5	6,3		

Les différentes allocations versées pour les familles sont regroupées ici. La comparaison est faite avec les statistiques de la Caisse nationale des allocations familiales. Le champ couvert par la Cnaf est plus large que celui de SILC : les bénéficiaires ne vivant pas en ménage ordinaire y sont inclus. Cette différence de champ ne peut cependant expliquer l'ensemble de l'écart.

La Cnaf ne publie que le nombre de bénéficiaires au 31 décembre 2007 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2007.

4.1.6.1 Aides au logement (HY070N)

	EU-SILC		Cnaf	
Année de revenus	2005	2006	2005	2006
Allocations logements (milliards d'€)	11,6	14,4	12,8	13,2
Bénéficiaires (millions)	5,9	6,5		

La Cnaf ne publie que le nombre de bénéficiaires au 31 décembre 2007 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2007.

4.1.7 Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation).

	EU-SILC		ERF	
Année de revenus	2005	2006	2005	2006
Impôts (milliards d'€)	53,6	53,7	53,8	53,9

Les deux enquêtes diffèrent de 0,4% sur la masse de l'impôt sur le revenu. La période de référence n'est pas la même : l'ERF retient l'impôt de l'année des revenus tandis que le panel collecte l'impôt de l'année précédant les revenus. Mais ceci n'explique qu'une partie de la différence observée.

4.1.8 Loyers

Les comparaisons montrent que les loyers collectés dans EU-SILC sont de bonne qualité.

Un loyer est attribué à chaque locataire dans EU-SILC. Le loyer est collecté par enquête. En cas de non-réponse ou de loyer en dessous des prix du marché, un loyer fictif lui est attribué (cf. partie 2.6).

Nous présentons les comparaisons avec l'enquête logement 2006.

	Enquête Logement 2006	EU-SILC 2007
Effectif	10 353 930	9 880 337
Moyenne	400	409
Q0	14	1
Q5	175	166
Q10	211	205
Q25	269	271
Q50	356	370
Q75	475	498
Q90	617	639
Q95	728	750
Q100	3 951	3 033

Un loyer fictif est attribué à chaque ménage propriétaire, accédant à la propriété ou logé gratuitement (cf. la description de la méthode et les résultats obtenus dans la partie 2.6).

4.1.9 Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)

Dans le panel EU-SILC, les revenus du patrimoine financier sont calculés à partir des stocks collectés. Six postes sont distingués lors de la collecte, les livrets d'épargne exonérés, les livrets bancaires soumis à l'impôt sur le revenu, l'épargne logement, les revenus de valeurs mobilières, les produits d'assurance vie, les autres placements financiers. Un taux de rendement issu de l'enquête Patrimoine de 2004 est appliqué aux stocks collectés afin d'estimer le revenu.

La comparaison entre l'ERF et EU-SILC n'est pas possible, tant les champs semblent différents. Un grand nombre de produits sont en effet exonérés fiscalement (livret A, bleu ...) ou placés dans des enveloppes elles-mêmes exonérées (PEA, PEP ...). En outre l'ERF ne recueille qu'une partie des revenus de placements financiers soumis au prélèvement libératoire.

Nous pouvons cependant observer que le nombre de bénéficiaires et la masse totale sont, comme attendu, supérieurs dans EU-SILC.

	EU-SILC		ERF	
Année des revenus	2005	2006	2005	2006
Revenus du patrimoine financier (milliards d'€)	19,44	20,21	17,03	22,9
Bénéficiaires (millions)	20,12	19,84	7,95	8,4
Moyenne	966	1 019	2 143	2 702
Minimum	2	2	1	1
D10	25	25	11	12
D20	64	67	35	44
D30	92	92	79	106
D40	218	253	152	206
D50	275	325	269	363
D60	499	513	450	606
D70	768	788	770	1 037
D80	1 238	1 371	1 377	1 940
D90	2 483	2 680	3 196	4 190
Maximum	19 724	20 040	600 425	990 638

4.1.10 Revenu disponible du ménage et niveau de vie des individus.

	EU-SILC		ERF	
Année des revenus	2005	2006	2005	2006
Revenus disponible des ménages (milliards d'€)	745,10	759,62	749,66	842,52
Bénéficiaires (millions)	25,99	26,27	25,66	25,88
Moyenne	28 669	28 911	29 180	32 557
Minimum	-25 007	52	-965 192	9
D10	10 836	11 238	10 999	12 207
D20	14 138	14 386	14 340	15 707
D30	17 146	17 450	17 573	19 127
D40	20 626	21 006	21 045	22 950
D50	24 461	24 905	24 804	27 152
D60	28 651	29 107	29 076	31 701
D70	33 345	33 861	33 883	36 617
D80	39 474	40 442	40 254	43 782
D90	49 495	50 469	50 897	56 290
Maximum	351 380	1 041 917	749 129	2 146 463

	EU-SILC		ERF	
Revenus	2005	2006	2005	2006
Niveau de vie des individus (milliards d'€)	473,69	487,44	468,44	537,07
Bénéficiaires (millions)	59,40	59,57	59,78	59,46
Moyenne	18 226	18 588	18 258	20 605
Minimum	-16 671	52	-419 649	9
D10	8 970	9 321	9 078	9 724
D20	11 086	11 488	11 239	12 091
D30	12 855	13 192	12 951	13 962
D40	14 527	14 803	14 575	15 756
D50	16 186	16 623	16 274	17 598
D60	18 026	18 440	18 140	19 685
D70	20 258	20 861	20 361	11 241
D80	23 475	24 073	23 412	25 803
D90	29 131	29 627	29 310	33 194
Maximum	195 211	520 959	499 419	1 430 975

Ces comparaisons montrent que les revenus estimés dans EU-SILC sont relativement proches des revenus estimés par l'ERF.

Annexe méthodologique

En présence de non-réponse, les estimateurs de sondage ne sont pas calculables tels quels. Une possibilité est de procéder à des imputations. On distingue souvent les méthodes dites « déterministe » ou « aléatoire ». Dans le cadre d'un modèle linéaire, une imputation aléatoire simple, par opposition à l'imputation déterministe, consiste à ajouter un nombre aléatoire simulé à la moyenne linéaire en les covariables, ce nombre correspond à une réalisation du résidu. Une imputation déterministe revient à remplacer la donnée manquante par la moyenne de la loi (dans le cas du modèle linéaire la moyenne est une moyenne conditionnelle à l'observation de covariables).

On peut justifier une telle approche lorsque l'on dit que l'on souhaite effectuer une prédiction de l'estimateur de sondage non calculable. Fournir cette prédiction est optimale au sens que nous obtenons in fine l'estimateur le moins risqué⁴. De ce point de vue la prédiction de données individuelles ne donne pas le résultat que l'on souhaite lorsque l'estimateur de sondage est non linéaire en les variables d'intérêt⁵. L'estimateur de sondage d'une médiane, d'un quantile ou d'un indicateur d'inégalité sont des quantités non linéaires en les variables d'intérêt. Le choix d'une imputation aléatoire, autrement dit par simulation, a donc été retenu. En effet, produire l'estimateur qui correspondrait à la moyenne d'estimateurs de sondage calculés sur plusieurs jeux indépendants de données simulées approcherait la prédiction de l'estimateur de sondage. Afin de satisfaire la demande d'Eurostat nous avons uniquement procédé à une imputation aléatoire simple. Lorsque le taux de valeurs simulées est très inférieur à celui des valeurs renseignées et conservées, la différence est négligeable. La différence est également d'autant plus faible que nous utilisons des informations auxiliaires par conditionnement (informations en fourchettes et jusqu'à un certain point l'information fournie par des covariables).

Nous allons détailler désormais les différentes étapes des simulations des montants manquants. La démarche est de procéder en deux étapes. La première étape est une étape d'estimation d'un modèle. La seconde étape est une étape de simulation des variables d'intérêt manquantes conformément au modèle préalablement estimé.

Nous avons modélisé les montants comme issus de tirages dans une loi log-normale dont la moyenne du logarithme est linéaire en des covariables et avec un facteur d'échelle constant. Les montants sont en général observés soit en clair soit en tranches. Bien qu'il soit possible de faire une estimation par maximisation de la vraisemblance il nous a semblé préférable de ne mener l'estimation que sur le groupe des répondants en clair avec une procédure d'estimation robuste (dans l'état actuel de nos connaissances nous ne savions pas mener une estimation robuste utilisant ces deux informations). Nous faisons implicitement l'hypothèse, qu'au sein des deux groupes, la loi (conditionnelle à l'observation des covariables) du montant est la même. Il est possible d'imaginer la présence de biais de sélection, mais nous avons pris le parti de les négliger. Les biais de sélection sont d'autant plus faibles que nous avons utilisé le conditionnement par des observations de covariables.

Le phénomène le plus préoccupant dans l'enquête EU-SILC est la présence d'erreurs. Une seule erreur pourrait en théorie induire un biais « infini » avec des estimations de type maximum de vraisemblance ou MCO. Nous avons donc utilisé des méthodes de régressions robustes à une contamination inférieure à 50% par des erreurs. La comparaison de différentes méthodes disponibles sous SAS IML nous a fait opter pour le CALL LMS (Least Median Squares de P.J. Rousseeuw). En quelque sorte, les paramètres correspondent à ceux d'un modèle qui rassemblerait les 50 % des valeurs les plus crédibles de l'échantillon. L'estimation a été faite sur des sous-groupes lorsque cela était nécessaire (défaut d'additivité) et car, pour des problèmes de complexité algorithmique, le call LMS nécessite de restreindre drastiquement le nombre de covariables (à peu près 15 covariables dichotomisées). Cette étape d'estimation a été menée sur le sous-échantillon des réponses en clair qui n'étaient pas déclarées suspectes au vu de règles comptables prenant en compte les autres déclarations faites par le ménage.

Dans un second temps nous avons fait l'approximation que les estimateurs des paramètres sont les vrais paramètres et nous avons produit des valeurs simulées.

⁴ Pour le risque quadratique calculé avec l'aléa de modèle.

⁵ Dans certains cas seulement la différence peut être négligeable.

Si aucune information n'est disponible sur la valeur du montant manquant, la simulation est faite dans la loi log-normale. Lorsqu'une plage de valeurs est disponible, ou reconstituée à partir de la connaissance d'autres variables, la simulation se fait dans la loi précédente conditionnelle au fait que la variable est dans la plage de valeurs prescrite. Dans ce second cas la simulation revient à la simulation de lois normales tronquées. La simulation dans la loi normale tronquée est effectuée par inversion de la fonction de répartition dès que cela est possible. Il est arrivé parfois que des problèmes numériques (plage loin de la valeur prédite par le modèle...) nous fassent utiliser une méthode d'acceptation/rejet avec optimisation de la probabilité d'acceptation. Cette méthode est une variante de la méthode de la macro « résidus simulés » mais est beaucoup plus efficace⁶.

⁶ Elle revient à tirer conjointement dans une loi uniforme et dans une famille à un paramètre de lois exponentielles translatées en un point de la troncature ou dans une loi uniforme dans certains cas où il y a troncature à gauche et à droite. Le critère d'arrêt découle du résultat général du principe d'acceptation-rejet. L'optimisation sur le paramètre permet de maximiser la probabilité d'acceptation en 1 coup. Dans ce cas à peu près 5 itérations suffisent pour obtenir une simulation pour toutes les données manquantes sans avoir à envisager de traitement adapté en cas de « non convergence ». Du point de vu du temps de calcul la simulation est quasiment instantanée mais légèrement plus longue que celle basée sur l'inversion de la fonction de répartition.