

**Rapport intermédiaire  
sur la qualité des données  
françaises  
EU-SILC 2009**

## TABLE DES MATIÈRES

<b>Table des matières .....</b>	<b>2</b>
<b>1 Indicateurs transversaux communs de l'Union européenne .....</b>	<b>4</b>
1.1 Indicateurs communautaires transversaux communs fondés sur la dimension transversale des statistiques EU-SILC .....	4
1.1.1 Taux de pauvreté après transferts sociaux .....	4
1.1.2 Inégalité de la distribution des revenus Q80/Q20, ratio interquintile .....	9
1.1.3 Intensité de la pauvreté .....	9
1.1.4 Dispersion autour du seuil de pauvreté .....	9
1.1.5 Taux de pauvreté ancré à une date fixée (2006).....	10
1.1.6 Taux de pauvreté avant transferts .....	10
1.1.7 Inégalité de la distribution des revenus : coefficient de Gini .....	11
1.2 Autres indicateurs .....	11
1.2.1 Niveau de vie médian .....	11
1.2.2 Niveau de vie moyen .....	11
1.2.3 Écart de rémunération entre les sexes non ajusté .....	11
<b>2 Précision .....</b>	<b>12</b>
2.1 Plan d'échantillonnage .....	12
2.1.1 Type de plan d'échantillonnage .....	12
2.1.2 Unités d'échantillonnage.....	12
2.1.3 Critères de stratification et de sous-stratification.....	12
2.1.4 Taille de l'échantillon et critères d'attribution.....	15
2.1.5 Systèmes de sélection de l'échantillon .....	15
2.1.6 Répartition de l'échantillon au cours du temps.....	15
2.1.7 Renouvellement de l'échantillon : groupes de rotation.....	15
2.1.8 Pondération.....	15
2.1.9 Substitutions .....	23
2.2 Erreurs d'échantillonnage .....	24
2.3 Erreurs non dues à l'échantillonnage.....	27
2.3.1 Erreurs dans la base de sondage et la couverture .....	27
2.3.2 Erreurs de mesure et de traitements .....	28
2.3.3 Erreurs de non-réponse .....	31
2.4 Mode de collecte des données .....	38
2.5 Durée de l'entretien.....	39
<b>3 Comparabilité .....</b>	<b>44</b>
3.1 Concepts et définitions de base.....	44

3.2	Composantes du revenu .....	45
3.2.1	Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC	45
3.2.2	La source ou la procédure utilisée pour la collecte des variables du revenu .....	48
3.2.3	La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes .....	48
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) .....	48
<b>4</b>	<b>Cohérence .....</b>	<b>50</b>
4.1	Comparaison des variables cibles avec des sources extérieures .....	50
4.1.1	Salaires et revenus assimilés (PY010N) .....	50
4.1.2	Les allocations chômage (PY090N) .....	51
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension au survivant (PY110N) .....	52
4.1.4	Revenus sociaux (HY050N, HY060N et HY070N) .....	53
4.1.5	Prestations familiales (HY050N), .....	53
4.1.6	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation). .....	54
4.1.7	Loyers .....	54
4.1.8	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N) .....	55
4.1.9	Revenu disponible du ménage et niveau de vie des individus .....	55

# 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 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 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 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 globale pour les autres revenus. L'impact de l'appariement avec la source fiscale avait été évalué à 0,2 point sur le taux global 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

	Total			0-17	0 - 64			18+			18 - 64		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Pauvres	12,9	13,7	12,0	17,3	13,3	14,1	12,4	11,6	12,4	10,8	11,9	12,5	11,2
Non pauvres	87,1	86,3	88,0	82,7	86,7	85,9	87,6	88,4	87,6	89,2	88,1	87,5	88,8

	18 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	21,2	21,8	20,6	11,2	12,1	10,2	9,0	9,5	8,5	10,7	11,9	9,1
Non pauvres	78,8	78,2	79,4	88,8	87,9	89,8	91,0	90,5	91,5	89,3	88,1	90,9

#### Effectifs non pondérés

	Total			0 - 17	0 - 64			18+		
	Total	F	H	Total	Total	F	H	Total	F	H
Total	25611	13284	12327	6081	21563	11046	10517	19530	10240	9290
Pauvres	3189	1758	1431	1019	2786	1505	1281	2170	1212	958
Non pauvres	22422	11526	10896	5062	18777	9541	9236	17360	9028	8332

	18 - 64			18 - 24			25 - 49			50 - 64		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	15482	8002	7480	2240	1094	1146	8156	4241	3915	5086	2667	2419
Pauvres	1767	959	808	444	218	226	885	500	385	438	241	197
Non pauvres	13715	7043	6672	1796	876	920	7271	3741	3530	4648	2426	2222

	65+		
	Total	F	H
Total	4048	2238	1810
Pauvres	403	253	150
Non pauvres	3645	1985	1660

*Répartition de la population totale selon le sexe<sup>1</sup>*

Total	F	H
100,0	51.6	48.4

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

	0 - 17	0 - 64	18+	18 - 64	18 - 24	25 - 49	50 - 64	65+	Total
Total	21.7	83.4	78.3	61.7	8.6	33.6	19.5	16.6	100,0
F	20,6	81,4	79,4	60.8	8,2	33,1	19.5	18,6	100,0
H	22,9	85,6	77,1	62.7	9,0	34.2	19.5	14,4	100,0

*Répartition de la population pauvre selon le sexe*

Total	F	H
100,0	55.0	45.0

*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	29.2	86.2	70.8	57.1	14.2	29.2	13.7	13.8	100,0
F	28.0	83.9	72,0	55.8	13,1	29.2	13.5	16.1	100,0
H	30.6	89.1	69.4	58.5	15,5	29.2	13.9	10.9	100,0

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

	Total			En emploi			Sans emploi : total			Sans emploi : Chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	11.7	12.6	10.7	6.7	6.1	7.2	17.2	18.5	15.5	37.8	33.5	42.5
Non pauvres	88,3	87,4	89,3	93,3	93,9	92,8	82,8	81,5	84,5	62,2	66,5	57,5

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Pauvres	8.7	9.3	8.1	25.7	26.8	23.2
Non pauvres	91,3	90,7	91,9	74,3	73,2	76,8

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.

<sup>1</sup> Les tableaux en % sont pondérés avec la variable rb050

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	19849	10428	9421	10343	4974	5369	9506	5454	4052	876	463	413
Pauvres	2231	1256	975	668	292	376	1563	964	599	314	149	165
Non pauvres	17618	9172	8446	9675	4682	4993	7943	4490	3453	562	314	248

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Total	5087	2580	2507	3543	2411	1132
Pauvres	404	213	191	845	602	243
Non pauvres	4683	2367	2316	2698	1809	889

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

	Total	En emploi	Sans emploi : total	Sans emploi : chômeurs	Sans emploi : retraités	Sans emploi : autres inactifs
F	100,0	47,5	52,5	4,7	26,8	21,0
H	100,0	57,5	42,5	4,6	27,6	10,4
Total	100,0	52,2	47,8	4,6	27,2	16,0

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

	Total	En emploi	Sans emploi : total	Sans emploi : chômeurs	Sans emploi : retraités	Sans emploi : autres inactifs
F	100,0	23,1	76,9	12,5	19,7	44,7
H	100,0	38,5	61,5	18,2	20,8	22,5
Total	100,0	29,8	70,2	15,0	20,2	35,0

### 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	10.1	16.3	6.9	7.6	7.2
Non pauvres	89,9	83,7	93,1	92,4	92,8

	Avec enfant : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Pauvres	15.4	29.4	5.6	10.6	20.9	24.9
Non pauvres	84.6	70,6	94,4	89,4	79,1	75,1

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	11331	3024	3860	2878	1542
Pauvres	1066	499	264	198	102

	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
Non pauvres	10265	2525	3596	2680	1440

Total	Avec enfant : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Total	14280	1536	2838	5028	3285	1523
Pauvres	2123	449	180	500	622	339
Non pauvres	12157	1087	2658	4528	2663	1184

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

	F	H	< 65	65+
Pauvres	16.9	15.5	16.9	15.5
Non pauvres	83,1	84,5	83,1	84,5

*Effectifs non pondérés*

	F	H	< 65	65+
Total	1834	1190	1717	1307
Pauvres	311	188	299	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
48.1	15.3	16.2	11.1	5.5

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
51.9	5.9	12.1	17.9	10.5	5.5

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

F	H	< 65	65+
58.0	42.0	60.2	39.8

*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
37.8	19.7	8.8	6.7	3.1

Avec enfants : total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
62.2	13.5	5.4	14.9	17.2	10.7

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

F	H	< 65	65+
60.1	39.9	62.2	37.8

#### 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	12.9	21.9	8.2
Non pauvres	87,1	78,1	91,8

#### Effectifs non pondérés

	Total	Locataires	Propriétaires ou logés à titre gratuit
Total	25611	7954	17657
Pauvres	3189	1765	1424
Non pauvres	22422	6189	16233

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	33.9	66.1

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	42.4	57.6

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

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

	Sans enfant $W=0$	Sans enfant $0<W<1$	Sans enfant $W=1$	Avec enfants $W=0$	Avec enfants $0<W<0,5$	Avec enfants $0,5<W<1$	Avec enfants $W=1$
Pauvres	16.1	10.7	4.4	68.3	52.3	19.8	5.9
Non pauvres	83,9	89,3	95,6	31,7	47,7	80,2	94,1

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	1826	2771	3308	709	664	4315	8583
Pauvres	270	259	129	479	332	828	482
Non pauvres	1556	2512	3179	230	332	3487	8101

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

Total	Sans enfant $W=0$	Sans enfant $0<W<1$	Sans enfant $W=1$	Avec enfants $W=0$	Avec enfants $0<W<0,5$	Avec enfants $0,5<W<1$	Avec enfants $W=1$
100,0	9.0	13.1	17.3	3.0	2.9	18.4	36.4



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

Total	Sans enfant W=0	Sans enfant 0<W<1	Sans enfant W=1	Avec enfants W=0	Avec enfants 0<W<0,5	Avec enfants 0,5<W<1	Avec enfants W=1
100,0	11.2	10.8	5.9	15.8	11.6	28.1	16.6

#### 1.1.1.6 Seuil de pauvreté (valeurs indicatives)

En 2008, le seuil de pauvreté s'élève à 11 856 euros par an. Une personne seule est pauvre si elle a un revenu disponible inférieur à 988 euros par mois et un couple avec deux enfants de moins de 14 ans, s'il a un revenu disponible inférieur à 24 898 euros (soit 2 075 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 4,4 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	
18,4	18,3	18,5	17,7	18,6	18,5	18,8	19,8	19,8	19,7	14,7	14,5	15,5	

#### Effectifs non pondérés

Total			0 - 17		18+			18 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H	
3189	1758	1431	1019	2170	1212	958	1767	959	808	403	253	150	

#### 1.1.4 Dispersion autour du seuil de pauvreté

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

	Total	F	H
<i>40 % de la médiane</i>			
Pauvres	3.2	3.4	3.0
Non pauvres	96.8	96.6	97.0
<i>50 % de la médiane</i>			
Pauvres	6.8	7.2	6.3
Non pauvres	93.2	92.8	93.7
<i>70 % de la médiane</i>			
Pauvres	21.4	22.4	20.2
Non pauvres	78.6	77.6	79.8

### Effectifs non pondérés

	total	F	H
Total	25611	13284	12327
<i>40 % de la médiane</i>			
Pauvres	750	412	338
Non pauvres	24861	12872	11989
<i>50 % de la médiane</i>			
Pauvres	1681	931	750
Non pauvres	23930	12353	11577
<i>70 % de la médiane</i>			
Pauvres	5351	2907	2444
Non pauvres	20260	10377	9883

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

Il s'agit du taux de pauvreté calculé en 2008, 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	43.3	45.9	40.6	36.2	45.3	48.2	42.1	34.0	35.8	32.2	87.4	88.6	85.6
Non pauvres	56,7	54,1	59,4	63,8	54,7	51,8	57,9	66,0	64,2	67,8	12,6	11,4	14,4

### Effectifs non pondérés

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Total	25611	13284	12327	6081	19530	10240	9290	15482	8002	7480	4048	2238	1810
Pauvres	12312	7394	4918	2150	8658	4786	3872	5163	2820	2343	3495	1966	1529
Non pauvres	13299	5890	7409	3931	10872	5454	5418	10319	5182	5137	553	272	281

#### 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	23,8	24,6	22,9	34,8	20,7	21,7	19,6	22,6	23,6	21,5	13,6	15,2	11,4
Non pauvres	76,2	75,4	77,1	65,2	79,3	78,3	80,4	77,4	76,4	78,5	86,4	84,8	88,6

### Effectifs non pondérés

	Total			0 - 17	18+			18 - 64			65+		
	Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
Total	25611	13284	12327	6081	19530	10240	9290	15482	8002	7480	4048	2238	1810
Pauvres	6044	3243	2801	2068	3976	2179	1797	3473	1863	1610	503	316	187
Non pauvres	19567	10041	9526	4013	15554	8061	7493	12009	6139	5870	3545	1922	1623

#### 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 2008.

## 1.2 Autres indicateurs

### 1.2.1 Niveau de vie médian

En 2008, le niveau de vie médian des individus s'élève à 19 760 euros par unité de consommation, soit 1 647 euros par mois.

### 1.2.2 Niveau de vie moyen

En 2008, le niveau de vie des individus s'élève en moyenne à 23 336 euros par unité de consommation, soit 1 945 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

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

#### 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. Dans tout échantillon entrant, puisqu'on est exhaustif au dernier degré de sondage, on peut dire qu'il s'agit d'un sondage d'individus en grappes.

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

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 »),
- 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.

**Synthèse sur la constitution de l' Echantillon-Maître1999 (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 $N_{gcom} = 1332$ -	Groupes de communes  Commune $N_{gcom} = 284$	<i>Groupes de communes de l'EM</i>  D'au moins 1/6 de la taille de l'UP Tri sur la taille des communes $N_{gcom} = 209$	<i>Groupes de communes de l'EM</i>  Groupe ou arrondissement d'au moins 100 districts Tri sur la taille des communes $N_{gcom} = 392$	<i>Groupes de communes de l'EM</i>  Par département, groupe ou arrondissement d'au moins 100 districts Tri sur la taille des communes $N_{gcom} = 254$
Unités secondaires	Groupes de communes	Groupes de communes	<i>Groupes de districts de l'EM</i> Au moins 100 logements Tri sur la taille des districts dans le groupe de communes $N_{gdist} = 13642$	<i>Districts de l'EM</i>  District $N_{dist} = 101919$	<i>Districts de l'EM</i>  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 $n_{gcom}$ variable selon l'enquête	Exhaustif $n_{gcom} = 284$	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 gdist à tirer par gcom $n_{gdist} = 2790$	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 gcom $n_{dist} = 5098$	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 gcom $n_{dist} = 2465$
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

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%)

#### 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 2009 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).

#### 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.

#### 2.1.6 Répartition de l'échantillon au cours du temps

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

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

Pour la collecte de 2009, 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 2008 au moins une personne répondante et n'appartenant pas à l'échantillon sortant. Il comprenait 10 021 logements.

#### 2.1.8 Pondération

La pondération se fait au niveau logement.

En 2009, 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  $W_i$ , 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  $W_i$  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 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 pour construire des classes pour lesquelles on suppose un taux de réponse homogène.

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

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

où  $r_{t,k}$  est l'ensemble des répondants de l'échantillon  $a_{t,k}$ .

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 et 2008 : le nombre de groupe et les motifs principaux de non réponse variant à la marge. Par ailleurs en 2004, compte tenu de la taille d'échantillon plus grande (tous les échantillons étaient entrants, ce sont cinq modèles de non-réponses qui ont été estimés et non trois (cf. Rapport final sur la qualité des données françaises EU-SILC 2006).

Pour l'ensemble des ménages, PI est comprise entre 0,3 et 0,9. Le poids corrigé de la non-réponse initiale (WI/PI) varie dans un rapport de 1 à 10.

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

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 2009. 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 en  $t - k + 1$ ,  $\theta_i(t, k)$  est estimée en utilisant les informations individuelles et ménage collectée 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 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 2009, il faut en estimer 5.

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 2009 :

$$W_i^{*2009} = \frac{W_l(2009,k)}{\Phi_l(2009,k)} \times \frac{1}{\sum_{\substack{j \in I \\ j \in r_{t,k}}} \theta_j(2009,k)}$$

$$\times \frac{1}{7 \times Nchamp_l^{2004}(2009) + Nchamp_l^{2005,2008}(2009) + Npers_l(2009)}$$

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  $k^{\text{jème}}$  interrogation depuis 2004.

En 2009 par exemple, on estime cinq (2009-2004) modèles sur tous les individus qui répondent en  $k^{\text{jème}}$  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
- 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

- 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
- 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
- un modèle pour les individus interrogés en 6ème vague :
  - o sur les entrants 2004 réinterrogés en 2009

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 2ème vague sachant qu'on a répondu à la première*

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

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

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, les adultes réinterrogés en 2ème vague ont été réparti en 7 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(2ème vague) vaut donc :

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

La probabilité PR(2<sup>ème</sup> vague/1<sup>ère</sup> vague) est comprise entre 0,5 et 0,9.

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

Un modèle de non-réponse a été estimé sur les adultes réinterrogés en 3<sup>ème</sup> vague. L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait qu'il y ait eu un ou plusieurs départs dans le ménage depuis la dernière vague,
- le fait d'habiter dans une maison ou non
- le quartile de niveau de vie (1<sup>er</sup> ou plus ) du ménage ,
- la localisation géographique (agglomération parisienne/autre),
- la situation familiale (seule/non),
- avoir un contrat à durée indéterminée ou non

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer

la robustesse des résultats). Au final, nous avons réparti les individus dans 7 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(3<sup>ème</sup> vague/1<sup>ère</sup> vague) vaut donc :

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

La probabilité PR(3<sup>ème</sup> vague/1<sup>ère</sup> vague) est comprise entre 0,5 et 0,9.

*Probabilité de répondre en 4<sup>ème</sup> vague sachant qu'on a répondu à la 1<sup>ère</sup> vague*

Un modèle de non-réponse a été estimé sur les adultes panel interrogés en 4<sup>ème</sup> vague et répondants en 1<sup>ère</sup> vague . L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait qu'il y ait eu un ou plusieurs départs dans le ménage depuis la dernière vague,
- la nationalité (magrébine/autre),
- le quartile de niveau de vie (1<sup>er</sup> ou plus ) du ménage,

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les individus dans 5 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(4<sup>ème</sup> vague/1<sup>ère</sup> vague) vaut donc :

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

La probabilité PR(4<sup>ème</sup> vague/1<sup>ère</sup> vague) est comprise entre 0,5 et 0,9.

*Probabilité de répondre en 5<sup>ème</sup> vague sachant qu'on a répondu à la 1<sup>ère</sup> vague*

Un modèle de non-réponse a été estimé sur les adultes panel interrogés en 5<sup>ème</sup> vague et répondants en 1<sup>ère</sup> vague . L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir déménagé depuis 2004,
- le fait qu'il y ait eu un ou plusieurs départs dans le ménage depuis la dernière vague,
- avoir en contrat à durée indéterminée ou non
- le niveau de diplôme

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les individus dans 5 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(5<sup>ème</sup> vague/1<sup>ère</sup> vague) vaut donc :

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

La probabilité PR(5<sup>ème</sup> vague/1<sup>ère</sup> vague) est comprise entre 0,3 et 0,9.

*Probabilité de répondre en 6<sup>ème</sup> vague sachant qu'on a répondu à la 1<sup>ère</sup> vague*

Un modèle de non-réponse a été estimé sur les adultes panel interrogés en 6<sup>ème</sup> vague et répondants en 1<sup>ère</sup> vague . L'information utilisée est celle de la première vague. Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir déménagé depuis 2004,

- la nationalité (française / pays du Maghreb, /autre),
- la localisation géographique (agglomération parisienne/autre),

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les individus dans 5 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité PR(6ème vague/1ère vague) vaut donc :

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

La probabilité PR(6<sup>ème</sup> vague/1<sup>ère</sup> vague) est comprise entre X et Y.

### **Partage des poids**

Le partage des poids est une étape spécifique à la pondération transversale. Elle a pour but 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 l'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 (ç'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>2</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>3</sup>  $l$  du  $k^{\text{ième}}$  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  $S_l$ .

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.

<sup>2</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>3</sup> Il s'agit du poids de tirage multiplié par 9

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  $Li(t)$  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),  $Li(t)$  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 2009, est, comme toutes les enquêtes après des ménages de l'Insee, calée sur les marges issues de l'enquête Emploi de 2008. En effet, toutes les enquêtes ménages d'une année civile  $N$  sont calées sur la dernière enquête Emploi disponible au début de l'année  $N$ . Toutes les enquêtes dont la collecte a lieu pendant l'année civile  $N$  sont donc calées sur les marges constituées à partir des 4 enquêtes Emploi trimestrielles de l'année  $N-1$ .

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

Le calage 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 31/12/2008.

- 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).

#### 2.1.9 Substitutions

Sans objet dans le cas de la France.

## 2.2 Erreurs d'échantillonnage

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

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

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

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



Indicateurs	Valeur	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Taux de pauvreté après prestations sociales (en pourcentage)						
Actifs occupés	6.7	0.3	4.5	1.17	10323	8823
Chômeurs	37.7	2.0	5.3	1.29	873	677
Retraités	8.7	0.4	4.6	1.38	5087	3686
Autres inactifs	26.6	1.2	4.5	1.11	2876	2591
Hommes, actifs occupés	7.1	0.4	5.6	1.18	5351	4535
Hommes, chômeurs	42.4	2.3	5.4	1.22	412	338
Hommes, retraités	8.1	0.5	6.2	1.25	2507	2006
Hommes, autres inactifs	26.6	1.2	4.5	1.11	2876	2591
Femmes, actives occupées	6.1	0.4	6.6	1.14	4972	4361
Femmes, chômeuses	33.4	2.5	7.5	1.37	461	336
Femmes, retraitées	9.3	0.5	5.4	1.28	2580	2016
Femmes, autres inactives	27.4	1.5	5.5	1.15	2051	1783
Personnes seules, < 65 ans	16.9	0.8	4.7	1.21	1717	1419
Personnes seules, 65+ ans	15.5	0.9	5.8	1.18	1307	1108
Personnes seules, hommes	15.5	0.8	5.2	1.18	1190	1008
Personnes seules, femmes	16.9	0.8	4.7	1.21	1834	1516
Personnes seules, total	16.3	0.6	3.7	1.25	1834	1467
2 adultes, sans enfant, tous les deux < 65	6.9	0.5	7.2	1.26	3860	3063
2 adultes, sans enfant, au moins 1 65+	7.6	0.7	9.2	1.5	2878	1919
Autres ménages sans enfant	7.2	1.3	18.1	0.93	1542	1658
Familles monoparentales	29.4	2.1	7.1	1.14	1536	1347
2 adultes, 1 enfant	5.6	0.9	16.1	1.10	2838	2580
2 adultes, 2 enfants	10.6	1.0	9.4	1.21	5028	4155
2 adultes, 3+ enfants	20.9	2.2	10.5	1.13	3285	2907
Autres ménages avec enfants	24.9	5.3	21.3	1.20	1523	1269
Ménages sans enfant	10.1	0.3	3.0	1.26	11304	8971
Ménages avec enfants	15.1	0.6	4.0	1.12	14210	12688
Propriétaires	8.2	0.4	4.9	1.27	17657	13903
Locataires	21.9	0.8	3.7	1.24	7954	6415
Ménages sans enfant, w = 0	16.9	1.0	5.9	1.02	1826	1790
Ménages sans enfant, 0 < w < 1	10.7	0.7	6.5	1.23	2771	2253
Ménages sans enfant, w = 1	4.4	0.4	9.1	1.13	3308	2927
Ménages avec enfants, w = 0	68.3	3.6	5.3	1.10	709	645
Ménages avec enfants, 0 < w < 0,5	52.3	4.8	9.2	1.19	664	558
Ménages avec enfants, 0,5 < w < 1	19.8	1.5	7.6	1.24	4315	3480
Ménages avec enfants, w = 1	5.9	0.7	11.9	1.13	8583	7596
Médiane des niveaux de vie (en euros par unité de consommation)	19760	59.1	0.3	1.26	25611	20326
Seuil de pauvreté - personnes seules (en euros)	11856	0.6	0.0	1.25	25611	20489
Seuil de pauvreté - 2 adultes, 2 enfants (en euros)	24898	1.0	0.0	1.21	25611	21166

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

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

Deux bases de sondages ont été utilisées conjointement :

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

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

### 2.3.2 Erreurs de mesure et de traitements

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

Jusqu'en 2007, l'ensemble des données étaient récupérées 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 2008), 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 2008 (pré-remplie par l'administration fiscale, reçue en mai 2009) 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 2008 il a effectué, et s'il consent à s'y reporter. Pour les revenus d'activité et de remplacement, le questionnement prend deux formes différentes selon la réponse :

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

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

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

#### *Codification de la profession*

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

#### *Questionnement longitudinal*

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

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

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

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

#### *Tests du questionnaire*

Le questionnaire transversal de 2009 a bénéficié des bilans des collectes 2004 à 2008. Le bilan de la collecte 2009 a été effectué grâce aux synthèses rédigées par les directions régionales et des comptes-rendus des réunions de bilans organisées en Alsace, Haute Normandie et Nord.

Le questionnaire de 2009 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2008 dans les directions régionales de Champagne et Limousin. Il a été mené auprès de 141 ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2008 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2009. 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 2009), l'équipe de conception ne forme pas directement les enquêteurs mais les gestionnaires responsables de l'enquête dans les directions régionales. Les gestionnaires des directions régionales forment ensuite les enquêteurs SILC. Cependant, la formation des enquêteurs reproduit à l'identique la formation des gestionnaires. L'équipe de conception SILC fournit aux gestionnaires des directions régionales (en plus des

documents de collecte) les transparents et les exercices CAPI. La formation dispensée aux gestionnaires et aux enquêteurs a duré trois jours pour ceux qui ne connaissaient pas encore l'enquête et 1,5 jours pour ceux qui avaient déjà participé à la collecte précédente.

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

#### *Documents de collecte*

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

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

### 2.3.3 Erreurs de non-réponse

#### 2.3.3.1 Taille de l'échantillon obtenu

Les deux tableaux ci-dessous fournissent le nombre de ménages, puis le nombre d'individus correspondants, pour lesquels l'entretien a été accepté suite à la collecte 2009. 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 2009. 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 2009. L'échantillon pour lequel cette variable vaut 9 correspond à l'échantillon entrant en 2009.

<b>Nombre de ménages pour lesquels un entretien est accepté pour la base de données</b>		
Ventilation par sous-échantillon		
Nombre d'années restantes dans le panel	DB075 (sous-échantillon)	Nombre de ménages
1	9	677
2	8	627
3	7	664
4	6	672
5	1	1250
6	2	1518
7	3	1613
8	4	1730
9	5	1852
Total		10 603

<b>Nombre de personnes âgées de 16 ans et plus, membres des ménages dont l'entretien est accepté pour la base de données, pour lesquels un entretien personnel est achevé,</b> Ventilation par sous-échantillon		
Nombre d'années restantes dans le panel	DB075 (sous-échantillon)	Nombre de personnes
1	9	1282
2	8	1200
3	7	1250
4	6	1300
5	1	2403
6	2	2904
7	3	3095
8	4	3245
9	5	3446
Total		20 125

### 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 16,9 %.

Adresses contactées avec succès DB120=11	12698
Total DB120	13214
Adresse inexistante DB120=23	454
Adresses validées sélectionnées	12760
<b>Taux de contact à l'adresse Ra</b>	99,5 %
Entretiens de ménages achevés et acceptés pour la base de données DB135=1	10 603
Ménages éligibles aux adresses contactées	12 698
Total DB130	
<b>Proportion d'entretiens de ménages achevés et acceptés pour la base de données Rh</b>	83,5 %
<b>Taux de non-réponse des ménages NRh (<math>= [1 - (Ra * Rh)] * 100</math>)</b>	16,9 %

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

Nombre entretiens individuels achevés RB250 = 11	20 125
Nombre d'individus éligibles dans les ménages dont les entretiens ont été achevés et acceptés pour la base de données RB245 = 1	20 232
Rp	99,5 %
<b>Taux de non-réponse individuelle NRp</b>	0,5 %
<b>Taux global de non-réponse individuelle *NRp</b>	17,3 %

On obtient ainsi un taux global de non-réponse individuelle de 17,3 % très proche du taux de non-réponse des ménages à 16,9 %.



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

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

Nombre d'années restantes dans le panel	DB075	Enregistrement du contact à l'adresse			Total :
		Adresse contactée	Impossible de localiser l'adresse	Adresse inexistante	
1	9	811 100,0 %	0 0,90%	0 0,0 %	811
2	8	764 99,5 %	4 0,5 %	0 0,0 %	768
3	7	802 99,9 %	1 0,1 %	0 0,0 %	803
4	6	787 99,5 %	4 0,5 %	0 0,0 %	791
5	1	1586 99,9 %	1 0,1 %	0 0,0%	1587
6	2	1695 99,8 %	4 0,2%	0 0,0%	1699
7	3	1797 99,9 %	2 0,1%	0 0,0%	1799
8	4	1925 99,8 %	3 0,2 %	0 0,0%	1928
9	5	2531 83,6 %	43 1,4 %	454 15,0 %	3028
<b>Total</b>		12698 96,1 %	62 0,5 %	454 3,4 %	13214

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

DB130 Résultat du questionnaire "ménage"							Total
Nombre d'années restantes dans le panel	DB075	Questionnaire rempli	Refus de coopérer	Ensemble du ménage absent	Ménage incapable de répondre	Autres raisons	
1	9	678 83,6 %	78 9,6 %	18 2,2 %	18 2,2 %	19 2,3 %	811
2	8	628 82,2 %	81 10,6 %	25 3,3 %	10 1,3 %	20 2,56%	764
3	7	665 82,9 %	86 10,2 %	12 1,5 %	22 2,7 %	17 2,1 %	802
4	6	673 85,5 %	56 7,1 %	27 3,4 %	12 1,5 %	19 2,4 %	787
5	1	1253 79,0 %	240 15,1 %	25 1,6 %	30 1,9 %	38 2,4 %	1586
6	2	1523 89,9 %	78 4,6 %	32 1,9 %	26 1,5%	36 2,1 %	1695
7	3	1614 89,8 %	69 3,8 %	44 1,9 %	24 1,3 %	56 3,1 %	1797
8	4	1735 90,1 %	79 4,1 %	37 1,9 %	29 1,5 %	45 2,3 %	1925
9	5	1854 73,3 %	340 13,4 %	208 8,2 %	93 3,7 %	36 1,4 %	2531
<b>Total</b>		10623 83,7 %	1107 8,7 %	418 3,3 %	264 2,1 %	286 2,3 %	12698

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

Nombre d'années dans le panel	DB135	Acceptation de l'entretien		Total
	DB075	1 : Entretien accepté pour la base de données	2 : Entretien rejeté pour la base de données	
1	9	677 99,8 %	1 0,2 %	678
2	8	627 99,8 %	1 0,2 %	628
3	7	664 99,8 %	1 0,2 %	665
4	6	672 99,8 %	1 0,2 %	673
5	1	1250 99,8 %	3 0,2 %	1253
6	2	1518 99,7 %	5 0,3 %	1523
7	3	1613 99,9 %	1 0,1 %	1614
8	4	1730 99,7 %	5 0,3 %	1735
9	5	1852 99,9 %	2 0,1 %	1854
<b>Total</b>		10603 99,8 %	20 0,2 %	10623

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

	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Ménages</i>			
Revenu disponible total (HY020)	100.0	0.8	96.4
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et des pensions au survivant (HY022)	99.4	9.5	82.8
Revenu disponible des ménages avant transferts sociaux (HY023)	98.3	4.4	80.9
Revenus tirés de la location d'une propriété ou de terres (HY040N)	13,7	1,9	2,0
Allocation famille/enfant ( HY050N)	21.8	-	7.7
Exclusion sociale non classée ailleurs (HY060N)	7.0	2.9	0.7
Aides au logement (HY070N)	24.0	2.2	12.5
Transferts interménages perçus régulièrement en espèces (HY080N)	3.7	0.8	-
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	89.7	0.3	2.1
Revenus perçus par les personnes de moins de 16 ans (HY110N)	2.0	-	-
Impôts réguliers sur la fortune (HY120N)	1.7	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	6.2	4.1	1.5
Remboursements consécutifs à des ajustements fiscaux (HY145N)	87.8	9.1	64.9

	% de personnes de plus de 16 ans...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Individus</i>			
Salaire ou revenu assimilé (PY010N)	56.7	3.7	0.6
Avantages en nature (PY020N)	12.5	-	-
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4.8	0.2	0.3
Allocations chômage(PY090N)	8.7	0.8	7.8
Retraites (PY100N)	29.0	-	9.3
Allocations au survivant (PY110N)	0.5	-	0.2
Indemnités de maladie (PY120N)	12.5	-	-
Pensions d'invalidité (PY130N)	3.3	-	0.1
Allocations d'éducation (PY140N)	1.8	-	1.8

### 2.3.3.6 Nombre total des non-réponses

La non-réponse se décompose en non-réponse partielle (non-réponse à une question) et en non-réponse totale. La non-réponse au niveau ménage (cf. partie 2.3.3.2) concerne 2 694 ménages, soit 17,5 % des ménages.

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

	Taille d'échantillon obtenu (nb de personnes prises en compte dans le calcul)	Non réponse partielle	Hors champ	Non-réponse totale des individus	Taux de non réponse individuelle
Taux de pauvreté après transferts selon l'âge et le sexe	25 611	0	0	0	0 %
Taux de pauvreté après transferts selon le statut professionnel	19 849	0	5 655 <sup>1</sup>	107 <sup>2</sup>	0.5%
Taux de pauvreté après transferts selon le type de ménage	25 611	0	0	0	0 %
Taux de pauvreté après transferts selon le statut d'occupation du logement	25 611	0	0	0	0 %
Taux de pauvreté après transferts selon l'intensité d'emploi	22 176	0	3 147 <sup>3</sup>	288 <sup>4</sup>	1.3 %
Médiane du revenu disponible par unité de consommation - après transferts sociaux	25 611	0	0	0	0 %
Seuil de pauvreté - après transferts sociaux - personnes seules	25 611	0	0	0	0 %
Seuil de pauvreté - après transferts sociaux - 2 adultes, 2 enfants	25 611	0	0	0	0 %
Inégalité de la distribution de revenu - Rapport du partage de revenus S80/S20	25 611	0	0	0	0 %
Écart moyen relatif du taux de pauvreté	3 189	0	22 422 <sup>4</sup>	0	0 %
Dispersion autour du seuil de pauvreté	25 611	0	0	0	0 %
Taux de pauvreté avant transferts sociaux autres que retraites	25 611	0	0	0	0 %
Taux de pauvreté avant tout transfert social	25 611	0	0	0	0 %
Coefficient de Gini	25 611	0	0	0	0 %
Revenu disponible par uc moyen	25 611	0	0	0	0 %

<sup>1</sup> dont 5 379 enfants de moins de 16 ans et 276 individus qui n'ont pas gardé le même statut professionnel plus de la moitié de l'année.

<sup>2</sup> 107 individus n'ont pas rempli de questionnaire individuel, leur statut d'occupation est donc inconnu.

<sup>3</sup> 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> 288 individus appartenant à un ménage répondant dont un individu n'a pas rempli de questionnaire individuel.

<sup>4</sup> 22422 non-pauvres.

## 2.4 Mode de collecte des données

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,5 %. En effet, les ménages répondants ont tous rempli au moins un questionnaire individuel et les proxys sont autorisés.

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

RB250 Statut des données					
Nombre d'années restantes dans le panel	DB075		Information collectée par enquête	Pas de contact	Total
1	9	Nombre Pourcentage	1282 99,5 %	7 0,35%	1289
2	8	Nombre Pourcentage	1200 98,9 %	13 1,1 %	1213
3	7	Nombre Pourcentage	1250 99,5 %	6 0,5 %	1256
4	6	Nombre Pourcentage	1300 99,2 %	11 0,8 %	1311
5	1	Nombre Pourcentage	2403 99,5 %	11 0,5 %	2414
6	2	Nombre Pourcentage	2904 99,4 %	16 0,6 %	2920
7	3	Nombre Pourcentage	3095 99,7 %	10 0,3 %	3105
8	4	Nombre Pourcentage	3245 99,6 %	14 0,4 %	3259
9	5	Nombre Pourcentage	3446 99,5 %	19 0,5 %	3465
	Total	Nombre Pourcentage	20125 99,5 %	107 0,5%	20232

## 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 27,5.

RB260 Type d'entretien				
Nombre d'années dans le panel	DB075		2	5
			Face à face CAPI	Proxy
1	9	Nombre Pourcentage	903 70,4 %	379 29,6 %
2	8	Nombre Pourcentage	856 71,3 %	344 28,7 %
3	7	Nombre Pourcentage	900 72,0 %	350 28,0 %
4	6	Nombre Pourcentage	939 72,2 %	361 27,8 %
5	1	Nombre Pourcentage	1713 71,3 %	690 28,7 %
6	2	Nombre Pourcentage	2097 72,2 %	807 27,8 %
7	3	Nombre Pourcentage	2248 72,6 %	847 27,4 %
8	4	Nombre Pourcentage	2333 71,9 %	912 28,1 %
9	5	Nombre Pourcentage	2592 75,2 %	854 24,8 %
	Total	Nombre Pourcentage	14581 72,5	5544 27,5

### 2.5 Durée de l'entretien

La durée moyenne de l'entretien pour la collecte 2009 est de 50 minutes (55 minutes pour les entrants et 49 minutes pour les réinterrogés).

### 2.6 Les imputations

L'enquête réalisée en 2008 est différente des précédentes dans le mode de collecte des revenus. En effet, à de très rares exceptions 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, dans le cas où les adresses ne correspondent 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.

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 ceux-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.

### **Salaires 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 manquante, 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 élevés 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**

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, 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 est un revenu trimestriel que nous ne connaissons pas. Le RMI imputé est égal à la différence entre le plafond du RMI et les revenus de l'année. 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, un étude sur l'impact de l'intéressement du RMI 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 2002. L'estimation porte sur les logements locatifs du parc privé hors loi de 1948. Les variables explicatives sont les caractéristiques du logement (surface, confort, sanitaires, équipement, état du logement ...) et de localisation (tranche de taille d'agglomération, zone climatique, typologie socio-économique de Nicole Tabard ...). Deux équations distinctes sont estimées, l'une pour les appartements (variable expliquée : le logarithme du loyer au 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.

---

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

## 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	2 057 643	1166762	1 327 906	3 313 533	3 912 269	3 313 533	3 068 758
Moyenne	703	694	675	641	480	337	480	521
Minimum	57	122	115	87	32	9	32	1
Q5	260	273	281	278	230	163	230	230
Q10	314	325	332	337	269	200	269	280
Q25	425	421	432	401	345	255	345	387
Q50	580	581	588	551	430	313	430	490
Q75	855	850	809	766	552	400	552	603
Q90	1 183	1 160	1 081	1 034	691	500	691	790
Q95	1 438	1 464	1 376	1 283	827	578	827	916
Maximum	10 438	3 387	3 968	5 479	3 951	1 600	3 951	3 250

## 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	8 879 848	3 969 054	4 599 352	820 293	1 152 399	1 504 239	1 618 574
Moyenne	482	477	644	631	350	373	501	543
Minimum	31	37	49	100	37	17	36	0
Q5	158	153	260	271	159	124	189	200
Q10	201	193	329	337	197	194	260	282
Q25	293	279	452	446	267	281	352	376
Q50	415	408	590	593	344	363	462	504
Q75	581	581	767	760	418	450	605	650
Q90	799	797	997	988	487	567	760	812
Q95	986	953	1 188	1 114	540	640	875	990
Maximum	7 384	4 218	5 700	3 018	2 251	1 250	3 160	6 254

### 3 COMPARABILITÉ

#### 3.1 Concepts et définitions de base

##### *Population de référence*

La population de référence de l'enquête est constituée par l'ensemble des ménages ordinaires (hors institutions) dont la résidence principale se situe en France métropolitaine. En 1999, 2,2 % de la population de la France métropolitaine vivait en collectivité, en institution ou était sans domicile fixe<sup>5</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 2009. L'année de référence est 2008. On collecte au niveau individuel et ménage les revenus perçus au cours de l'année 2008. On collecte au niveau du ménage les impôts payés en 2008 au titre des revenus perçus au cours de l'année 2007. Les cotisations sociales sont relatives aux revenus perçus en 2008.

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

On considère le montant de l'impôt de solidarité sur la fortune payé en 2008. Il est relatif au patrimoine détenu au 1<sup>er</sup> janvier 2008.

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

Les variables actuelles (par exemple les variables relatives au coût du logement) portent sur la période de collecte, c'est-à-dire les mois de mai et juin 2009, ce qui rajoute une troisième période de référence après l'année 2008 (revenus déclarés) et l'année 2007 (assiette des impôts sur le revenu).

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

La collecte SILC s'est déroulée du 9 mai au 25 juin 2009.

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

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

Le codage en ISCO se fait à partir du code des professions et catégories socioprofessionnelles (PCS) à quatre chiffres, de la nomenclature d'activité française en 712 postes et du nombre de salariés de l'établissement. Dans 95 % des cas, la matrice de passage élaborée pour l'enquête Emploi permet de coder la profession de l'enquêté en un code ISCO à deux positions. Sinon, la codification en ISCO a

---

<sup>5</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*.

dû se faire par imputation statistique. Le code ISCO est alors affecté aléatoirement en fonction de la répartition du code ISCO par PCS à deux chiffres dans l'enquête Emploi du deuxième trimestre de 2007.

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

### **3.2 Composantes du revenu**

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

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

#### **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 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 effectué 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 déclarés 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 alimentés par déduction.

La variable PY100N concernent, 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 ceux 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 ainsi estimés la contribution sociale de solidarité (CSG) et la contribution de remboursement de la dette sociale (CRDS) non déductibles, la CSG déductible, les cotisations du régime général (maladie, vieillesse, veuvage, chômage et famille), les cotisations aux retraites complémentaires salariées (reçues par l'Agirc) et cadres (reçues par l'Arrco), le versement transport, la cotisation au Fond national d'aide au logement, la taxe d'apprentissage, la participation des employeurs à la formation professionnelle. Afin de contrôler les cotisations obtenues, des cas types ont été testés sur le logiciel de l'Union de recouvrement des cotisations de sécurité sociale et d'allocations familiales (URSSAF) donnant des résultats satisfaisants.

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

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



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

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

**Loyers imputés (HY030)**

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

**Prestations familiales (HY050) et allocations logement (HY070)**

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

**Cotisations à un régime de retraite privé (PY035), valeur des biens produits pour la consommation propre (PY070), indemnités 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 barèmes jusqu'en 2005 puis collectées auprès des organismes sociaux depuis 2006.

Nous avons aussi mobilisé les statistiques de la Caisse nationale des Allocations familiales. Enfin certains dossiers sont venus compléter cette aide extérieure, entre autres le dossier sur les revenus sociaux 2008 publié par la Direction de la recherche, des études, de l'évaluation et des statistiques.

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

Année de revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Salaires (milliards d'€)	505	527	527	553
Bénéficiaires (millions)	28.4	27.8	26.8	27.1
Moyenne	17793	18 957	19 680	20 411
Minimum	3	10	1	1
D10	2664	3 408	3 660	3 927
D20	7007	7 904	8 260	8 846
D30	11137	12 160	12 676	13 210
D40	14408	14 923	15 372	15 908
D50	15993	17 000	17 532	18 179
D60	18217	19 306	19 849	20 453
D70	21009	22 190	22 653	23 365
D80	24628	26 063	26 630	27 600
D90	31698	33 321	34 161	35 465
Maximum	594 622	360 554	2 282 103	875 204

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 exclues de l'ERFS alors qu'elles sont couvertes par SRCV.

#### 4.1.2 Les allocations chômage (PY090N)

	EU-SILC		ERFS	
	2007	2008	2007	2008
Revenus				
Chômage (milliards d'€)	25.6	24,2	24,28	23,14
Bénéficiaires (millions)	4.63	4.37	4,22	4.05
Moyenne	5 586	5 535	5 751	5 720
Minimum	1	22	1	1
D10	651	583	634	651
D20	1191	1 170	1 337	1 363
D30	1922	2 020	2 243	2 110
D40	2806	2 833	3 112	3 062
D50	3790	3 970	4 145	4 136
D60	5094	5 191	5 328	5 282
D70	6157	6 382	6 455	6 502
D80	8367	8 475	8 736	8 671
D90	11144	11 206	12 178	11 857
Maximum	433 555	240 864	113 039	330 013

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 du salaire SRCV.

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

Année de revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Retraites (milliards d'€)	230	242	212	221
Bénéficiaires (millions)	14,9	15,0	15,0	15,2
Moyenne	15 483	16 176	14 115	14 525
Minimum	22	86	4	22
D10	4230	4 322	2 738	2 815
D20	7639	7 954	5 362	5 608
D30	9857	10 171	8 149	8 343
D40	12058	12 494	10 243	10 525
D50	14030	14 586	12 547	12 904
D60	16099	16 644	14 817	15 179
D70	18396	19 155	17 273	17 736
D80	21949	22 630	20 768	21 309
D90	27128	28 148	26 399	27 164
Maximum	247 768	155 593	234 108	315 199

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>6</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 l'enquête).

<sup>6</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.

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

Année de revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Revenus d'indépendants (milliards d'€)	53,1	51,7	56,1	56,9
Bénéficiaires (millions)	2,33	2,45	2,00	2,01
Moyenne	22 780	21 074	28 109	28 315
Minimum	2	1	1	1
D10	1200	1 091	1 432	1 383
D20	3500	2 362	3 824	4 176
D30	6000	4 332	7 007	7 363
D40	9394	6 936	10 185	10 899
D50	12568	10 339	14 521	15 055
D60	16294	14 681	20 715	20 350
D70	21000	20 000	29 228	28 319
D80	29500	29 000	41 210	41 271
D90	49469	47 010	66 610	66 959
Maximum	1 500 000	1 303 181	1 414 689	1 476 829

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	
	2007	2008	2007	2008
Sociaux (milliards d'€)	43,7	45,0	44,9	45,5
Bénéficiaires (millions)	11,4	11,7	12,1	11,1

Les bénéficiaires se comptent en ménages.

### 4.1.5 Prestations familiales (HY050N),

Année de revenus	EU-SILC		Cnaf	
	2007	2008	2007	2008
Prestations familiales (milliards d'€)	24,9	24,6	26,0	26,8
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 2008 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2008.

#### 4.1.5.1 Aides au logement (HY070N)

Année de revenus	EU-SILC		Cnaf	
	2007	2008	2007	2008
Allocations logements (milliards d'€)	12.4	13.6	13.2	14.2
Bénéficiaires (millions)	6.8	7.0		

La Cnaf ne publie que le nombre de bénéficiaires au 31 décembre 2008 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2008.

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

Année de revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Impôts (milliards d'€)	54.6	50.5	58.5	57.2

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	9 753 721
Moyenne	400	433
Q0	14	1
Q5	175	169
Q10	211	220
Q25	269	281
Q50	356	383
Q75	475	507
Q90	617	650
Q95	728	750
Q100	3 951	2930

#### 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.

Année des revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Revenus du patrimoine financier (milliards d'€)	92.9	104.3	79.6	86.1
Bénéficiaires (millions)	22.3	24.0	24.2	25.3
Moyenne	4162	4354	3295	3407
Minimum	1	1	1	1
D10	45	47	41	55
D20	121	136	115	146
D30	247	277	236	284
D40	461	486	412	470
D50	778	825	679	742
D60	1324	1310	1073	1129
D70	2320	2083	1728	1794
D80	4292	3734	3076	2849
D90	9527	8254	6658	5972
Maximum	424560	622000	1768651	757006

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

Année des revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Revenus disponible des ménages (milliards d'€)	930.70	983.42	870,01	917.63
Bénéficiaires (millions)	26,7	27.0	26,8	27.15
Moyenne	34 829	36 450	32 462	33 803
Minimum	249	10	- 67 626	-61 274
D10	13 142	13 765	11 725	12 035
D20	16 930	17 630	15 495	16 024
D30	20 535	21 335	18 994	19 716
D40	24 601	25 537	22 894	23 712
D50	28 811	29 964	27 142	28 091
D60	33 466	34 755	31 791	33 010
D70	39 057	10 502	37 269	38 693
D80	46 807	48 235	44 655	46 161
D90	60 605	62 619	57 224	58 983
Maximum	866 696	993 412	1 438 606	1 415 623

Revenus	EU-SILC		ERFS	
	2007	2008	2007	2008
Niveau de vie des individus (milliards d'€)	597.60	635.80	557,38	587.03
Bénéficiaires (millions)	60,33	60.76	60,55	60.19
Moyenne	22 360	23 550	20 859	22 085
Minimum	-272 643	-33 656	- 67 626	-696
D10	10 630	11 071	9 709	10 523
D20	13 209	13 509	12 208	13 120
D30	15 298	15 659	14 198	15 198
D40	17 137	17 766	16 077	17 099
D50	18 962	19 737	18 031	18 986
D60	21 016	21 983	20 216	21 122
D70	23 945	24 727	22 864	23 791
D80	27 695	28 990	26 533	27 632
D90	35 809	37 308	33 716	35 549
Maximum	577 797	622 275	939 683	943 749

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, auto-consommation), 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>7</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>8</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. Bien qu'il soit possible de faire une estimation par maximisation de la vraisemblance il nous a semblé préférable de ne mener l'estimation que sur le groupe des répondants en clair avec une procédure d'estimation robuste (dans l'état actuel de nos connaissances nous ne savions pas mener une estimation robuste utilisant ces deux informations). Nous faisons implicitement l'hypothèse, qu'au sein des deux groupes, la loi (conditionnelle à l'observation des covariables) du montant est la même. Il est possible d'imaginer la présence de biais de sélection, mais nous avons pris le parti de les négliger. Les biais de sélection sont d'autant plus faibles que nous avons utilisé le conditionnement par des observations de covariables.

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

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

<sup>7</sup> Pour le risque quadratique calculé avec l'aléa de modèle.

<sup>8</sup> Dans certains cas seulement la différence peut être négligeable.

Si aucune information n'est disponible sur la valeur du montant manquant, la simulation est faite dans la loi log-normale. Lorsqu'une plage de valeurs est disponible, ou reconstituée à partir de la connaissance d'autres variables, la simulation se fait dans la loi précédente conditionnelle au fait que la variable est dans la plage de valeurs prescrite. Dans ce second cas la simulation revient à la simulation de lois normales tronquées. La simulation dans la loi normale tronquée est effectuée par inversion de la fonction de répartition dès que cela est possible. Il est arrivé parfois que des problèmes numériques (plage loin de la valeur prédite par le modèle...) nous fassent utiliser une méthode d'acceptation/rejet avec optimisation de la probabilité d'acceptation. Cette méthode est une variante de la méthode de la macro « résidus simulés » mais est beaucoup plus efficace<sup>9</sup>.

---

<sup>9</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.