

**Rapport intermédiaire  
sur la qualité des données  
françaises  
EU-SILC 2010**

## TABLE DES MATIÈRES

<b>Table des matières .....</b>	<b>2</b>
<b>1 Indicateurs transversaux communs de l'Union européenne .....</b>	<b>5</b>
1.1 Indicateurs communautaires transversaux communs fondés sur la dimension transversale des statistiques EU-SILC .....	5
1.1.1 Taux de pauvreté après transferts sociaux .....	5
1.1.2 Inégalité de la distribution des revenus Q80/Q20, ratio interquintile .....	11
1.1.3 Intensité de la pauvreté .....	11
1.1.4 Dispersion autour du seuil de pauvreté .....	11
1.1.5 Taux de pauvreté ancré à une date fixée (2006).....	12
1.1.6 Taux de pauvreté avant transferts .....	12
1.1.7 Inégalité de la distribution des revenus : coefficient de Gini .....	13
1.2 Autres indicateurs .....	13
1.2.1 Niveau de vie médian .....	13
1.2.2 Niveau de vie moyen .....	13
1.2.3 Écart de rémunération entre les sexes non ajusté .....	13
<b>2 Précision .....</b>	<b>14</b>
2.1 Plan d'échantillonnage .....	14
2.1.1 Type de plan d'échantillonnage .....	14
2.1.2 Unités d'échantillonnage.....	14
2.1.3 Critères de stratification et de sous-stratification.....	14
2.1.4 Taille de l'échantillon et critères d'attribution.....	18
2.1.5 Systèmes de sélection de l'échantillon .....	18
2.1.6 Répartition de l'échantillon au cours du temps.....	18
2.1.7 Renouvellement de l'échantillon : groupes de rotation.....	19
2.1.8 Pondération.....	19
2.1.9 Substitutions .....	26
2.2 Erreurs d'échantillonnage .....	27
2.3 Erreurs non dues à l'échantillonnage.....	30
2.3.1 Erreurs dans la base de sondage et la couverture .....	30
2.3.2 Erreurs de mesure et de traitements .....	31
2.3.3 Erreurs de non-réponse .....	35
2.4 Mode de collecte des données .....	41
2.5 Durée de l'entretien.....	42
<b>3 Comparabilité .....</b>	<b>47</b>
3.1 Concepts et définitions de base.....	47
3.2 Composantes du revenu .....	48

3.2.1	Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC	48
3.2.2	La source ou la procédure utilisée pour la collecte des variables du revenu.....	51
3.2.3	La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes.....	51
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).....	51
<b>4</b>	<b>Cohérence.....</b>	<b>53</b>
4.1	Comparaison des variables cibles avec des sources extérieures.....	53
4.1.1	Salaires et revenus assimilés (PY010N).....	53
4.1.2	Les allocations chômage (PY090N).....	54
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension au survivant (PY110N).....	55
4.1.4	Revenus sociaux (HY050N, HY060N et HY070N).....	56
4.1.5	Prestations familiales (HY050N),.....	56
4.1.6	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation). .....	57
4.1.7	Loyers.....	57
4.1.8	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N).....	58
4.1.9	Revenu disponible du ménage et niveau de vie des individus.....	58



## 1 INDICATEURS TRANSVERSAUX COMMUNS DE L'UNION EUROPÉENNE

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 par convention 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.

A partir des données de revenus 2007, soit pour la collecte ayant eu lieu en 2008, les données d'enquête ont été appariées avec les fichiers administratifs fiscaux et sociaux. Auparavant, les données de revenu étaient collectées par voie d'enquête. Le changement de collecte des revenus permet une amélioration de la qualité des données, un allègement de la charge des enquêtes et une meilleure cohérence avec l'enquête de référence de l'INSEE sur les revenus, soit l'enquête Revenus fiscaux et sociaux ERFS (enquête emploi<sup>1</sup> appariée avec les fichiers fiscaux et sociaux).

Ce changement de collecte permet de disposer en particulier d'une meilleure information sur les revenus du patrimoine qui étaient mal recensés par voie de collecte et une meilleure qualité pour les autres revenus. L'impact de l'appariement avec la source fiscale avait été évalué à plus ou moins 0,2 point sur le taux de pauvreté (cf. [Comparability Study](#), Laurence Dauphin) mais celui de l'appariement avec les données des organismes gestionnaires de prestations sociales (Caisse nationale d'allocations familiales, Caisse nationale d'allocation vieillesse, Caisse Centrale de Mutualité Sociale Agricole) n'a pas pu être évalué.

#### 1.1.1.1 Taux de pauvreté selon l'âge et le sexe en 2010

	Total			0-64			0-17	18+			18-64		
	Total	F	H	Total	F	H	Total	Total	F	H	Total	F	H
Pauvres	13,3	13,9	12,6	13,8	14,3	13,3	17,9	12,0	12,8	11,2	12,4	13,0	11,7
Non pauvres	86,7	86,1	87,4	86,2	85,7	86,7	82,1	88,0	87,2	88,8	87,6	87,0	88,3

	18-24			25-49			50-64			65+		
	Total	F	H									
Pauvres	23,5	24,1	22,9	11,8	12,9	10,8	8,3	8,6	8,1	10,6	12,0	8,7
Non pauvres	76,5	75,9	77,1	88,2	87,1	89,2	91,7	91,4	91,9	89,4	88,0	91,3

#### Effectifs non pondérés

	Total			0 - 64			0 - 17	18+			18 - 64		
	Total	F	H	Total	F	H	Total	Total	F	H	Total	F	H
Total	26531	13728	12803	22247	11376	10871	6161	20370	10658	9712	16086	8306	7780
Pauvres	3190	1728	1462	2796	1477	1319	1024	2166	1207	959	1772	956	816
Non pauvres	23341	12000	11341	19451	9899	9552	5137	18204	9451	8753	14314	7350	6964

	18 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	2293	1110	1183	8341	4310	4031	5452	2886	2566	4284	2352	1932
Pauvres	461	231	230	893	498	395	418	227	191	394	251	143

<sup>1</sup> Labour Force Survey

Non pauvres	1832	879	953	7448	3812	3636	5034	2659	2375	3890	2101	1789
-------------	------	-----	-----	------	------	------	------	------	------	------	------	------

Répartition de la population totale selon le sexe<sup>2</sup>

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

Répartition de la population totale selon l'âge et le sexe (% en ligne)

	0-17	0-64	18+	18-64	18-24	25-49	50-64	65+	Total
Total	21,5	83,4	78,5	61,9	8,7	33,6	19,5	16,6	100
F	20,5	81,3	79,5	60,9	8,3	32,9	19,7	18,7	100
H	22,6	85,5	77,4	62,9	9,2	34,3	19,4	14,5	100

Répartition de la population pauvre selon le sexe

Total	F	H
100,0	54,0	46,0

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	29,0	86,7	71,0	57,7	15,4	30,0	12,3	13,3	100
F	26,7	83,8	73,3	57,1	14,4	30,5	12,2	16,2	100
H	31,6	90,0	68,4	58,4	16,6	29,4	12,4	10,0	100

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	11,8	12,8	10,8	6,2	5,8	6,5	17,6	18,8	16,1	33,1	29,5	36,1
Non pauvres	88,2	87,2	89,2	93,8	94,2	93,5	82,4	81,2	83,9	66,9	70,5	63,9

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Pauvres	8,4	9,4	7,3	27,1	28,0	25,4
Non pauvres	91,6	90,6	92,7	72,9	72,0	74,6

Sont exclus du champ les personnes de moins de 18 ans au 1<sup>er</sup> 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	20595	10783	9812	10460	5072	5388	10135	5711	4424	1063	494	569
Pauvres	2171	1221	950	622	282	340	1549	939	610	318	129	189
Non pauvres	18424	9562	8862	9838	4790	5048	8586	4772	3814	745	365	380

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Total	5451	2783	2668	3621	2434	1187

<sup>2</sup> Les tableaux en % sont pondérés avec la variable rb050

Pauvres	398	225	173	833	585	248
Non pauvres	5053	2558	2495	2788	1849	939

*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 : inactifs
F	100	46,4	53,6	5,1	27,0	21,5
H	100	55,2	44,8	6,6	26,8	11,4
Total	100	50,6	49,4	5,8	26,9	16,7

*Erreur ! Liaison incorrecte. 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 : inactifs
F	100	21,1	78,9	11,7	20,0	47,1
H	100	33,1	66,9	22,1	18,1	26,7
Total	100	26,4	73,6	16,3	19,2	38,2

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	10,9	17,7	8,3	6,8	6,5
Non pauvres	89,1	82,3	91,7	93,2	93,5

	Avec enfant : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Pauvres	15,6	34,6	6,7	10,6	20,5	17,9
Non pauvres	84,4	65,4	93,3	89,4	79,5	82,1

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	11874	3192	3964	3070	1590
Pauvres	1047	514	272	188	73
Non pauvres	10827	2678	3692	2882	1517

	Avec enfant : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Total	14657	1579	2979	5124	3363	1562
Pauvres	1865	512	207	468	678	236
Non pauvres	12792	1067	2772	4656	2685	1326

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

	F	H	< 65	65+
Pauvres	17,9	17,3	18,9	15,7
Non pauvres	82,1	82,7	81,1	84,3

#### Effectifs non pondérés

	F	H	< 65	65+
Total	1945	1247	1825	1367
Pauvres	320	194	314	200
Non pauvres	1523	1002	1418	1107

#### 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
48,8	15,8	16,3	10,8	5,8

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
45,7	6,1	11,7	17,0	10,9	5,5

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

F	H	< 65	65+
57,6	42,4	61,1	38,9

#### 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
40,0	21,2	10,3	5,6	2,9

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
52,6	16,0	5,9	13,6	17,0	7,5

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

F	H	< 65	65+
58,6	41,4	65,4	34,6

#### 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	13,3	24,2	7,4
Non pauvres	86,7	75,8	92,6

**Erreur ! Liaison incorrecte. Effectifs non pondérés**

	Total	Locataires	Propriétaires ou logés à titre gratuit
Total	26531	8156	18375
Pauvres	3190	1919	1271
Non pauvres	23341	6237	17104

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100	34,8	65,2

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100	63,5	36,5

#### 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 ménage où les adultes ont travaillé toute l'année à temps plein.

	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	18,5	10,5	4,5	68,2	48,0	18,7	5,4
Non pauvres	81,5	89,5	95,5	31,8	52,0	81,3	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	1998	2895	3349	781	744	4472	8646
Pauvres	278	242	130	509	348	798	474
Non pauvres	1720	2653	3219	272	396	3674	8172

**Erreur ! Liaison incorrecte.**

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

Total	Sans enfant			Avec enfants			
	W=0	0<W<1	W=1	W=0	0<W<0,5	0,5<W<1	W=1
100	9,6	13,6	16,9	3,7	3,2	18,5	34,6

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

Total	Sans enfant			Avec enfants			
	W=0	0<W<1	W=1	W=0	0<W<0,5	0,5<W<1	W=1
100	13,4	10,7	5,7	18,8	11,4	26,1	14,0

1.1.1.6 Seuil de pauvreté

En 2009, le seuil de pauvreté s'élève à 12 034 euros par an. Une personne seule est pauvre si elle a un revenu disponible inférieur à 1003 euros par mois et un couple avec deux enfants de moins de 14 ans, s'il a un revenu disponible inférieur à 25 273 euros (soit 2 106 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 4,5 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 relatif 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			
20,2	19,7	20,3	18,0	20,8	20,5	21,3	22,4	22,6	22,1	14,0	13,4	14,3			

*Effectifs non pondérés*

Total			0 - 17				18+			18 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H			
26531	1728	1462	1024	26531	1207	959	1772	956	816	394	251	143			

**Erreur ! Liaison incorrecte.**

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
40% de la médiane	Pauvres	3,8	3,9	3,6
	Non pauvres	96,2	96,1	96,4
50% de la médiane	Pauvres	7,5	7,8	7,1
	Non pauvres	92,5	92,2	92,9
70% de la médiane	Pauvres	21,6	22,6	20,5
	Non pauvres	78,4	77,4	79,5

*Effectifs non pondérés*

		Total	F	H
Total		26531	13728	12803
40% de la médiane	Pauvres	800	432	368
	Non pauvres	25731	13296	12435
50% de la médiane	Pauvres	1744	948	796

	Non pauvres	24787	12780	12007
70% de la médiane	Pauvres	5354	2915	2439
	Non pauvres	21177	10813	10364

### 1.1.5 Taux de pauvreté ancré à une date fixée (2006)

Il s'agit du taux de pauvreté calculé en 2009, en prenant comme seuil celui de 2006, corrigé de l'inflation entre ces deux dates.

Compte tenu de la rupture de série sur les revenus 2007, cet indicateur n'est pas pertinent pour la France.

En effet depuis la collecte 2008 sur les revenus 2007, pour une grande partie des ressources des ménages, la collecte par voie d'enquête a été remplacée par un recours aux données administratives, déclarations fiscales et fichiers de prestations des organismes gestionnaires (Cnaf, MSA et Cnav). Par ailleurs, afin de se rapprocher de la source nationale ERFS (Enquête sur les revenus fiscaux et sociaux), des revenus financiers ont été imputés dans SRCV quand ceux-ci ne sont pas recensés par la source fiscale. Par ailleurs, le changement de mode de collecte permet également de disposer de revenus que les individus omettaient de déclarer lors de la collecte (il s'agit en particulier de revenus de faibles montants) ou sous-estimaient (revenus financiers et fonciers en particulier). Enfin, la collecte permet de continuer à recueillir des revenus non imposables.

Ces éléments conduisent à une augmentation du revenu disponible des ménages et à une rupture de série à partir de 2008 sur les données de revenu.

En conséquence, les résultats sur le taux de pauvreté ancré en 2006 ne peuvent être calculés pour la France de 2008 à 2010.

### 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	44,6	46,9	42,1	37,8	46,5	49,3	43,4	35,7	37,4	33,9	86,6	87,8	84,8		
Non pauvres	55,4	53,1	57,9	62,2	53,5	50,7	56,6	64,3	62,6	66,1	13,4	12,2	15,2		

#### Effectifs non pondérés

	Total			0 - 17			18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H		
Total	26531	13728	12803	6161	20370	10658	9712	16086	8306	7780	4284	2352	1932		
Pauvres	11329	6133	5196	2192	4093	5034	4103	5453	2980	2473	3684	2054	1630		
Non pauvres	15202	7595	7607	3969	16277	5624	5609	10633	5326	5307	600	298	302		

#### 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	25,2	25,8	24,5	36,4	22,1	23,0	21,1	24,5	25,4	23,5	13,1	15,0	10,5		
Non pauvres	74,8	74,2	75,5	63,6	77,9	77,0	78,9	75,5	74,6	76,5	86,9	85,0	89,5		

### Effectifs non pondérés

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Total	26531	13728	12803	6161	20370	10658	9712	16086	8306	7780	4284	2352	1932
Pauvres	6201	3301	2900	2108	4093	2243	1850	3603	1929	1674	490	314	176
Non pauvres	20330	10427	9903	4053	16277	8415	7862	12483	6377	6106	3794	2038	1756

#### 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 29,8 en 2009.

## 1.2 Autres indicateurs

### 1.2.1 Niveau de vie médian

En 2009, le niveau de vie médian des individus s'élève à 20 058 euros par unité de consommation, soit 1 671 euros par mois.

### 1.2.2 Niveau de vie moyen

En 2009, le niveau de vie des individus s'élève en moyenne à 23 532 euros par unité de consommation, soit 1 961 euros par mois.

### 1.2.3 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

Jusqu'à l'enquête 2009, le système d'échantillonnage de l'enquête SILC en France reposait exclusivement sur l'échantillon-maître (EM) issu des données du dernier recensement exhaustif de 1999 et la base de logements neufs (BLSN). Depuis 2004, la France est passée à une nouvelle méthode : le recensement rotatif continu. Désormais, les communes de moins de 10000 habitants ou « petites communes », sont recensées exhaustivement tous les cinq ans par roulement : pour cela, cinq groupes de rotation ont été définis aléatoirement, dans lesquels ont été réparties ces petites communes. Pour ce qui est des communes comprenant 10000 habitants ou plus, ou « grandes communes », elles font l'objet d'une enquête de recensement plus complexe, par sondage chaque année au taux moyen de 8%. Les adresses de ces communes sont réparties aléatoirement entre cinq groupes de rotation disjoints : chaque année, des logements appartenant à un échantillon d'adresse puisé dans le groupe de rotation « actif » sont recensés. Au total, cinq groupes de rotations ont ainsi été constitués.

Ce nouveau recensement conduit à une modification radicale des méthodes employées pour la construction des échantillons des enquêtes. Un nouvel échantillon-maître, OCTOPUSSE<sup>3</sup>, adapté au contexte original du nouveau recensement a donc été mis au point<sup>4</sup>.

Pour l'enquête SILC de 2010, cohabitent deux échantillons :

- Le premier, tiré à partir de l'ancien échantillon-maître et de la BLSN, concerne les individus déjà présents dans l'échantillon en 2009. L'ancien échantillon-maître était constitué par 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 ».
- Le second, tiré depuis le nouvel échantillon-maître, concerne seulement le sous échantillon entrant. L'échantillon-maître OCTOPUSSE est constitué à partir de deux listes de logements issues de la dernière enquête annuelle de recensement disponible. La première liste est celle relative aux logements situés dans une grande commune, la seconde celle relative aux logements localisés dans une petite commune.

#### 2.1.2 Unités d'échantillonnage

L'unité d'échantillonnage gérée dans ces bases de sondage est le logement. 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.

#### 2.1.3 Critères de stratification et de sous-stratification

Jusqu'en 2009, les critères de stratification et de sous-stratification étaient les suivants.

Le partage rural/urbain constitue une première stratification préalable au premier degré de tirage. Elle est constituée de cinq strates :

- communes rurales,
- 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 »),

---

<sup>3</sup> Organisation Coordinée de Tirages Optimisés Pour une Utilisation Statistique des Échantillons.

<sup>4</sup> M. Christine et S. Faivre « OCTOPUSSE : un système d'Échantillon-Maître pour le tirage des échantillons dans la dernière Enquête Annuelle de Recensement » *Journées de méthodologie statistique 2009*.

- unités urbaines de Paris

Les trois premières strates sont ventilées en sous-strates selon la région administrative. Comme il y a 22 régions administratives, la stratification préalable au premier degré comporte 66 sous-strates. Dans ces sous-strates, les unités primaires sont déterminées à partir d'un tirage à probabilités inégales proportionnellement à leur nombre de résidences principales dans la strate de tirage (Tirage équilibré sur des critères de revenu et d'âge au niveau de groupes de régions).

Pour les deux dernières strates (unités urbaines de plus de 100 000 habitants et unité urbaine de Paris), l'ensemble des unités primaires des strates sont conservées soit 52 pour la première et une seule pour l'unité urbaine de Paris.

Ensuite, il n'y a pas de stratification préalable au second degré en zone rurale et pour les unités urbaines de moins de 20 000 habitants. 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. Le tirage est stratifié par groupes de communes pour garantir la dispersion géographique des districts. Il s'agit d'un tirage à probabilités égales des districts dans l'UU et équilibré par la méthode du Cube sur les critères de revenu et d'âge.

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,

Depuis l'échantillon entrant 2006, les logements vacants sont distingués entre « rural » et « urbain », ce qui constitue cinq jeux de poids initiaux au lieu de quatre pour les échantillons des années précédentes.

Avec le nouvel échantillon-maître OCTOPUSSE, il n'est pas possible de réaliser une stratification des unités primaires par type d'espace (rural-urbain). Pour l'échantillon entrant en 2010, les critères de stratification et de sous-stratification sont donc différents de ceux des 8 autres sous-échantillons.

Le partage en régions administratives constitue la première stratification. Dans le cas de l'île de France, une sous-stratification permet de séparer la grande de la petite couronne. Les unités primaires sont les zones d'action enquêteurs (ZAE)<sup>5</sup>. Le tirage est également équilibré sur des *taux régionaux*: Il est nécessaire d'équilibrer non seulement au niveau de l'ensemble de la ZAE mais aussi de chacun des 5 groupes de rotation, de manière à avoir chaque année une base de sondage « représentative ».

Chaque grande commune (commune de plus de 10 000 habitants) constitue à elle seule une ZAE. Pour les petites communes, les ZAE doivent contenir au moins 300 résidences principales.

---

<sup>5</sup> Ces nouvelles unités primaires diffèrent donc des unités primaires constituées à partir de l'ancien recensement.

Synthèse sur la constitution de l' Echantillon-Maitre1999 (EM99) et le tirage d'un échantillon pour une enquête ménage

TIRAGE DE L'EM99	STRATE 0	STRATE 1	STRATE 2	STRATE 3	STRATE 4
Stratification préalable au premier degré $N_h = 66$	Régions administratives	Régions administratives	Régions administratives	-	-
Unités Primaires $N_{UP} = 3435$	Groupes de communes contiguës dans le même canton de 1800 à 3600 logements $N_{UP} = 2227$	Groupes d'unités urbaines de 1800 à 3600 logements $N_{UP} = 975$	Unités Urbaines $N_{UP} = 180$	Unités Urbaines $N_{UP} = 52$	Unité Urbaine de Paris $N_{UP} = 1$
Méthode de tirage des UP $n_{UP} = 349$ (8 super régions)	Allocations proportionnelles à la taille de l'UP dans la strate Equilibré par super région sur le revenu, les classes d'âge, le nombre d'UP à tirer par région $n_{UP} = 128$	Allocations proportionnelles à la taille de l'UP dans la strate Equilibré par super-région sur le revenu, les classes d'âge, le nombre d'UP à tirer par région $n_{UP} = 75$	Allocations proportionnelles à la taille de l'UP dans la strate Equilibré par super-région sur le revenu, les classes d'âge, le nombre d'UP à tirer par région $n_{UP} = 93$	Exhaustif $n_{UP} = 52$	Exhaustif $n_{UP} = 1$
Stratification préalable au second degré	-	-	Groupes de communes Au moins 1/6 de la taille de l'UP Tri sur la taille des communes $N_{grcom} = 209$	Groupes de communes (ou arrondissement) Au moins 100 districts Tri sur la taille des communes $N_{grcom} = 392$	Par département, groupes de communes (ou arrondissement) Au moins 100 districts Tri sur la taille des communes $N_{grcom} = 254$
Unités secondaires	-	-	Groupes de districts D'au moins 100 logements Tri sur la taille des districts dans le groupe de communes $N_{grdist} = 13642$	Districts $N_{dist} = 101919$	Districts $N_{dist} = 49276$
Méthode de tirage des US	-	-	30 groupes de districts par UP Allocations proportionnelles à la taille du groupe de communes dans l'UP Equilibré par UP sur le revenu, les classes d'âge, la taille en P et S, le nb de grdist à tirer par grcom $n_{grdist} = 2790$	Tirage au 1/20 des districts dans les groupes de communes Equilibré par UP sur le revenu, les classes d'âge, la taille en P et S, le nb de dist à tirer par grcom $n_{dist} = 5098$	Tirage au 1/20 des districts dans les groupes de communes Equilibré par dep sur le revenu, les classes d'âge, la taille en P et S, le nb de dist à tirer par grcom $n_{dist} = 2465$
Désignation des logements $n = 202889$ (7% du RP)	Exhaustif sur l'UP $n = 451776$ (6%)	Exhaustif sur l'UP $n = 444978$ (9%)	Exhaustif sur l'US $n = 497962$ (13%)	Exhaustif sur l'US $n = 399845$ (5%)	Exhaustif sur l'US $n = 228328$ (5%)

**Synthèse sur la constitution de l' Echantillon-Maitre1999 (EM99) et le tirage d'un échantillon pour une enquête ménage**

<i>TIRAGE D'UNE ENQUETE</i>	STRATE 0	STRATE 1	STRATE 2	STRATE 3	STRATE 4
Stratification préalable au second degré	Groupes de communes D'au moins 100 logements dans une UP ou fraction d'UP Tri sur la taille des communes $N_{grcom} = 1332$ -	Groupes de communes  Commune $N_{grcom} = 284$	<u>Groupes de communes de l'EM</u> D'au moins 1/6 de la taille de l'UP Tri sur la taille des communes $N_{grcom} = 209$	<u>Groupes de communes de l'EM</u> Groupe ou arrondissement d'au moins 100 districts Tri sur la taille des communes $N_{grcom} = 392$	<u>Groupes de communes de l'EM</u> Par département, groupe ou arrondissement d'au moins 100 districts Tri sur la taille des communes $N_{grcom} = 254$
Unités secondaires	Groupes de communes	Groupes de communes	<u>Groupes de districts de l'EM</u> Au moins 100 logements Tri sur la taille des districts dans le groupe de communes $N_{grdist} = 13642$	<u>Districts de l'EM</u>  District $N_{dist} = 101919$	<u>Districts de l'EM</u>  District $N_{dist} = 49276$
Méthode de tirage des US	Allocations par UP fonction de l'enquête considérée Proportionnelles à la taille du groupe de communes dans l'UP $n_{grcom}$ variable selon l'enquête	Exhaustif $n_{grcom} = 284$	30 groupes de districts par UP Allocations proportionnelles à la taille du groupe de communes dans l'UP Equilibré par UP sur le revenu, les classes d'âge, la taille en P et S, le nb de dist à tirer par grcom $n_{grdist} = 2790$	Tirage au 1/20 des districts dans les groupes de communes Equilibré par UP sur le revenu, les classes d'âge, la taille en P et S, le nb de dist à tirer par grcom $n_{dist} = 5098$	Tirage au 1/20 des districts dans les groupes de communes Equilibré par dep sur le revenu, les classes d'âge, la taille en P et S, le nb de dist à tirer par grcom $n_{dist} = 2465$
Stratification préalable au 3 <sup>ème</sup> degré de tirage	Catégorie de logements	Catégorie de logements	Catégorie de logements	Catégorie de logements	Catégorie de logements
Désignation des logements enquêtés	Systematique sur fichier trié	Systematique sur fichier trié	Systematique sur fichier trié	Systematique sur fichier trié	Systematique sur fichier trié

**Vocabulaire**

- o Le terme « taille » employé est le nombre de résidences principales recensées au niveau géographique considéré.
- o Définition des strates de gestion :
  - Strate 0 : Communes rurales au RP99
  - Strate 1 : Communes appartenant à des Unités Urbaines (UU) ayant moins de 20 000 habitants au RP99
  - Strate 2 : Communes appartenant à des UU ayant entre 20 000 et 100 000 habitants au RP99.
  - Strate 3 : Communes appartenant à des UU de plus de 100 000 habitants au RP99, sauf l'UU de PARIS.
  - Strate 4 : UU de PARIS

#### 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 entrants des années 2005 à 2010 en comprenaient chacun 3 000.

#### 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 qui sont tirés au taux  $f/2$  jusqu'en 2005 (à partir de 2006 taux  $f$  pour les logements vacants urbains et  $f/18$  pour les logements vacants ruraux). Le nouvel échantillon-maître permet de bénéficier de la « fraîcheur » des informations fournies par les enquêtes annuelles de recensement. Il permet notamment de minimiser le nombre « de transformations » d'une résidence principale en résidence secondaire.

#### **Pour les individus déjà présents en 2009 :**

##### **Cas de la strate rurale**

Le tirage des 128 unités primaires de l'échantillon-maître parmi les 2 227 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 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.

#### **Pour l'échantillon entrant en 2010 :**

Le tirage des logements dans les ZAE est effectué de la manière suivante : la « base de sondage utile » de la ZAE est constituée de l'ensemble des logements de la ZAE appartenant à la dernière campagne de recensement et échantillonnables. Ces logements sont triés selon 5 critères : département, commune, nombre d'occupants, type d'habitation, position professionnelle. Les logements sont ensuite tirés par tirage systématique à probabilités égales.

#### 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. Par construction, les tirages des sous-échantillons étaient toujours effectués dans les mêmes unités primaires jusqu'en 2009.

### 2.1.7 Renouvellement de l'échantillon : groupes de rotation

Pour la collecte de 2010, 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 2009 au moins une personne répondante et n'appartenant pas à l'échantillon sortant : Il comprenait 10 554 logements.

### 2.1.8 Pondération

La pondération se fait au niveau logement.

En 2010, nous avons estimé pour la cinquième année 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  $W_j$ , 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-échantillons, les pondérations initiales  $W_j$  varient 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 2009

Soit  $\Phi_l(t, k)$  la probabilité de répondre en première interrogation ( $k$  étant le ssech en  $k^{\text{ième}}$  interrogation,  $k=1$  ici) des ménages du logement  $l$ , c'est-à-dire en  $t - k + 1$ . Elle est estimée chaque année pour l'échantillon entrant en utilisant les données de l'échantillon. On estime un modèle *logit* sur tous les ménages entrants dans le champ de l'enquête ; puis afin de limiter la dispersion des poids, des taux de non-réponse sont calculés pour des *catégories* homogènes définies par le croisement de variables apparaissant comme les plus discriminantes.

On attribue à tous les individus appartenant à un ménage répondant le poids corrigé de la non-réponse initiale suivant :

$$W_i^{*t} = \frac{\sum_{\substack{j \in l \\ j \in r_{i,k}}} \frac{W_j(t, k)}{\Phi_j(t, k)}}{\sum_{j \in l} L_i(t)} = \frac{W_l(t, k)}{\Phi_l(t, k)} \times \frac{Npan_l(t)}{\sum_{j \in l} L_i(t)}$$

- où  $W_l(t, k)$  est le poids de sondage brut du logement<sup>6</sup>  $l$  du sous-échantillon l'année  $t$  en  $k^{\text{ième}}$  interrogation.

<sup>6</sup> Il s'agit du poids de tirage multiplié par 9

- $N_{pan_l}(t)$  est le nombre d'individus panel du logement  $l$  en  $t$ .
- $L_i(t)$  est le nombre d'année d'appartenance au champ ou le nombre de sous-échantillon dans lequel l'individu  $i$  (panel comme non panel) aurait pu être tiré pour appartenir à  $S_t$ .

Tous les membres d'un même logement ont donc le même poids.

Les modèles de non-réponse ont été estimés sur les logements de l'échantillon qui faisaient partie du champ en 2009 : 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 2009 :

- les résidences principales au dernier recensement de la population (RP) et en 2009,
- 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 2009,
- 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 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, 2006, 2007, 2008 et 2009 : 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).

#### **La correction de la non-réponse en réinterrogation**

La non-réponse est définie au niveau du ménage. Nous calculons cependant la probabilité de réponse en réinterrogation au niveau des individus parce que c'est la seule unité stable dans le temps. Nous ne l'estimons que pour les individus panel parce que nous ne disposons d'information homogène que pour ces individus. Les individus non panel ont par ailleurs un poids de base nul.

Une probabilité de réponse est donc attribuée à chaque individu panel qui est encore dans le champ en 2010. Est considéré comme répondant tout individu qui appartient à un ménage répondant (DB135='1').

Soit  $\theta_i(t, k)$  la probabilité de répondre en  $t$  sachant qu'on appartient à l'échantillon  $a_{t,k}$  et qu'on a répondu en  $t - k + 1$ .

Cette probabilité est estimée pour tous les individus panel  $i \in a_{t,k} \cap \Omega_t$ .

Si  $k = 1$ ,  $\theta_i(t, k) = 1$ .

Pour les adultes,  $\theta_i(t, k)$  est estimée en utilisant les informations individuelles et ménage collectées en  $t - k + 1$ , le fait d'avoir déménagé, d'avoir été dans un ménage éclaté entre  $t - k + 1$  et  $t$ .

Jusqu'en 2006, on attribuait aux enfants âgés de moins de 16 ans en  $t - k + 1$  la proportion d'enfants en  $t - k + 1$  de l'échantillon  $a_{t,k}$  répondants en  $t$  comme probabilité de réponse. Puis on attribuait aux enfants panel nés après la collecte de  $t - k + 1$  le poids de leur mère.

A partir de 2007, on attribue aux enfants la probabilité de réponse de leur mère ou à défaut de leur père : en effet, par définition la « non réponse » n'existe pas pour eux car le questionnaire individuel n'est pas posé aux enfants : le fait de disposer de réponse aux questions sur les enfants dépend donc de la bonne volonté des parents.

Enfin, les individus non panels entrants dans un ménage panel se voient attribuer un taux de réinterrogation correspondant à la moyenne observée du sous-échantillon : par construction, ils ne sont pas au même rang d'interrogation que les individus panel du ménage interrogé mais l'on considèrera que le taux de réponse des individus panels influence celui des non panels entrants dans le ménage.

Si  $t - 8 \geq 2004$ , il faut donc estimer 8 modèles de réponse en réinterrogation. En 2010, il faut en estimer 6.

On attribue à chaque individu le poids corrigé de la non-réponse suivant :

$$W_i^{*t} = \frac{\sum_{\substack{j \in I \\ j \in r_{t,k}}} \frac{W_j(t, k)}{\Phi_l(t, k) \times \theta_j(t, k)}}{\sum_{j \in I} L_i(t)} = \frac{W_l(t, k)}{\Phi_l(t, k)} \times \frac{\sum_{\substack{j \in I \\ j \in r_{t,k}}} \frac{1}{\theta_j(t, k)}}{\sum_{j \in I} L_i(t)}$$

$$= \frac{W_l(t, k)}{9} \times \frac{1}{\Phi_l(t, k)} \times \frac{1}{Npan_l(t)} \sum_{\substack{j \in I \\ j \in r_{t,k}}} \frac{1}{\theta_j(t, k)} \times \frac{9 \times Npan_l(t)}{\sum_{j \in I} L_i(t)}$$

Soit en 2010 :

$$W_i^{*2010} = \frac{W_l(2010, k)}{\Phi_l(2010, k)} \times \frac{1}{\sum_{\substack{j \in I \\ j \in r_{t,k}}} \theta_j(2010, k)}$$

$$\times \frac{1}{3 \times Nchamp_l^{2004}(2010) + Nchamp_l^{2005, 2009}(2010) + Npers_l(2010)}$$

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  $t$  et du partage des poids.

En régime de croisière, le calage transversal nécessite donc d'estimer chaque année 9 modèles sur des échantillons qui peuvent devenir restreints.

Pour éviter d'estimer les modèles sur des champs restreints, on peut estimer un modèle sur tous les individus qui répondent en  $k^{\text{ième}}$  interrogation depuis 2004.

En 2010 par exemple, on estime six (2010-2004) modèles sur tous les individus qui répondent en  $k^{\text{ième}}$  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
  - o sur les entrants 2007 encore dans le champ 2008
  - o sur les entrants 2008 encore dans le champ 2009
  - o sur les entrants 2009 encore dans le champ 2010
- 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
  - o sur les entrants 2006 encore dans le champ 2008
  - o sur les entrants 2007 encore dans le champ 2009
  - o sur les entrants 2008 encore dans le champ 2010
- un modèle pour les individus interrogés en 4ème vague :
  - o sur les entrants 2004 réinterrogés en 2007
  - o sur les entrants 2005 réinterrogés en 2008
  - o sur les entrants 2006 réinterrogés en 2009
  - o sur les entrants 2007 réinterrogés en 2010
- un modèle pour les individus interrogés en 5ème vague :
  - o sur les entrants 2004 réinterrogés en 2008
  - o sur les entrants 2005 réinterrogés en 2009
  - o sur les entrants 2006 réinterrogés en 2010
- un modèle pour les individus interrogés en 6ème vague :
  - o sur les entrants 2004 réinterrogés en 2009
  - o sur les entrants 2005 réinterrogés en 2010
- un modèle pour les individus interrogés en 7ème vague :
  - o sur les entrants 2004 réinterrogés en 2010

Cela suppose une stabilité des comportements dans le temps (ce qui effectivement est observé jusqu'ici) et permettrait d'améliorer au fur et à mesure des vagues les modèles de non-réponse (d'année en année les modèles par rang d'interrogation sont sensiblement identiques).

*Probabilité de répondre en Nième vague sachant qu'on a répondu à la première*

Les modèles de non-réponse sont estimés sur les adultes réinterrogés en  $n^{\text{ème}}$  vague. L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont observés parmi les variables suivantes :

- 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,
- avoir un contrat à durée indéterminée ou non,
- le fait qu'il y ait eu un ou plusieurs départs dans le ménage depuis la dernière vague,
- la nationalité (magrébine/autre),

À partir de ces résultats, différentes sous-populations de répondants sont formés en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif sont regroupés pour assurer la robustesse des résultats). Au final, les adultes réinterrogés en n<sup>ème</sup> vague sont répartis en plusieurs groupes et nous supposons le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité  $PR(n^{ème} \text{ vague})$  vaut donc :

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

### **Partage des poids**

Le partage des poids est une étape spécifique ayant pour but dans la pondération transversale de recalculer un poids pour les ménages dans lesquels se sont ajoutés un ou plusieurs nouveaux membres après la sélection de l'échantillon. Le poids doit être alloué de façon à ce que l'échantillon demeure représentatif de la population transversale. Pour ce faire, il faut déterminer le statut de chaque personne dans l'échantillon transversal selon que la personne est un individu longitudinal ou un cohabitant.

Un cohabitant est un individu non panel qui se joint à un sous-échantillon après sa sélection initiale. Les poids de base des individus non-panel sont donc nuls.

Le principe se traduit par le partage entre tous les membres adultes du ménage de la somme des poids de base des adultes (personnes de 17 ans et plus), que ces derniers soient individus panel ou individus non-panel, et selon le nombre d'années que l'individu a appartenu au champ d'enquête soit, dans le cadre de cette enquête, résidant en ménage ordinaire et en métropole (c'est-à-dire selon le nombre d'années où il aurait pu être tiré dans un sous-échantillon entrant).

En faisant l'hypothèse que l'individu panel n'apparaît que dans un seul sous-échantillon<sup>7</sup>, la formule est la suivante :

$$W_i^t = \frac{W_l(t, k) \times Npan_l(t)}{\sum_{j \in l} L_j(t)} \text{ où}$$

- $W_l(t, k)$  est le poids de sondage brut du logement<sup>8</sup>  $l$  du k<sup>ième</sup> sous-échantillon l'année  $t$ .
- $Npan_l(t)$  est le nombre d'individus panel du logement  $l$  en  $t$ .
- $L_i(t)$  est le nombre d'année d'appartenance au champ ou le nombre de sous-échantillon dans lequel l'individu  $i$  (panel comme non panel) aurait pu être tiré pour appartenir à  $S_l$ .

<sup>7</sup> Cependant il peut y avoir des cas où l'individu panel a été tiré deux fois ou plus sur neuf années consécutives : du fait des déménagements essentiellement mais aussi des changements de base de sondage, cela reste possible mais jusqu'ici nous n'avons jamais rencontré de cas.

<sup>8</sup> Il s'agit du poids de tirage multiplié par 9

L'année initiale de l'enquête, soit en 2004, il n'y a pas de partage de poids puisque, par définition, il n'y a que des individus panel.

Ensuite en 2005, le calcul de  $L_i(t)$  est :

$$L_i(t) = 8 \times Nchampi^{2004}(2005) + Npers_i(2005)$$

où  $Npers_i(2005)$  est le nombre total d'individus dans le logement  $l$  en 2005 et  $Nchampi^{2004}(2005)$  est le nombre d'individus habitant le logement  $l$  en 2005 qui étaient déjà dans le champ en 2004.

$$\text{En 2006, } L_i(t) = 7 \times Nchampi^{2004}(2006) + Nchampi^{2005}(2006) + Npers_i(2006)$$

$$\text{En 2007, } L_i(t) = 6 \times Nchampi^{2004}(2007) + Nchampi^{2005,2006}(2007) + Npers_i(2007)$$

Où  $Nchampi^{2005,2006}(2007)$  est la somme des années passées dans le champ d'enquête entre 2005 et 2006 par les individus appartenant au logement  $l$  en 2007.

$$\text{En 2008, } L_i(t) = 5 \times Nchampi^{2004}(2008) + Nchampi^{2005,2007}(2008) + Npers_i(2008)$$

$$\text{En 2009, } L_i(t) = 4 \times Nchampi^{2004}(2009) + Nchampi^{2005,2008}(2009) + Npers_i(2009)$$

$$\text{En 2010, } L_i(t) = 3 \times Nchampi^{2004}(2010) + Nchampi^{2005,2009}(2010) + Npers_i(2010)$$

$$\text{En 2011, } L_i(t) = 2 \times Nchampi^{2004}(2011) + Nchampi^{2005,2010}(2011) + Npers_i(2011)$$

$$\text{A partir de 2012, } L_i(t) = Nchampi^{t-8,t-1}(t) + Npers_i(t)$$

En l'absence d'informations complètes sur la présence dans le champ ou non, pour les échantillons entrants  $L_i(t)$  serait automatiquement = 9. Dans le cas d'informations sur la présence dans le champ ou non les années précédentes (par exemple nombre d'années passées à l'étranger ou dans les DOM-TOM),  $L_i(t)$  pourra varier.

Pour appliquer la formule, il est donc nécessaire de comptabiliser le nombre d'années de présence de l'individu dans le champ. Pour cela, on prendra en compte l'année de naissance, la date d'arrivée sur le territoire français et le nombre d'années passées à l'étranger ou dans les DOM-TOM.

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

L'enquête, réalisée en mai-juin 2010, est, comme beaucoup d'enquêtes auprès des ménages de l'Insee, calée sur les marges issues de l'enquête Emploi<sup>9</sup> de 2009 pour une représentativité transversale. En effet, toutes les enquêtes ménages dont la collecte a lieu pendant l'année civile N sont calées sur la dernière enquête Emploi disponible au début de l'année N, soit 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.

<sup>9</sup> Labour Force Survey

Le calage a été effectué 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 31/12/2009.
- 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 ;

#### 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).

Les pondérations finales varient dans un rapport de 1 à 96 après prise en compte de l'ensemble des opérations de redressement de la non réponse et de calage réalisés.

#### *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. Jusqu'en 2009, ceux-ci pouvaient être linéarisés puis estimés en utilisant le logiciel Poulpe développé à l'INSEE : ce logiciel était conçu pour l'estimation des intervalles de confiance des enquêtes tirées dans l'échantillon maître de 1999. A partir de 2010, les échantillons entrants sont tirés dans la base de sondage conçue à partir du recensement en continu de la population (cf. 2.1). La cohabitation de données issues de deux bases de sondage différentes complexifie les calculs de précision et ne permet plus l'utilisation du logiciel Poulpe. Nous ne sommes pas en mesure de fournir cette année ces écarts-type : des investissements théoriques et méthodologiques sont en cours.

Indicateurs	Valeur2009	Valeur 2008	Écart-type 2008	Coefficient de variation 2008	Effet de sondage 2008	Effectif atteint 2008	Taille effective de l'échantillon 2008
Taux de pauvreté après prestations sociales (en pourcentage)							
Total	<b>13,5</b>	12.9	0.4	3.1	1.18	25611	21704
Hommes total	<b>12,8</b>	12.0	0.4	3.3	1.15	12327	10719
Femmes total	<b>14,1</b>	13.7	0.4	2.9	1.18	13284	11258
0-17 ans	<b>18,4</b>	17.3	0.8	4.6	1.11	6081	5478
18-24 ans	<b>24,5</b>	21.2	1.0	4.7	1.18	2240	1898
25-49 ans	<b>12,3</b>	11.2	0.4	3.6	1.25	8156	6525
50-64 ans	<b>8,4</b>	9.0	0.4	4.4	1.03	5086	4938
65+ ans	<b>9,7</b>	10.7	0.6	5.6	1.39	4048	2912
18+ ans	<b>12,1</b>	11.6	0.3	2.6	1.23	19530	15878
18-64 ans	<b>12,8</b>	11.9	0.3	2.5	1.19	15482	13010
0-64 ans	<b>14,2</b>	13.3	0.4	3.0	1.15	21563	18750
Hommes 18-24 ans	<b>23,5</b>	20.6	1.3	6.3	1.18	1146	971
Hommes 25-49 ans	<b>11,2</b>	10.2	0.5	4.9	1.31	3915	2989
Hommes 50-64 ans	<b>8,1</b>	8.5	0.5	5.9	1.05	2419	2304
Hommes 65+ ans	<b>8,0</b>	9.1	0.7	7.7	1.28	1810	1414
Hommes 18+ ans	<b>11,3</b>	10.8	0.3	2.8	1.24	9290	7492
Hommes 18-64 ans	<b>12,0</b>	11.2	0.4	3.6	1.23	7480	6081
Hommes 0-64 ans	<b>13,6</b>	12.5	0.4	3.2	1.15	10517	9145
Femmes 18-24 ans	<b>25,5</b>	21.9	1.3	5.9	1.24	1094	882
Femmes 25-49 ans	<b>13,3</b>	12.1	0.4	3.3	1.19	4241	3564
Femmes 50-64 ans	<b>8,7</b>	9.5	0.5	5.3	1.05	2667	2540
Femmes 65+ ans	<b>10,8</b>	11.9	0.7	5.9	1.32	2238	1695
Femmes 18+ ans	<b>12,9</b>	12.4	0.3	2.4	1.19	10240	8605
Femmes 18-64 ans	<b>13,5</b>	12.6	0.4	3.2	1.23	8002	6506
Femmes 0-64 ans	<b>14,8</b>	14.1	0.4	2.8	1.15	11046	9605

Indicateurs	Valeur 2009	Valeur 2008	Écart-type 2008	Coefficient de variation 2008	Effet de sondage 2008	Effectif atteint 2008	Taille effective de l'échantillon 2008
Taux de pauvreté après prestations sociales (en pourcentage)							
Actifs occupés	<b>6,6</b>	6.7	0.3	4.5	1.17	10323	8823
Chômeurs	<b>33,5</b>	37.7	2.0	5.3	1.29	873	677
Retraités	<b>4,9</b>	8.7	0.4	4.6	1.38	5087	3686
Autres inactifs	<b>29,9</b>	26.6	1.2	4.5	1.11	2876	2591
Hommes, actifs occupés	<b>6,9</b>	7.1	0.4	5.6	1.18	5351	4535
Hommes, chômeurs	<b>36,3</b>	42.4	2.3	5.4	1.22	412	338
Hommes, retraités	<b>4,8</b>	8.1	0.5	6.2	1.25	2507	2006
Hommes, autres inactifs	<b>28,7</b>	26.6	1.2	4.5	1.11	2876	2591
Femmes, actives occupées	<b>6,3</b>	6.1	0.4	6.6	1.14	4972	4361
Femmes, chômeuses	<b>30,1</b>	33.4	2.5	7.5	1.37	461	336
Femmes, retraitées	<b>5,1</b>	9.3	0.5	5.4	1.28	2580	2016
Femmes, autres inactives	<b>30,4</b>	27.4	1.5	5.5	1.15	2051	1783
Personnes seules, < 65 ans	<b>19,1</b>	16.9	0.8	4.7	1.21	1717	1419
Personnes seules, 65+ ans	<b>13,2</b>	15.5	0.9	5.8	1.18	1307	1108
Personnes seules, hommes	<b>16,6</b>	15.5	0.8	5.2	1.18	1190	1008
Personnes seules, femmes	<b>16,9</b>	16.9	0.8	4.7	1.21	1834	1516
Personnes seules, total	<b>16,8</b>	16.3	0.6	3.7	1.25	1834	1467
2 adultes, sans enfant, tous les deux < 65	<b>8,7</b>	6.9	0.5	7.2	1.26	3860	3063
2 adultes, sans enfant, au moins 1 65+	<b>6,7</b>	7.6	0.7	9.2	1.5	2878	1919
Autres ménages sans enfant	<b>6,5</b>	7.2	1.3	18.1	0.93	1542	1658
Familles monoparentales	<b>35,7</b>	29.4	2.1	7.1	1.14	1536	1347
2 adultes, 1 enfant	<b>7,2</b>	5.6	0.9	16.1	1.10	2838	2580
2 adultes, 2 enfants	<b>10,8</b>	10.6	1.0	9.4	1.21	5028	4155
2 adultes, 3+ enfants	<b>21,1</b>	20.9	2.2	10.5	1.13	3285	2907
Autres ménages avec enfants	<b>18,3</b>	24.9	5.3	21.3	1.20	1523	1269
Ménages sans enfant	<b>10,6</b>	10.1	0.3	3.0	1.26	11304	8971
Ménages avec enfants	<b>15,9</b>	15.1	0.6	4.0	1.12	14210	12688
Propriétaires	<b>34,9</b>	8.2	0.4	4.9	1.27	17657	13903
Locataires	<b>65,1</b>	21.9	0.8	3.7	1.24	7954	6415
Ménages sans enfant, w = 0	<b>18,3</b>	16.9	1.0	5.9	1.02	1826	1790
Ménages sans enfant, 0 < w < 1	<b>10,9</b>	10.7	0.7	6.5	1.23	2771	2253
Ménages sans enfant, w = 1	<b>4,8</b>	4.4	0.4	9.1	1.13	3308	2927
Ménages avec enfants, w = 0	<b>66,0</b>	68.3	3.6	5.3	1.10	709	645
Ménages avec enfants, 0 < w < 0,5	<b>49,5</b>	52.3	4.8	9.2	1.19	664	558
Ménages avec enfants, 0,5 < w < 1	<b>19,5</b>	19.8	1.5	7.6	1.24	4315	3480
Ménages avec enfants, w = 1	<b>5,9</b>	5.9	0.7	11.9	1.13	8583	7596
Médiane des niveaux de vie (en euros par unité de consommation)	<b>20046</b>	19760	59.1	0.3	1.26	25611	20326
Seuil de pauvreté - personnes seules (en euros)	<b>12028</b>	11856	0.6	0.0	1.25	25611	20489

Seuil de pauvreté - 2 adultes, 2 enfants (en euros)	25259	24898	1.0	0.0	1.21	25611	21166
---	-------	-------	-----	-----	------	-------	-------

Indicateurs	Valeur 2009	Valeur 2008	Écart-type	Coefficient de variation	Effet de sondage	Effectif Atteint	Taille effective de l'échantillon
Inégalité de la distribution des revenus : rapport Q80/Q20	4,5	4.4	0.1	2.3	1.08	25611	23714
Écart de risque de pauvreté (en %)							
Total	19,5	18.4	0.8	4.3	1.22	3189	2614
Hommes total	19,5	18.5	1.0	5.4	1.21	1431	1183
Femmes total	19,3	18.3	0.9	4.9	1.21	1758	1453
0-17 ans	17,2	17.7	1.4	7.9	1.22	1019	835
18-64 ans	21,9	19.7	0.9	4.6	1.19	1909	1604
65+ ans	11,5	14.7	1.2	8.2	1.51	403	267
18+ ans	20,3	18.6	0.7	3.8	1.20	2170	1808
Hommes, 18-64 ans	21,7	19.6	1.1	5.6	1.17	871	744
Hommes, 65+ ans	10,9	15.5	1.4	9.0	1.27	150	118
Hommes, 18+ ans	20,8	18.8	0.9	4.8	1.22	958	785
Femmes, 18-64 ans	21,9	19.8	0.9	4.5	1.14	959	841
Femmes, 65+ ans	12,1	14.5	1.3	9.0	1.50	253	169
Femmes, 18+ ans	20,1	18.8	0.8	4.3	1.2	1212	1010
Taux de pauvreté selon le seuil de pauvreté (en %)							
40%	3,7	0.2	7.4	1.33	25611	19256	2.7
60%	7,4	0.3	4.9	1.28	25611	20009	6.1
70%	21,6	0.4	2.0	1.14	25611	22466	19.8
Taux de pauvreté avant prestations sociales autres que retraites (en %)							
Total	25,0	23.8	0.4	1.7	1.17	25611	21890
Homes total	24,4	22.9	0.4	1.7	1.16	12627	10885
Femmes total	25,6	24.6	0.4	1.6	1.15	13284	11551
0-17 ans	36,5	34.8	0.9	2.6	1.14	6081	5334
18-64 ans	24,5	22.6	0.4	1.8	1.16	15482	13347
65+ ans	12,4	13.6	0.6	4.4	1.49	4048	2717
18+ ans	24,5	20.7	0.3	1.4	1.20	19530	16275
Hommes, 18-64 ans	23,5	21.5	0.4	1.9	1.18	7480	6339
Hommes, 65+ ans	10,2	11.4	0.7	6.1	1.39	1810	1302
Hommes, 18+ ans	23,5	19.6	0.4	2.0	1.23	9290	7553
Femmes, 18-64 ans	25,5	23.6	0.4	1.7	1.14	8002	7019
Femmes, 65+ ans	14,0	15.2	0.7	4.6	1.42	2238	1576
Femmes, 18+ ans	22,8	21.7	0.3	1.4	1.16	10240	8828
Coefficient de Gini (en pourcentage)	29,9	29.8	0.4	1.3	1.05	25611	24391
Niveau de vie moyen (en euros par unité de consommation)	26519	23336				25611	

<b>Composantes du revenu disponible</b>	Moyenne par ménage 2009 (en euros)	Moyenne par ménage 2008 (en euros)	Écart-type 2008	Coefficient de variation 2008	Effet de Sondage 2008	Effectif Atteint 2008	Taille effective de l'échantillon 2008
Revenu disponible total (HY020)	36437	36411	253	0.7	1.16	10603	9141
Revenus tirés de la location d'une propriété ou de terres (HY040N)	1015	964	47	4.9	1.08	10603	9818
Allocation famille/enfant ( HY050N)	932	911	11	1.2	1.12	10603	9467
Exclusion sociale non classée ailleurs (HY060N)	280	249	14	5.6	1.09	10603	9728
Aides au logement (HY070N)	472	505	8	1.6	1.21	10603	8763
Transferts interménages perçus régulièrement en espèces (HY080N)	162	164	9	5.5	1.15	10603	9220
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	3474	3863	148	3.8	1.14	10603	9301
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	6	1	16.7	0.7	10603	15147
Impôts réguliers sur la fortune (HY120N)	50	76	14	18.4	0.98	10603	10819
Transferts interménages versés régulièrement en espèces (HY130N)	269	244	13	5.3	1.05	10603	10098
Remboursements consécutifs à des ajustements fiscaux (HY145N)	1834	1816	53	2.9	0.94	10603	11280
Salaire ou revenu assimilé (PY010N)	19530	19536	154	0.8	1.28	10603	8284
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	1777	1915	133	6.9	0.93	10603	11401
Allocations chômage(PY090N)	1051	896	33	3.7	1.00	10603	10603
Retraites (PY100N)	9104	8878	73	0.8	1.11	10603	9552
Allocations au survivant (PY110N)	67	65	7	10.8	1.12	10603	9467
Indemnités de maladie (PY120N)	240	226	13	5.8	1.20	10603	8836
Pensions d'invalidité (PY130N)	445	329	17	5.2	1.19	10603	8910
Allocations d'éducation (PY140N)	38	41	3	7.3	1.10	10603	9639

## 2.3 Erreurs non dues à l'échantillonnage

### 2.3.1 Erreurs dans la base de sondage et la couverture

#### Pour les individus déjà présents en 2009

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.

### **Pour l'échantillon entrant en 2010**

La base de sondage annuelle OCTOPUSSE est constituée à partir des listes de logements suivantes issues de la dernière Enquête Annuelle de Recensement disponible :

- Dans les ZAE grandes communes (ZAEGC) tirées (ou exhaustives), liste des logements de la commune recensés lors de la dernière Enquête Annuelle de Recensement (environ 8% des logements de la commune)
- Dans les ZAE petites communes (ZAEPC) tirées, liste des logements du recensement exhaustif des communes appartenant à la fraction recensée de la ZAEPC (commune de la ZAEPC appartenant au groupe de rotation impacté par la dernière Enquête Annuelle de Recensement disponible).

Toutefois, dans les grandes communes, une phase intermédiaire est nécessaire. En effet :

- la 1<sup>ère</sup> phase RP (affectation des adresses aux groupes de rotation) conduit à une affectation inégale et « semi-déterministe » (et non à probabilités égales 1/5) des adresses dans les différents groupes de rotation (notamment les grandes adresses) : il est donc nécessaire de reconstituer une « pseudoprobabilité » d'affectation.
- lors de la 2<sup>ème</sup> phase RP (tirage de logements à recenser dans le groupe de rotation annuel en grande commune), il y a une surreprésentation des « adresses neuves » et des « grandes adresses » (recensées d'office dans chaque groupe de rotation d'adresses).

Ainsi, la probabilité qu'un logement soit recensé varie suivant le type d'adresse (grande, neuve ou autre) : cela nécessite donc une opération statistique préalable (**rééchantillonnage** des logements situés en grandes adresses et en adresses neuves) pour disposer d'une base effective de logements à poids identiques. En pratique, l'objectif du rééchantillonnage est de ne conserver dans la base de sondage qu'une partie des logements recensés en grandes adresses et en adresses neuves de manière à éliminer la surreprésentation de ces strates par le Recensement.

### *2.3.2 Erreurs de mesure et de traitements*

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

Jusqu'en 2007, l'ensemble des données était récupéré par voie d'enquête. Pour limiter les erreurs de mesures (cf. rapports intermédiaires des années 2004 à 2007) et gagner en qualité sur la mesure des revenus individuels dans SRCV, les données de revenus (imposables) et les prestations sociales sont désormais récupérées par appariement avec les données fiscales et sociales (DGFIP, CNAF, CNAV et CCMSA), sur le modèle de l'Enquête Revenus Fiscaux et sociaux en France.

L'appariement n'est cependant pas exhaustif : du fait que les appariements se font à partir des adresses, les jeunes adultes âgés de 18 à 25 ans faisant leur déclaration de revenus sur celles de leurs parents à une adresse autre que celle de la collecte ne peuvent être retrouvés (pour ces jeunes, le questionnaire restera sous la même forme que les années précédentes afin de récupérer les données de revenu). De la même façon, les personnes ayant récemment déménagé peuvent être difficilement retrouvées. Pour ces cas de non appariements (environ 4% des individus), les erreurs de mesures constatées les années précédentes restent (confusion euros/francs, confusion entre les différentes périodes de référence dans l'enquête...).

Enfin, en ce qui concerne les revenus, seuls les revenus imposables sont récupérés par appariement. Par conséquent, le questionnaire comporte encore quelques questions sur les revenus exonérés d'impôt sur le revenu.

**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<sup>er</sup> 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.

Grâce à l'appariement, on a pu alléger sensiblement les questionnaires (ménage et individu), en ce qui concerne les "prestations familiales", les "aides au logement", le bloc "impôt sur le revenu", "prime pour l'emploi", le bloc "impôts locaux", salaires, "Allocations de chômage, allocations de préretraite, pensions et retraites" ou "Aide sociale".

Le temps dégagé du fait de l'appariement a permis d'introduire des questions sur les conditions de vie des ménages (insertion sociale, loisirs, culture, santé, travail...) permettant ainsi de développer les mesures de l'exclusion/inclusion sociale dans SRCV, conformément aux objectifs de SILC.

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 ;
  - Indicateurs sociaux : Santé, participation sociale et relations professionnelles, vie associative.

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 certains 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*

L'appariement avec les sources fiscales et sociales permet de garantir une bonne exhaustivité et de limiter les doubles comptes. Pour les revenus non imposables et certains individus non appariés (jeunes adultes), la stratégie de questionnement sur les revenus reste la même que les années précédentes. Comme l'enquêteur n'est pas en mesure de connaître à priori 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 2009), 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'enquêté 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 du type de revenus et des documents utilisés. Lorsque le questionnement complet est passé : parmi ces documents, la ou les déclarations des revenus de l'année de référence 2009 (pré-remplie par l'administration fiscale, reçue en mai 2010) 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) est homogène avec celui mesuré dans l'enquête Revenus fiscaux et sociaux (enquête de référence à l'Insee pour la distribution des revenus).

Dans le questionnaire individuel, l'enquêté indique combien de déclarations de revenus 2009 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*

Un logiciel de codification de l'activité de l'établissement est intégré dans le questionnaire CAPI. Il permet de reconnaître les libellés d'activité de l'établissement et de le recoder automatiquement. Dans le cadre de la profession, une reconnaissance du libellé est effectuée à partir de la base des professions. Les enquêteurs sont avertis lorsque le libellé renseigné est inconnu de la base des professions. Ils peuvent alors saisir d'autres libellés : ceci permet de corriger les erreurs de saisie et de préciser les libellés. Lorsque le libellé n'est pas connu, il est recodé par une équipe spécialisée : cela concerne 15 % des libellés. Le codage de la profession s'effectue ensuite en production aval à l'aide d'un programme de chiffrage automatique développé par l'Insee.

#### *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 l'erreur de mesure. 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 2010 a bénéficié des bilans des collectes 2004 à 2009. Le bilan de la collecte 2010 a été effectué grâce aux synthèses rédigées par les directions régionales et des comptes-rendus des réunions de bilans organisées dans les régions Rhône-Alpes et Pays de la Loire.

Le questionnaire de 2010 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2009 dans les directions régionales de Languedoc Roussillon et du Nord/Pas-de-Calais. Il a été mené auprès de 134 ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2009 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2010. 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-juin 2010), 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 obtenir la coopération/participation des ménages),
- 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 2010. 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 2010. 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 2010. L'échantillon pour lequel cette variable vaut 9 correspond à l'échantillon entrant en 2010.

**Erreur ! Liaison incorrecte.**

**Erreur ! Liaison incorrecte.**

#### 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 16,4 %.

**Erreur ! Liaison incorrecte.**

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

**Erreur ! Liaison incorrecte.**

On obtient ainsi un taux global de non-réponse individuelle de 17,1% très proche du taux de non-réponse des ménages à 16,4 %.

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)**

**Erreur ! Liaison incorrecte.**

**Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130)**  
**Erreur ! Liaison incorrecte.**

**Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) :**  
**Erreur ! Liaison incorrecte.**

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.

##### **Ventilation des non-réponses à certaines questions**

**Erreur ! Liaison incorrecte.** \* Les écarts à 100% concernent des ménages avec un revenu disponible calculé à zéro.

**Erreur ! Liaison incorrecte.**

### 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.

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

##### **Erreur ! Liaison incorrecte.**

<sup>1</sup> dont 5 464 enfants de moins de 16 ans et 467 individus qui n'ont pas gardé le même statut professionnel plus de la moitié de l'année.

<sup>2</sup> 161 individus n'ont pas rempli de questionnaire individuel, leur statut d'occupation est donc inconnu.

<sup>3</sup> 3245 personnes dans un ménage d'étudiants ou dans un ménage avec seulement des enfants dépendants ou des personnes âgées de plus de 65 ans.

<sup>4</sup> 401 individus appartenant à un ménage répondant dont un individu n'a pas rempli de questionnaire individuel.

<sup>5</sup> 23 258 non-pauvres.

#### **2.4 Mode de collecte des données**

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,2 %. 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**

**Erreur ! Liaison incorrecte.**

## **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 27,6 %.

**Erreur ! Liaison incorrecte.**

### **2.5 Durée de l'entretien**

La durée moyenne de l'entretien pour la collecte 2010 est de 48 minutes (56 minutes pour les entrants et 47 minutes pour les réinterrogés).

### **2.6 Les imputations**

Depuis l'enquête réalisée en 2008, à de très rares exceptions près, les ménages ne sont plus questionnés sur le montant des revenus. Ces derniers sont obtenus à partir des fichiers fiscaux de la Direction Générale des Finances Publiques ou des organismes gestionnaires de prestations pour les revenus sociaux. Le rapprochement du fichier d'enquête et du fichier fiscal est opéré à l'aide de l'adresse du ménage et d'informations sur le déclarant (nom, prénom, date de naissance, département de naissance, pays de naissance). Dans le cas où les adresses ne correspondraient pas, les revenus ne seront pas retrouvés. C'est le cas des jeunes en particulier qui peuvent déclarer leur impôt avec leurs parents s'ils sont âgés de moins de 21 ans ou bien s'ils sont étudiants de moins de 25 ans. Si ces jeunes ont décohabité, le risque de ne pas retrouver leurs revenus est élevé, c'est pourquoi nous les interrogeons toujours suivant l'ancienne méthode (collecte en face à face).

Les prestations familiales et les minima sociaux sont obtenus auprès des organismes les détenant, c'est toujours l'adresse du ménage qui permet la recherche et les caractéristiques démographiques de l'allocataire (sexe, date de naissance) : pour diverses raisons certaines adresses ne correspondent pas, ces revenus sont alors imputés.

L'imputation est d'abord nécessaire parce qu'il existe des données manquantes ou en tranches. Les données manquantes proviennent des individus dont la déclaration fiscale n'a pu être retrouvée. Les données en tranches concernent les enfants rattachés fiscalement au foyer fiscal de leurs parents et ne résidant pas avec eux-ci. Les revenus dans l'EU-SILC 2008 sont relatifs à plusieurs dates. L'impôt payé a pour assiette les revenus imposables perçus au cours de l'année 2006. Les revenus appariés et collectés sont relatifs à l'année 2007. 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 (pour les jeunes) ou non appariés. Dans le second cas, nous utilisons le revenu donné par l'individu à une date précédente 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 puis nous comparons le salaire imputé avec les maxima observés dans l'ERFS en tenant compte du sexe du salarié et de sa catégorie socioprofessionnelle (sur une position). Nous attribuons aussi un salaire aux individus déclarant en percevoir et qui ne figurent pas dans le fichier des impôts, ainsi qu'aux individus ayant répondu par un montant en tranches (cas des jeunes).

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. 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 : nous leur attribuons le montant moyen de préretraites d'individus ayant des caractéristiques similaires.

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

Lorsque le montant de la pension est manquant, 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.

Les retraites étant quasi stables, il est possible par ailleurs de contrôler le montant imputé. 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 de revenu antérieures, toujours sous la même hypothèse nous comparons les impôts sur le revenu déclarés que nous calculons. Deux contrôles sont donc possibles avant de prendre une décision.

Par ailleurs, depuis les revenus 2008, a été imputée une majoration de retraite (laquelle est non imposable) pour les individus ayant eu ou adopté 3 enfants au moins au cours de leur vie (la majoration n'est dans la réalité accordée que pour ceux ayant élevé les 3 enfants jusqu'à l'âge de 7 ans) représentant 10% du montant des pensions de retraites déclarées. Ceci n'est qu'une approximation car ce pourcentage varie selon les caisses de retraites mais, ne disposant pas d'information sur l'ensemble de la carrière de l'individu, il est difficile de construire des modèles permettant une estimation plus précise. Cependant, ce taux uniforme de 10% permet d'obtenir les masses financières cohérentes sur cette majoration avec des sources extérieures (EIR, Caisses de retraites).

#### **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 collecté par voie d'enquête est privilégié (les prélèvements privés que la personne a effectué sur les ressources de son activité d'indépendant pour ses besoins de consommation ou d'épargne). Ainsi si les deux revenus sont renseignés 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, depuis la collecte 2008 sur les prestations 2007, récupérées par appariement auprès des organismes de prestations sociales. Dans le cas où l'allocataire n'a pu être apparié, les prestations sont calculées sur barème. La principale difficulté est la période de référence des revenus pour les prestations sous conditions 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 résultats sont conformes aux données des CAF. Les valeurs calculées sont imputées si l'individu n'a pas été retrouvé dans les fichiers sociaux.

### **Aides au logement**

Pour les allocataires non appariés, 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 (ménages en institutions). 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'ERFS. Nous avons en effet choisi d'appliquer le barème locatif aux accédants, en nous aidant du loyer fictif imputé.

### **Minima sociaux**

Au 01/06/2009 le RMI est remplacé par le Revenu de Solidarité Active (RSA). Le RSA assure un revenu minimum à des personnes qui ne travaillent pas (RSA socle) ou à des personnes qui travaillent déjà et dont les revenus sont faibles (RSA socle ou/et RSA activité). Le RSA remplace le RMI, l'API et certaines aides forfaitaires comme la prime de retour à l'emploi. En cas de non-appariement, le montant est imputé si la personne a déclaré être bénéficiaire. Pour la collecte 2010 sur les revenus 2009, le calage sur les sources extérieures a été réalisé en considérant le montant cumulé des deux revenus.. Dans la suite nous dénommerons ce cumul sous le nom de RMI\_RSA.

Pour les allocataires non appariés, trois minima sociaux sont imputés dans SILC : le revenu minimum d'insertion (RMI), le minimum vieillesse et l'allocation de parent isolé (API). Selon les données brutes, SILC comprend 50 % des bénéficiaires du RMI\_RSA, 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'ERFS. Elles présentent une limite. Le revenu retenu pour le calcul du RMI\_RSA est un revenu trimestriel que nous ne connaissons pas. Le RMI\_RSA imputé est égal à la différence entre le plafond du RMI\_RSA et les revenus de l'année. 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.

N'ont pas été traités les dispositifs d'intéressement à la prise ou à la reprise d'un emploi. Néanmoins, une étude sur l'impact de l'intéressement du RMI\_RSA dans l'enquête ERFS a montré que les taux de pauvreté restaient inchangés.

### **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 2006. 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 m<sup>2</sup>) 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.
- 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	2 439 984	2153140	1166762	1269214	3 313 533	4001810	3 313 533	3187823
Moyenne	703	729	675	620	480	345	480	532
Minimum	57	86	115	79	32	9	32	5
Q5	260	277	281	253	230	168	230	250
Q10	314	338	332	319	269	210	269	300
Q25	425	436	432	403	345	260	345	386
Q50	580	612	588	545	430	320	430	495
Q75	855	852	809	743	552	400	552	620
Q90	1 183	1212	1 081	1013	691	509	691	800
Q95	1 438	1518	1 376	1196	827	590	827	930
Maximum	10 438	5679	3 968	3511	3 951	1626	3 951	3300

## 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	8 424 053	9101186	3 969 054	4480889	820 293	1302038	1 504 239	1698772
Moyenne	482	479	644	630	350	391	501	548
Minimum	31	40	49	106	37	7	36	29
Q5	158	152	260	277	159	150	189	235
Q10	201	196	329	343	197	213	260	300
Q25	293	287	452	449	267	300	352	398
Q50	415	408	590	592	344	380	462	510
Q75	581	579	767	762	418	468	605	660
Q90	799	795	997	958	487	590	760	850
Q95	986	984	1 188	984	540	652	875	990
Maximum	7 384	4914	5 700	2859	2 251	1200	3 160	1900

### 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 2008, 2,4 % de la population de la France métropolitaine vivait en collectivité, en institution ou était sans domicile fixe<sup>10</sup>. 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.

Moins de 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 2010. L'année de référence pour les revenus est 2009. On collecte au niveau individuel et ménage les revenus perçus au cours de l'année 2009. On collecte au niveau du ménage les impôts payés en 2009 au titre des revenus perçus au cours de l'année 2008 ; si l'information est manquante, on attribue l'impôt qui sera payé en 2010 sur les revenus 2009. Les cotisations sociales sont relatives aux revenus perçus en 2009.

##### *Période de référence pour les impôts sur la fortune*

On considère le montant de l'impôt de solidarité sur la fortune payé en 2009. Il est relatif au patrimoine détenu au 1<sup>er</sup> janvier 2009.

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

Certaines variables comme 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 2010, ce qui diffère de la période de référence des revenus portant sur l'année 2009 (revenus déclarés). Les impôts sur le revenu retenus quant à eux portent sur les revenus de l'année 2008 (assiette des impôts sur le revenu) payés en 2009.

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

La collecte SILC s'est déroulée du 10 mai au 26 juin 2010.

##### *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 un programme proche du système de codification automatique SICORE développé à l'INSEE. 85 % des professions sont codées automatiquement, les autres professions étant « reprises » à la main par une équipe spécialisée à l'Insee.

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

---

<sup>10</sup> 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*.

l'établissement. Dans 95 % des cas, la matrice de passage élaborée pour l'enquête Emploi<sup>11</sup> permet de coder la profession de l'enquêté en un code ISCO à deux positions à partir d'informations sur la catégorie socioprofessionnelle mais aussi de la NAF2008.

## **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 généralement les définitions des statistiques EU-SILC. Le contour des variables varie parfois à la marge car la source administrative utilisée pour le recueil des données ne s'accorde pas toujours avec la demande européenne.

### **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 2008. Il n'est pas possible d'isoler des salaires récupérés par appariement avec les fichiers administratifs 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 ou le RSA. La plupart sont 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 ;
- La prime de naissance de la prestation d'accueil du jeune enfant (PAJE). Elle est versée sous condition de ressources au 7<sup>ème</sup> mois de grossesse ;
- L'allocation de base de la PAJE. Sous condition de ressources, elle est versée jusqu'aux 3 ans du dernier-né ;
- Le complément de libre-choix d'activité de la PAJE. Il vise à compenser l'arrêt total ou partiel d'activité d'un des deux parents pour élever son enfant âgé de moins de 3 ans. La durée maximale de l'arrêt est de 6 mois pour un premier enfant et 3 ans pour les suivants. Son montant est modulé en fonction de la perception ou non de l'allocation de base ;
- 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 parent isolé (API) est un minimum social versé aux personnes 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 journalière de présence parentale (AJPP) pour enfants handicapés ou malades en cas d'arrêt total ou partiel de travail des parents.

L'inclusion du complément de mode de garde de la PAJE ou de l'aide à la famille pour l'emploi d'une assistante agréée (AFEAMA) ou de l'allocation de garde d'enfant à domicile (AGED) dans les

---

<sup>11</sup> Labour Force Survey

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. La mise en place de la PAJE se passe de la façon suivante : les ménages qui ont un nouveau-né après le 1<sup>er</sup> janvier 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 leRMI, leRSA (socle et activité) et les aides financières sociales versées par les collectivités locales. Ne disposant pas des sources de calage extérieures sur les aides locales, ce poste est sans doute sous-estimé par rapport à la réalité.

#### **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.

#### **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. Les enfants âgés de 14 à moins de 16 ans peuvent cependant recevoir des revenus d'apprentissage. Par ailleurs est inclus dans cette composante les 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 déclaré par les enquêtés lors de la collecte en face à face.

#### **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. Ces 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 (IR), la taxe d'habitation et les impôts payés à l'étranger. La prime pour l'emploi est comptée négativement. L'IR 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'IR.

#### **Salaires ou revenu assimilé (PY010N)**

Les salaires sont nets des cotisations sociales salariées. Depuis 2008, la majorité des salaires sont récupérées par appariement, ils comprennent les avantages en nature déclarés aux impôts et les allocations de préretraite imposables (ceux-ci ne peuvent pas être distingués des salaires). Mais nous disposons aussi de salaires non imposables (majorations de retraite pour avoir élevé trois enfants ou plus, indemnités de chômage ou maladie non imposables..) disponibles dans l'enquête mais non présents dans les fichiers fiscaux.

#### **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 privés que la personne a effectués sur les ressources de son activité d'indépendant pour ses besoins de consommation ou d'épargne. Le revenu fiscal est utilisé 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 et déclaratif.

### **Allocations de chômage (PY090N)**

Elles contiennent les allocations chômage déclarées aux impôts.

### **Allocations de vieillesse (PY100N)**

Du fait du mode de collecte des pensions de retraite (administratif), les imputations et le redressement se font au niveau du montant total des retraites. Les variables européennes PY100, PY110 et PY080 sont alimentées par déduction.

La variable PY100N concerne, pour les personnes ayant déclaré en face à face percevoir des pensions de droits directs (droits résultant de l'activité professionnelle de la personne), l'ensemble des montants retraites déclarés aux impôts (retraites de droit propre comme de droit dérivé, les préretraites (à l'exclusion de celles incluses dans les allocations chômage et les pensions d'invalidité) et retraites facultatives), lesquelles ne sont pas dissociables entre elles, auxquelles on rajoute l'allocation de solidarité aux personnes âgées (ASPA, ex-minimum vieillesse) servies aux personnes âgées de 65 ans ou plus. L'ASPA est un minimum social versé sous condition de ressources aux personnes ayant atteint 65 ans.

### **Pensions au survivant (PY110N)**

Ce sont les pensions de réversion, c'est-à-dire les pensions reçues par les survivants n'ayant pas atteint 65 ans. Cela ne concerne que les personnes n'ayant pas encore de droit propre, sinon ils seraient inclus dans la variable PY100N. Y sont inclus également les retraites facultatives de ces personnes, non dissociables dans la donnée administrative.

### **Retraites non obligatoires (PY080N)**

Ce sont les retraites facultatives perçues par les personnes ne percevant ni de droit direct ni de droit dérivé (soit avec PY100N et PY110N =0) et 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 et déclarées.

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

Les revenus sont récupérés par appariement avec les fichiers fiscaux et sociaux principalement et collectés par enquête pour les revenus non imposables ou pour les personnes pour lesquels nous savons que l'appariement ne sera pas possible (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 sont déclarées nettes des cotisations sociales mais y compris des 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 récupérés par appariement. Les cotisations sociales ont été imputées sur barème. Des valeurs brutes et nettes ont été estimées pour les agrégats suivants.

### **Salaire ou revenu assimilé (PY010G et PY030G)**

Le salaire est récupéré par appariement. Dans le cas où il est collecté, il est dans la plupart des cas celui 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 aussi 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 (cf. 2.6), 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 d'invalidités (PY120), indemnités maladie (PY120), bourses d'études (PY140), transferts entre ménage (HY080 et HY130), remboursement de prêts hypothécaires (HY100), aides sociales (HY060), bourses scolaires (HY110)**

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 des ménages 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 barème jusqu'en 2005 puis collectées auprès des organismes sociaux depuis 2006.

Nous avons aussi mobilisé les statistiques de la Caisse nationale des Allocations familiales et de la Direction de la recherche, des études, de l'évaluation et des statistiques sur les revenus sociaux.

#### 4.1.1 Salaires et revenus assimilés (PY010N)

Année de revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Salaires (milliards d'€)	527	533	553	554
Bénéficiaires (millions)	27,8	28,6	27,1	26,9
Moyenne	18 957	18650	20 411	20558
Minimum	10	4	1	1
D10	3 408	2476	3 927	3450
D20	7 904	6520	8 846	8291
D30	12 160	10857	13 210	12909
D40	14 923	14461	15 908	15935
D50	17 000	16921	18 179	18275
D60	19 306	19206	20 453	20697
D70	22 190	22069	23 365	23762
D80	26 063	26121	27 600	28030
D90	33 321	33069	35 465	36330
Maximum	360 554	1199980	875 204	831288

Les écarts entre les deux enquêtes sont de l'ordre de 4 à 5 % en masses. Ces variations en masse trouvent en partie leur origine du fait que la CSG non déductible et la CRDS sont comprises dans le salaire ERFS, tandis qu'elles ont été enlevées du salaire SRCV. Le plus grand nombre de bénéficiaires dans EU-SILC s'explique par un champ plus large pris en compte : les personnes vivant dans un ménage où aucune déclaration de revenu n'est retrouvée sont en partie exclues de l'ERFS alors qu'elles sont couvertes par SRCV.

#### 4.1.2 Les allocations chômage (PY090N)

	EU-SILC		ERFS	
	2008	2009	2008	2009
Revenus				
Chômage (milliards d'€)	24,2	28,7	23,1	27,6
Bénéficiaires (millions)	4,4	5,0	4,1	4,5
Moyenne	5 535	5739	5720	6142
Minimum	22	3	1	1
D10	583	630	651	748
D20	1 170	1355	1 363	1645
D30	2 020	2162	2 110	2599
D40	2 833	3169	3 062	3550
D50	3 970	4325	4 136	4676
D60	5 191	5437	5 282	5626
D70	6 382	7035	6 502	7282
D80	8 475	9094	8 671	9643
D90	11 206	11555	11 857	12320
Maximum	240 864	105980	330 013	114124

Les écarts entre les deux enquêtes sont faibles, les données étant récupérées par appariement dans les deux enquêtes.. Comme pour les salaires, la CSG non déductible et la CRDS sont comprises dans les allocations chômage ERFS, tandis qu'elles ont été enlevées des allocations SILC.

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

Année de revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Retraites (milliards d'€)	242	251	221	230
Bénéficiaires (millions)	15,0	15,2	15,2	15,4
Moyenne	16 176	16530	14 525	14930
Minimum	86	48	22	11
D10	4 322	4492	2 815	2951
D20	7 954	8116	5 608	5886
D30	10 171	10493	8 343	8765
D40	12 494	12682	10 525	10950
D50	14 586	14762	12 904	13370
D60	16 644	16831	15 179	15627
D70	19 155	19235	17 736	18284
D80	22 630	22976	21 309	21746
D90	28 148	29014	27 164	27476
Maximum	155 593	168196	315 199	469270

L'appariement avec les données fiscales a permis d'améliorer la collecte des revenus de retraites plus sensibles aux effets mémoires des enquêtés. Par contre, en 2008, les montants de retraites sont ceux déclarés aux impôts et récupérés par appariement ; ils ne contiennent pas la majoration de retraite pour trois enfants ou plus qui n'est pas imposable. Dans SILC, 10 %<sup>12</sup> de majoration a donc été rajoutée aux retraités ayant eu au cours de leur vie 3 enfants ou plus : le nombre d'individus percevant cette majoration et la masse des majorations est conforme à ce qui est connu par ailleurs. Cela amène une différence avec l'enquête ERFS qui ne tient pas compte de cette majoration (la descendance n'étant pas connu dans cette enquête).

<sup>12</sup> Si le montant de la majoration versée par la caisse nationale d'assurance vieillesse, qui verse les retraites des salariés du privé, est fixée à 10 %, ce n'est pas le cas de toutes les caisses de retraite : certaines appliquent une majoration variable selon le nombre d'enfants. Nous ne disposons cependant pas dans l'enquête du nombre et du nom des caisses qui versent une pension aux retraités : seuls le dernier emploi est connu, alors qu'il faudrait idéalement disposer de l'ensemble du parcours professionnel pour estimer une majoration. Seule une estimation grossière pourrait être faite.

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

Année de revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Revenus d'indépendants (milliards d'€)	51,7	48,5	56,9	53,7
Bénéficiaires (millions)	2,45	2,5	2,01	2,06
Moyenne	21 074	19426	28 315	26123
Minimum	1	9	1	1
D10	1 091	1051	1 383	1140
D20	2 362	2190	4 176	3241
D30	4 332	4245	7 363	6060
D40	6 936	7366	10 899	9451
D50	10 339	11417	15 055	13461
D60	14 681	14400	20 350	19082
D70	20 000	19200	28 319	26850
D80	29 000	26634	41 271	39775
D90	47 010	45108	66 959	64795
Maximum	1 303 181	1000000	1 476 829	2299911

Les revenus des indépendants sont abordés de façon différente dans l'ERFS 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'ERFS 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. Ce résultat peut être négatif, le cas échéant il n'a pas été retenu pour la comparaison.

### 4.1.4 Revenus sociaux (HY050N, HY060N et HY070N)

Année de revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Sociaux (milliards d'€)	45,0	50,0	45,5	49,6
Bénéficiaires (millions)	11,7	11,5	11,1	11,7

Les bénéficiaires se comptent en ménages.

### 4.1.5 Prestations familiales (HY050N),

Année de revenus	EU-SILC		Cnaf	
	2008	2009	2008	2009
Prestations familiales (milliards d'€)	24,6	25,4	28,0	28,9
Bénéficiaires (millions)	6,8	6,8		

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 2009 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2009 qui est automatiquement supérieur, les allocataires de l'année pouvant ne plus percevoir de prestations en décembre.

#### 4.1.5.1 Aides au logement (HY070N)

Année de revenus	EU-SILC		Cnaf	
	2008	2009	2008	2009
Allocations logements (milliards d'€)	13,6	12,9	14,8	15,1
Bénéficiaires (millions)	7,0	6,7		

La Cnaf ne publie que le nombre de bénéficiaires au 31 décembre 2009 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2009.

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

Année de revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Impôts (milliards d'€)	50,5	53,4	57,2	61,4

Les deux enquêtes diffèrent sur la masse de l'impôt sur le revenu. La période de référence n'est pas la même : l'ERFS retient l'impôt relatif à l'année des revenus qui est payé un an plus tard, tandis que le panel collecte l'impôt de l'année payé pendant l'année de référence des revenus et relatif donc aux revenus de l'année précédente. Mais ceci n'explique qu'une partie de la différence observée.

#### 4.1.7 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 2009
Effectif	10 353 930	9 843 061
Moyenne	400	434
Q0	14	1
Q5	175	170
Q10	211	224
Q25	269	290
Q50	356	395
Q75	475	530
Q90	617	690
Q95	728	800
Q100	3 951	6254

#### 4.1.8 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 et flux appariés. 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 données de la Banque de France est appliqué aux stocks collectés ou imputés sur l'année de référence afin d'estimer les intérêts perçus durant l'année.

La différence sur les masses financières entre EU-SILC et ERFS résulte du fait que l'on tient compte aussi dans SILC de la donnée collectée sachant que tous les revenus financiers ne sont pas imposables : de fait lorsque le revenu financier estimé à partir de la donnée collectée est supérieure à celui de la donnée fiscale, c'est la première qui est retenue.

#### 4.1.9 Revenu disponible du ménage et niveau de vie des individus.

Année des revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Revenus disponible des ménages (milliards d'€)	983,42	944,44	917,63	923,24
Bénéficiaires (millions)	27,0	27,3	27,15	27,3
Moyenne	36 450	36437	33 803	33 818
Minimum	10	-130 436	-61 274	-52 210
D10	13 765	13 453	12 035	11 997
D20	17 630	17 169	16 024	16 026
D30	21 335	20 998	19 716	19 614
D40	25 537	25 258	23 712	23 704
D50	29 964	30 017	28 091	28 243
D60	34 755	35 326	33 010	33 170
D70	10 502	41 014	38 693	38 856
D80	48 235	49 567	46 161	46 429
D90	62 619	64 430	58 983	59 596
Maximum	993 412	1 262 585	1 415 623	2 153 787

Revenus	EU-SILC		ERFS	
	2008	2009	2008	2009
Niveau de vie des ménages (milliards d'€)	635,80	641,35	587,03	593
Bénéficiaires (millions)	26,7	27,3	27,15	27,3
Moyenne	23 550	23 500	22 085	21 723
Minimum	-33 656	-87 964	-696	-40 351
D10	11 071	10 985	10 523	9 830
D20	13 509	13 742	13 120	12 627
D30	15 659	15 786	15 198	14 786
D40	17 766	17 766	17 099	16 711
D50	19 737	19 951	18 986	18 743
D60	21 983	22 354	21 122	21 065
D70	24 727	25 405	23 791	23 873
D80	28 990	29 797	27 632	28 010
D90	37 308	38 150	35 549	35 537

Maximum	622 275	422 932	943 749	1 485 695
---------	---------	---------	---------	-----------

Les revenus estimés dans EU-SILC sont proches des revenus estimés par l'ERFS puisque les revenus sont collectés par appariement dans ERFS et c'est le cas pour la majorité des revenus dans SILC. Cependant, le champ des revenus couverts étant plus large par rapport à ERFS (revenus non imposables, autoconsommation), le niveau de vie des individus s'en trouve plus élevé également.

## 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 risqué<sup>13</sup>. 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'intérêt<sup>14</sup>. 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. Une estimation est réalisée sur le groupe des répondants et nous faisons implicitement l'hypothèse, qu'au sein des deux groupes (répondants/non répondants), 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.

La présence d'erreurs dans l'enquête EU-SILC 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.

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

---

<sup>13</sup> Pour le risque quadratique calculé avec l'aléa de modèle.

<sup>14</sup> Dans certains cas seulement la différence peut être négligeable.

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 efficace<sup>15</sup>.

---

<sup>15</sup> 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.