

**Rapport intermédiaire sur la qualité
des données françaises EU-SILC 2005**

TABLE DES MATIÈRES

1	Indicateurs transversaux communs de l'Union européenne	5
1.1	Indicateurs communautaires transversaux communs fondés sur la dimension transversale des statistiques EU-SILC	5
1.1.1	Taux de pauvreté après transferts sociaux	5
1.1.1.1	Taux de pauvreté selon l'âge et le sexe	5
1.1.1.2	Taux de pauvreté selon le statut professionnel le plus fréquent et le sexe	6
1.1.1.3	Taux de pauvreté selon le type de ménage	7
1.1.1.4	Taux de pauvreté selon le statut d'occupation du logement	9
1.1.1.5	Taux de pauvreté selon le type de ménage et l'intensité d'emploi	10
1.1.1.6	Seuil de pauvreté (valeurs indicatives)	10
1.1.2	Inégalité de la distribution des revenus Q80/Q20, ratio interquintile	11
1.1.3	Écart de risque de pauvreté	11
1.1.4	Dispersion autour du seuil de pauvreté	11
1.1.5	Taux de pauvreté avant transferts	12
1.1.5.1	Taux de pauvreté avant tout transfert social selon l'âge et le sexe	12
1.1.5.2	Taux de pauvreté avant transferts sociaux autres que retraites selon l'âge et le sexe	12
1.1.6	Inégalité de la distribution des revenus : coefficient de Gini	12
1.2	Autres indicateurs	12
1.2.1	Niveau de vie moyen	12
1.2.2	Écart de rémunération entre les sexes non ajusté	13
2	Précision	14
2.1	Plan d'échantillonnage	14
2.1.1	Type de plan d'échantillonnage	14
2.1.2	Unités d'échantillonnage	14
2.1.3	Critères de stratification et de sous-stratification	14
2.1.4	Taille de l'échantillon et critères d'attribution	15
2.1.5	Systèmes de sélection de l'échantillon	15
2.1.6	Répartition de l'échantillon au cours du temps et renouvellement de l'échantillon : groupes de rotation	15
2.1.8	Pondération	15
2.1.8.1	Effet du plan d'échantillonnage (entrants)	16
2.1.8.2	Ajustements pour non-réponse	16
2.1.8.3	Partage des poids	17
2.1.8.4	Ajustements aux données extérieures (niveau, variables utilisées et sources)	18
2.1.8.5	Pondération transversale finale	19
2.1.9	Substitutions	20
2.2	Erreurs d'échantillonnage	21

2.3	Erreurs non dues à l'échantillonnage.....	24
2.3.1	Erreurs dans la base de sondage et la couverture.....	24
2.3.2	Erreurs de mesure et de traitements.....	25
2.3.3	Erreurs de non-réponse.....	29
2.3.3.1	Taille de l'échantillon obtenue.....	29
2.3.3.2	Non-réponse des unités.....	30
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.....	31
2.3.3.4	Ventilation des unités remplacées.....	33
2.3.3.5	Non-réponse à certaines questions.....	34
2.3.3.6	Nombre total des non-réponses.....	35
2.4	Mode de collecte des données.....	36
2.5	Durée de l'entretien.....	37
2.6	Les imputations.....	37
3	Comparabilité.....	42
3.1	Concepts et définitions de base.....	42
3.2	Composantes du revenu.....	43
3.2.1	Les différences entre les définitions nationales et les définitions types des statistiques EU-SILC.....	43
3.2.2	La source ou la procédure utilisée pour la collecte des variables du revenu.....	46
3.2.3	La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes.....	46
3.2.4	La méthode utilisée pour obtenir les variables cibles du revenu sous la forme requise (c'est-à-dire en tant que valeur brute).....	46
4	Cohérence.....	48
4.1	Comparaison des variables cibles avec des sources extérieures.....	48
4.1.1	Salaire et revenu assimilé (PY010N).....	48
4.1.2	Les allocations chômage (PY090N).....	49
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et de survie (PY110N).....	49
4.1.4	Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante.....	50
4.1.5	Revenus sociaux (HY050G, HY060G et HY070G).....	50
4.1.5.1	Prestations familiales (HY050N).....	50
4.1.5.2	Aides au logement (HY070N).....	51
4.1.6	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation).	51
4.1.7	Loyers.....	51
4.1.8	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N).....	52
4.1.9	Revenu disponible du ménage et niveau de vie des individus.....	52
	Annexe méthodologique.....	54

1 INDICATEURS TRANSVERSAUX COMMUNS DE L'UNION EUROPEENNE

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

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

1.1.1 Taux de pauvreté après transferts sociaux

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

1.1.1.1 Taux de pauvreté selon l'âge et le sexe

	Total			0 - 15			0 - 64			16+			16 - 64		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	13,0	13,7	12,3	14,1	14,7	13,6	12,3	12,8	11,8	12,7	13,5	11,9	11,8	12,3	11,3
Non pauvres	87,0	86,3	87,7	85,9	85,3	86,4	87,7	87,2	88,2	87,3	86,5	88,1	88,2	87,7	88,7

	16 - 24			25 - 49			50 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	17,8	20,2	15,5	10,6	10,9	10,3	10,1	9,9	10,3	16,5	17,5	15,0
Non pauvres	82,2	79,8	84,5	89,4	89,1	89,7	89,9	90,1	89,7	83,5	82,5	85,0

Sont exclus du champ les enfants nés en 2005.

Effectifs non pondérés

	Total			0 - 15			0 - 64			16+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	24152	12448	11704	5204	2536	2668	20684	10513	10171	18948	9912	9036
Pauvres	3257	1763	1494	816	411	405	2708	1429	1279	2441	1352	1089
Non pauvres	20895	10685	10210	4388	2125	2263	17976	9084	8892	16507	8560	7947

	16 - 64			16 - 24			25 - 49			50 - 64		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	15480	7977	7503	2917	1484	1433	8129	4232	3897	4434	2261	2173
Pauvres	1892	1018	874	530	295	235	911	496	415	451	227	224
Non pauvres	13588	6959	6629	2387	1189	1198	7218	3736	3482	3983	2034	1949

	65+		
	Total	F	H
Total	3468	1935	1533
Pauvres	549	334	215
Non pauvres	2919	1601	1318

Répartition de la population totale selon le sexe

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

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

	0 - 15	0 - 64	16+	16 - 64	16 - 24	25 - 49	50 - 64	65+	Total
Total	19,1	83,6	80,9	64,5	11,7	34,8	18,1	16,4	100,0
F	18,1	81,5	81,9	63,4	11,3	34,2	17,9	18,5	100,0
H	20,1	85,9	79,9	65,7	12,0	35,4	18,3	14,1	100,0

Répartition de la population pauvre selon le sexe

Total	F	H
100,0	54,2	45,8

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

	0 - 15	0 - 64	16+	16 - 64	16 - 24	25 - 49	50 - 64	65+	Total
Total	20,7	79,3	79,3	58,5	16,0	28,4	14,1	20,7	100,0
F	19,4	76,3	80,6	56,9	16,7	27,2	13,0	23,7	100,0
H	22,3	82,7	77,7	60,4	15,2	29,8	15,4	17,3	100,0

1.1.1.2 Taux de pauvreté selon le statut professionnel le plus fréquent et le sexe

	Total			En emploi			Sans emploi : total			Sans emploi : chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	12,6	13,4	11,8	6,1	5,1	7,0	19,5	20,4	18,3	29,3	25,3	33,7
Non pauvres	87,4	86,6	88,2	93,9	94,9	93,0	80,5	79,6	81,7	70,7	74,7	66,3

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Pauvres	13,5	13,7	13,3	24,9	26,4	21,7
Non pauvres	86,5	86,3	86,7	75,1	73,6	78,3

Sont exclus du champ les personnes de moins de 16 ans au 1^{er} janvier de l'année de l'enquête et les individus qui n'ont pas gardé le même statut d'occupation pendant plus de la moitié de l'année.

Effectifs non pondérés

	Total			En emploi			Sans emploi : total			Sans emploi : chômeurs		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	18510	9690	8820	9519	4464	5055	8991	5226	3765	1027	547	480
Pauvres	2366	1312	1054	582	229	353	1784	1083	701	313	141	172
Non pauvres	16144	8378	7766	8937	4235	4702	7207	4143	3064	714	406	308

	Sans emploi : retraités			Sans emploi : autres inactifs		
	Total	F	H	Total	F	H
Total	4197	2103	2094	3767	2576	1191
Pauvres	539	272	267	932	670	262
Non pauvres	3658	1831	1827	2835	1906	929

Répartition de la population totale 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	45,7	54,3	5,8	25,3	23,2
H	100,0	57,1	42,9	5,7	25,4	11,8
Total	100,0	51,2	48,8	5,8	25,3	17,7

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	17,4	82,6	11,0	25,9	45,8
H	100,0	33,8	66,2	16,2	28,4	21,6
Total	100,0	24,7	75,3	13,3	27,0	34,9

1.1.1.3 Taux de pauvreté selon le type de ménage

	Sans enfants : total	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Sans enfants : autres	Avec enfants : total
Pauvres	13,0	19,6	8,4	12,9	10,0	13,0
Non pauvres	87,0	80,4	91,6	87,1	90,0	87,0

	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
Pauvres	25,2	8,5	9,4	19,7	15,5
Non pauvres	74,8	91,5	90,6	80,3	84,5

Effectifs non pondérés

	Sans enfants : total	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Sans enfants : autres	Avec enfants : total
Total	9933	2661	3334	2436	1502	14161
Pauvres	1247	537	274	296	140	2013
Non pauvres	8686	2124	3060	2140	1362	12148

Total	Familles monoparentales	2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
	1425	2937	5048	3169	1582
Pauvres	384	246	464	633	286
Non pauvres	1041	2691	4584	2536	1296

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

	F	H	< 65	65+
Pauvres	19,7	19,4	18,3	21,3
Non pauvres	80,3	80,6	81,7	78,7

Effectifs non pondérés

	F	H	< 65	65+
Total	1644	1017	1554	1107
Pauvres	339	198	290	247
Non pauvres	1305	819	1264	860

Répartition de la population totale selon le type de ménage

Sans enfants : total	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Sans enfants : autres	Avec enfants : total	Familles monoparentales
46,6	13,7	15,7	11,2	6,0	53,4	5,3

2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
12,9	21,3	8,9	5,1

Répartition des personnes seules selon l'âge ou le sexe

F	H	< 65	65+
61,3	38,7	57,8	42,2

Répartition de la population pauvre selon le type de ménage

Sans enfants : total	Personnes seules (total)	2 adultes, moins de 65	2 adultes, dont au moins un de 65 ou +	Sans enfants : autres	Avec enfants : total	Familles monoparentales
46,4	20,6	10,1	11,1	4,6	53,6	10,2

2 adultes, 1 enfant	2 adultes, 2 enfants	2 adultes, 3 enfants et +	Avec enfants : autres
8,4	15,4	13,4	4,6

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

F	H	< 65	65+
61,6	38,4	54,1	45,9

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

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

	Total	Locataires	Propriétaires ou logés à titre gratuit
Pauvres	13,0	20,3	9,2
Non pauvres	87,0	79,7	90,8

Effectifs non pondérés

	Total	Locataires	Propriétaires ou logés à titre gratuit
Total	24245	7729	16516
Pauvres	3271	1787	1484
Non pauvres	20974	5942	15032

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	34,5	65,5

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

Total	Locataires	Propriétaires ou logés à titre gratuit
100,0	53,8	46,2

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 16 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 enfants $W=0$	Sans enfants $0 < W < 1$	Sans enfants $W=1$	Avec enfants $W=0$	Avec enfants $0 < W < 0,5$	Avec enfants $0,5 < W < 1$	Avec enfants $W=1$
Pauvres	21,0	11,4	4,2	63,5	41,3	15,5	4,3
Non pauvres	79,0	88,6	95,8	36,5	58,7	84,5	95,7

Sont exclus du champ les ménages composés seulement de personnes de moins de 16 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 enfants $W=0$	Sans enfants $0 < W < 1$	Sans enfants $W=1$	Avec enfants $W=0$	Avec enfants $0 < W < 0,5$	Avec enfants $0,5 < W < 1$	Avec enfants $W=1$
Total	1678	2484	2706	786	881	4700	7616
Pauvres	347	250	111	509	392	749	324
Non pauvres	1331	2234	2595	277	489	3951	7292

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

Total	Sans enfants $W=0$	Sans enfants $0 < W < 1$	Sans enfants $W=1$	Avec enfants $W=0$	Avec enfants $0 < W < 0,5$	Avec enfants $0,5 < W < 1$	Avec enfants $W=1$
100,0	8,9	13,4	15,2	3,0	3,3	21,4	34,8

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

Total	Sans enfants $W=0$	Sans enfants $0 < W < 1$	Sans enfants $W=1$	Avec enfants $W=0$	Avec enfants $0 < W < 0,5$	Avec enfants $0,5 < W < 1$	Avec enfants $W=1$
100,0	15,4	12,6	5,3	15,6	11,2	27,4	12,5

1.1.1.6 Seuil de pauvreté (valeurs indicatives)

En 2004, le seuil de pauvreté s'élève à 9 562 euros par an. Une personne seule est pauvre si elle a un revenu disponible inférieur à 797 euros par mois et un couple avec deux enfants de moins de 14 ans, s'il a un revenu disponible inférieur à 20 081 euros (soit 1 673 euros par mois).

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

Le cinquième de la population le plus aisé possède 4,0 fois plus de niveaux de vie que le cinquième de la population le plus pauvre.

1.1.3 Écart de risque de pauvreté

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

Total			0 - 15	16+			16 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
16,6	16,5	16,7	14,8	16,9	16,7	16,9	17,4	16,7	18,8	14,8	17,1	13,0

Effectifs non pondérés

Total			0 - 15	16+			16 - 64			65+		
Total	F	H	Total	Total	F	H	Total	F	H	Total	F	H
3271	1772	1499	830	2441	1352	1089	1892	1018	874	549	334	215

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 dispersion des niveaux de vie autour du seuil de pauvreté.

	Total	F	H
<i>40 % de la médiane</i>			
Pauvres	2,6	2,8	2,5
Non pauvres	97,4	97,2	97,5
<i>50 % de la médiane</i>			
Pauvres	6,4	6,7	6,1
Non pauvres	93,6	93,3	93,9
<i>70 % de la médiane</i>			
Pauvres	22,0	23,2	20,7
Non pauvres	78,0	76,8	79,3

Effectifs non pondérés

	total	F	H
Total	24245	12500	11745
<i>40 % de la médiane</i>			
Pauvres	664	364	300
Non pauvres	23581	12136	11445
<i>50 % de la médiane</i>			
Pauvres	1608	866	742
Non pauvres	22637	11634	11003
<i>70 % de la médiane</i>			
Pauvres	5411	2925	2486
Non pauvres	18834	9575	9259

1.1.5 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.

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

	Total			0 - 15			16+			16 - 64			65+		
	Total	H	F	Total	H	F	Total	H	F	Total	H	F	Total	H	F
Pauvres	44,8	47,4	42,0	35,5	37,1	34,0	47,0	49,7	44,0	34,7	36,4	32,9	95,5	95,5	95,6
Non pauvres	55,2	52,6	58,0	64,5	62,9	66,0	53,0	50,3	56,0	65,3	63,6	67,1	4,5	4,5	4,4

Sont exclus du champ les personnes nées en 2005.

Effectifs non pondérés

	Total			0 - 15			16+			16 - 64			65+		
	Total	H	F	Total	H	F	Total	H	F	Total	H	F	Total	H	F
Total	24152	12448	11704	5204	2536	2668	18948	9912	9036	15480	7977	7503	3468	1935	1533
Pauvres	10657	5739	4918	1962	978	984	8695	4761	3934	5411	2928	2483	3284	1833	1451
Non pauvres	13495	6709	6786	3242	1558	1684	10253	5151	5102	10069	5049	5020	184	102	82

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

	Total			0 - 15			16+			16 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Pauvres	26,0	27,0	25,0	34,2	35,9	32,6	24,1	25,1	23,1	24,9	25,8	24,0	20,9	22,5	18,8
Non pauvres	74,0	73,0	75,0	65,8	64,1	67,4	75,9	74,9	76,9	75,1	74,2	76,0	79,1	77,5	81,2

Sont exclus du champ les personnes nées en 2005.

Effectifs non pondérés

	Total			0 - 15			16+			16 - 64			65+		
	Total	F	H	Total	F	H	Total	F	H	Total	F	H	Total	F	H
Total	24152	12448	11704	5204	2536	2668	18948	9912	9036	15480	7977	7503	3468	1935	1533
Pauvres	6586	3501	3085	1888	943	945	4698	2558	2140	4001	2133	1868	697	425	272
Non pauvres	17566	8947	8619	3316	1593	1723	14250	7354	6896	11479	5844	5635	2771	1510	1261

1.1.6 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 27,8 % en 2004.

1.2 Autres indicateurs

1.2.1 Niveau de vie moyen

En 2004, le niveau de vie des individus s'élève en moyenne à 18 199 euros par unité de consommation, soit 1 517 euros par mois.

1.2.2 *Écart de rémunération entre les sexes non ajusté*

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

2 PRÉCISION

2.1 Plan d'échantillonnage

2.1.1 Type de plan d'échantillonnage

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

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

2.1.2 Unités d'échantillonnage

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

2.1.3 Critères de stratification et de sous-stratification

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

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

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

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

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

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

2.1.4 Taille de l'échantillon et critères d'attribution

2.1.5 Systèmes de sélection de l'échantillon

L'échantillon de logements de SILC est tiré à probabilités égales (taux f) à l'exception des logements recensés secondaires qui sont tirés avec la probabilité $f/4$ et des logements recensés vacants en strate rurale qui sont tirés au taux $f/2$. L'échantillon 2004 comprenait 16 000 logements (ventilés en 9 sous-échantillons), et le sous-échantillon entrant en 2005 en comprenait 3 000. Les sous-échantillons entrants des années à venir devraient être calibrés à 3 000 logements.

Cas de la strate rurale

Le tirage des 128 unités primaires de l'échantillon-maître parmi les 227 recensées a été effectué proportionnellement à la taille définie en nombre de résidences principales, selon un algorithme de tirage équilibré (l'équilibrage a porté sur des structures par âge et sur le revenu déclaré aux services fiscaux). Le tirage de 652 unités secondaires de l'échantillon de EU-SILC sur les 3 217 unités de l'échantillon-maître a également été proportionnel à la taille en nombre de résidences principales. Le tirage des logements est effectué au sein du groupe de communes par un algorithme systématique sur fichier trié.

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

Cas des strates urbaines jusqu'à 100 000 habitants

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

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 et renouvellement de l'échantillon : groupes de rotation

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.

En 2005, 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 2004 au moins une personne répondante et n'appartenant pas à l'échantillon sortant. Il comprenait 8 928 logements.

2.1.8 Pondération

La pondération se fait au niveau logement.

En 2005, nous avons estimé pour la première fois des poids pour un échantillon comprenant à la fois des entrants et des réinterrogés. Nous avons suivi la méthodologie préconisée par Pascal Ardilly et Pierre Lavallée (Symposium 2003, Stat Canada). Quelques problèmes pratiques ont été identifiés,

notamment en ce qui concerne la dispersion des poids et la correction de la non-réponse des réinterrogés. Il s'agira d'étudier ces aspects et d'apporter d'éventuelles améliorations lors de la prochaine vague.

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. Les pondérations initiales 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).

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 le premier type de non-réponse et non pour la non-réponse en 2005. En effet, l'information utilisée pour corriger de la « non-réponse initiale », issue du recensement et valable au niveau du logement, n'est pas disponible en réinterrogation en cas de déménagement. Nous avons donc choisi de corriger de la non-réponse de seconde interrogation en effectuant un calage généralisé avec le logiciel Calmar 2.

La correction de la non-réponse initiale pour les réinterrogés

Les modèles de non-réponse ont été estimés sur les logements de l'échantillon qui faisaient partie du champ : les enquêteurs les ont reconnus comme résidence habituelle d'au moins une personne. Nous avons différencié cinq types de modèles de non-réponse totale selon les informations disponibles dans les bases de sondage ou recueillies par l'enquêteur en 2004 :

- un pour les ménages d'une personne habitant des résidences principales au dernier recensement de la population (RP) et en 2004,
- un pour les ménages de 2 personnes des résidences principales au RP et en 2004,
- un pour les ménages de 3 personnes et plus des résidences principales au RP et en 2004,
- un pour les résidences recensées secondaires, occasionnelles ou vacantes au RP,
- un pour les logements neufs (construits après mars 1999).

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

- la région,
- la densité d'habitat (rural ou urbain),
- la caractérisation en termes d'activité du quartier (chômage, activité),
- le type d'habitat au voisinage du logement (individuel ou collectif) en 2004,
- le fait d'habiter une maison ou un appartement en 2004,
- la taille du ménage (ou 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 10 273 ménages répondants dans 37 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur des sous-populations. Le nouveau poids d'un ménage répondant dans une sous-population donnée vaut donc :

$$\frac{\text{Nombre de ménages échantillonnés dans la sous population}}{\text{Nombre de ménages répondants dans la sous-population}} \times \text{poids d'échantillonnage du ménage}$$

Selon les configurations, la nouvelle pondération augmente les poids d'échantillonnage dans un rapport allant de 1,1 à 1,9.

La correction de la non-réponse initiale pour les entrants

Les modèles de non-réponse ont été estimés sur les logements de l'échantillon qui faisaient partie du champ en 2005 : 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 2005 :

- les résidences principales au dernier recensement de la population (RP) et en 2005,
- 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,
- la caractérisation en termes d'activité du quartier (chômage, activité),
- le fait d'habiter une maison ou un appartement en 2005,
- le nombre de pièces du logement en 1999,
- l'âge de la personne de référence (plus ou moins de 60 ans) en 1999,
- le fait d'avoir un gardien ou pas en 1999.

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les 1 820 ménages répondants dans 19 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. Le nouveau poids d'un ménage répondant dans une sous-population donnée vaut donc :

$$\frac{\text{Nombre de ménages échantillonnés dans la sous population} \times \text{poids d'échantillonnage du ménage}}{\text{Nombre de ménages répondants dans la sous-population}}$$

Selon les configurations, la nouvelle pondération augmente les poids d'échantillonnage dans un rapport allant de 1,2 à 1,8.

2.1.8.3 Partage des poids

Pour les entrants, la pondération issue de la non-réponse est modifiée comme suit. Le poids après le partage des poids est égal à :

Poids après partage des poids

$$= \frac{\text{poids initial corrigé de la non-réponse} \times \text{nombre de personnes dans le ménage en 2005}}{\text{nb de personnes en 2005} - \frac{8}{9} \times \text{nombre de nouveau-nés depuis la première enquête}}$$

Pour les réinterrogés, le poids après partage des poids est égal à :

$$= \frac{\text{poids initial corrigé de la non-réponse} \times \text{nombre de personnes dans le ménage en 2004 et 2005}}{\text{nombre de personnes en 2005} - \frac{8}{9} \times \text{nombre de nouveau-nés depuis la première enquête}}$$

Faute d'informations, nous ne tenons pas compte du fait que certains individus n'auraient pas pu être interrogés en 2004 parce qu'ils vivaient à l'étranger ou en institution. Cela sera fait à partir de la collecte de 2006.

Le partage des poids modifie les poids dans un rapport allant de 0,2 à 1,8.

2.1.8.4 Ajustements aux données extérieures (niveau, variables utilisées et sources)

L'enquête, réalisée en mai 2005, est, comme toutes les enquêtes ménage de l'Insee réalisées cette année, calée sur les marges issues de l'enquête Emploi de 2004. 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 (25 476 084 ménages, nombre issu de la série annuelle du compte satellite du logement). Il a donc fallu, dans un premier temps, reconstituer cette notion de ménage-logement et déterminer la personne de référence afin de pouvoir réaliser le calage.

Le calage et la correction de la non-réponse de seconde interrogation ont été effectués avec le logiciel Calmar 2. Les poids en entrée sont ceux issus du partage des poids. Les poids de l'échantillon entrant sont divisés par 9. Nous avons utilisé la fonction de distance « logit ».

Les variables et les modalités introduites dans le calage et la correction de la non-réponse sont les mêmes. 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 31/12/2004.
- nombre de ménages par tranche de densité d'habitat : rural, unité urbaine de moins de 20 000 habitants, unité urbaine de 20 000 à 100 000 habitants, unité urbaine de plus de 100 000 habitants, Région parisienne.
- nombre de ménages par type :
 - personne seule,
 - couple sans enfant,
 - couple avec 1 enfant,
 - couple avec 2 enfants ou plus,
 - famille monoparentale,
 - autre configuration ;
- nombre d'hommes par tranche d'âge (6 modalités, des moins de 15 ans aux 76 ans et plus) ;
- nombre de femmes par tranche d'âge (6 modalités) ;
- nombre de ménages selon le diplôme de la personne de référence :
 - sans diplôme, non déclaré,
 - diplôme inférieur baccalauréat (CAP, BEPC),
 - Baccalauréat, bac+2,
 - diplôme supérieur ;
- nombre de ménages selon la catégorie socioprofessionnelle de la personne de référence (activité actuelle ou ancienne activité) :
 - agriculteurs (retraités ou non),
 - indépendants et professions libérales (retraités ou non),
 - professeurs et instituteurs actifs,
 - professeurs et instituteurs retraités,
 - employés et ouvriers actifs,
 - employés et ouvriers retraités,
 - autres ;
- nombre de ménages par région :
 - Île-de-France et bassin parisien,
 - Nord, Est et Ouest,
 - Sud-Ouest, Centre-Est et Méditerranée.

Les rapports de poids avant/après calage évoluent entre 0,45 et 2,3. Les pondérations finales des logements varient quant à elles dans un rapport allant de 1 à 62.

2.1.8.5 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 0,9 %, l'impact de cette correction est très limité, de 0,2 % à 3,4 %.

Pour les répondants au module secondaire, le poids est modifié de la même manière. L'impact de la correction est très limité, de 0,3 % à 1,5 %.

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. L'impact de la correction est compris entre -8,0 % et +19,3 %

Il existe donc quatre 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, PM005 pour les questions du module secondaire 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 calculés en utilisant le logiciel Poulpe après linéarisation.

La variance du taux de pauvreté obtenue est supérieure de 11 % 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 pourcentages)						
Total	13,0	0,4	3,2	1,11	24 152	21 759
Hommes total	12,3	0,5	3,7	1,13	12 448	11 016
Femmes total	13,7	0,5	3,4	1,07	11 704	10 938
0-15 ans	14,1	0,8	5,7	1,05	5 204	4 956
16-24 ans	17,8	0,9	5,3	0,96	2 917	3 039
25-49 ans	10,6	0,4	4,2	1,12	8 129	7 258
50-64 ans	10,1	0,6	5,5	1,02	4 434	4 347
65+ ans	16,5	0,7	4,4	1,22	3 468	2 843
16+ ans	12,7	0,4	3,0	1,12	18 948	16 918
16-64 ans	11,8	0,4	3,5	1,07	15 480	14 467
0-64 ans	12,3	0,5	3,7	1,08	20 684	19 152
Hommes 16-24 ans	15,5	1,2	7,5	1,03	1 433	1 391
Hommes 25-49 ans	10,3	0,5	4,8	1,13	3 897	3 449
Hommes 50-64 ans	10,3	0,7	6,5	1,02	2 173	2 130
Hommes 65+ ans	15,0	0,9	5,9	1,15	1 533	1 333
Hommes 16+ ans	11,9	0,4	3,4	1,16	9 036	7 790
Hommes 16-64 ans	11,3	0,4	3,9	1,12	7 503	6 699
Hommes 0-64 ans	11,8	0,5	4,2	1,11	10 171	9 163
Femmes 16-24 ans	20,2	1,2	6,1	0,95	1 484	1 562
Femmes 25-49 ans	10,9	0,5	4,7	1,08	4 232	3 919
Femmes 50-64 ans	9,9	0,7	6,7	1,08	2 261	2 094
Femmes 65+ ans	17,5	0,8	4,6	1,22	1 935	1 586
Femmes 16+ ans	13,5	0,4	3,1	1,09	9 912	9 094
Femmes 16-64 ans	12,3	0,5	3,7	1,02	7 977	7 821
Femmes 0-64 ans	12,3	0,5	4,1	1,04	10 513	10 109

Indicateurs	Valeur	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Taux de pauvreté après prestations sociales (en pourcentages)						
Actifs occupés	6,1	0,4	6,1	1,09	9 519	8 733
Chômeurs	29,3	1,9	6,3	1,07	1 027	960
Retraités	13,5	0,6	4,4	1,13	4 197	3 714
Autres inactifs	24,9	1,2	4,7	1,05	3 767	3 588
Hommes, actifs occupés	7,0	0,4	6,0	1,11	5 055	4 554
Hommes, chômeurs	33,7	2,2	6,6	1,10	480	436
Hommes, retraités	13,3	0,7	5,3	1,12	2 094	1 870
Hommes, autres inactifs	21,7	1,5	6,8	1,09	1 191	1 093
Femmes, actives occupées	5,1	0,4	8,4	1,04	4 464	4 292
Femmes, chômeuses	25,3	2,2	8,7	1,05	547	521
Femmes, retraitées	13,7	0,7	5,1	1,12	2 103	1 878
Femmes, autres inactives	26,4	1,2	4,7	1,03	2 576	2 501
Personnes seules, < 65 ans	18,3	0,9	4,8	1,13	1 554	1 375
Personnes seules, 65+ ans	21,3	1,1	5,2	1,10	1 107	1 006
Personnes seules, hommes	19,4	1,1	5,5	1,13	1 017	900
Personnes seules, femmes	19,7	0,9	4,6	1,19	1 644	1 382
Personnes seules, total	19,6	0,7	3,7	1,13	2 661	2 355
2 adultes, sans enfants, tous les deux < 65	8,4	0,6	6,8	1,05	3 334	3 175
2 adultes, sans enfants, au moins 1 65+	12,9	0,9	7,1	1,19	2 436	2 047
Autres ménages sans enfants	10,0	1,4	13,8	0,96	1 502	1 565
Familles monoparentales	25,2	2,2	8,8	0,99	1 425	1 439
2 adultes, 1 enfant	8,5	0,9	10,8	1,19	2 937	2 468
2 adultes, 2 enfants	9,4	1,0	10,6	0,99	5 048	5 099
2 adultes, 3+ enfants	19,7	3,1	15,5	1,07	3 169	2 962
Autres ménages avec enfants	15,5	4,3	27,5	0,93	1 582	1 701
Ménages sans enfants	13,0	0,4	3,2	1,15	9 933	8 637
Ménages avec enfants	13,0	0,7	5,6	1,07	14 161	13 235
Propriétaires	9,2	0,4	4,5	1,03	16 516	16 035
Locataires	20,3	0,9	4,5	1,11	7 729	6 963
Ménages sans enfants, w = 0	21,0	1,4	6,4	1,00	1 678	1 678
Ménages sans enfants, 0 < w < 1	11,4	0,7	6,3	1,04	2 484	2 388
Ménages sans enfants, w = 1	4,2	0,5	11,2	1,09	2 706	2 483
Ménages avec enfants, w = 0	63,5	4,6	7,3	1,04	786	756
Ménages avec enfants, 0 < w < 0.5	41,3	4,0	9,8	1,00	881	881
Ménages avec enfants, 0.5 < w < 1	15,5	1,2	7,6	1,08	4 700	4 352
Ménages avec enfants, w = 1	4,3	0,6	14,7	1,08	7 616	7 052
Médiane des niveaux de vie (en euros par unité de consommation)	15 937	51	0,3	1,13	24 245	21 456
Seuil de pauvreté - personnes seules (en euros)	9 562	31	0,3	1,13	24 245	21 456
Seuil de pauvreté - 2 adultes, 2 enfants (en euros)	20 081	65	0,3	1,13	24 245	21 456

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	4,0	0,1	1,8	1,23	24 245	19 711
Écart de risque de pauvreté (en pourcentages)						
Total	16,6	0,9	5,1	1,09	3 271	3 001
Hommes total	16,7	0,9	5,6	1,12	1 499	1 338
Femmes total	16,5	0,9	5,5	1,04	1 772	1 704
0-15 ans	14,8	1,4	9,7	1,07	830	776
16-64 ans	17,4	1,0	5,5	1,06	1 892	1 785
65+ ans	14,8	1,1	7,2	1,07	549	513
16+ ans	16,9	0,8	4,7	1,10	2 441	2 219
Hommes, 16-64 ans	18,8	1,0	5,4	1,04	874	840
Hommes, 65+ ans	13,0	1,0	7,8	1,14	215	189
Hommes, 16+ ans	16,9	0,9	5,4	1,10	1 089	990
Femmes, 16-64 ans	16,7	1,1	6,3	1,04	1 018	979
Femmes, 65+ ans	17,1	1,3	7,8	1,14	334	293
Femmes, 16+ ans	16,7	0,9	5,1	1,10	1 352	1 229
Taux de pauvreté selon le seuil de pauvreté (en pourcentages)						
40 % de la médiane	2,6	0,2	7,7	1,09	24 245	22 243
50 % de la médiane	6,5	0,3	4,6	1,13	24 245	21 456
70 % de la médiane	22,0	0,5	2,0	1,04	24 245	23 313
Taux de pauvreté avant prestations sociales autres que retraites (en pourcentages)						
Total	26,0	0,4	1,6	1,09	6 586	6 042
Homes total	25,0	0,5	1,8	1,10	3 501	3 183
Femmes total	27,0	0,5	1,7	1,09	3 085	2 830
0-15 ans	34,2	1,0	2,8	1,02	3 316	3 251
16-64 ans	24,9	0,4	1,7	1,08	11 479	10 629
65+ ans	20,9	0,7	3,4	1,29	2 771	2 148
16+ ans	24,1	0,4	1,6	1,13	4 698	4 158
Hommes, 16-64 ans	24,0	0,5	2,0	1,09	1 868	1 714
Hommes, 65+ ans	18,8	0,9	4,7	1,12	272	243
Hommes, 16+ ans	23,1	0,4	1,9	1,10	2 140	1 945
Femmes, 16-64 ans	25,8	0,5	1,9	1,06	7 977	7 525
Femmes, 65+ ans	22,5	0,8	3,6	1,33	1 935	1 455
Femmes, 16+ ans	25,1	0,4	1,7	1,12	2 558	2 284
Coefficient de Gini (en pourcentages)	27,8	0,3	1,1	1,16	24 245	20 901
Niveau de vie moyen (en euros par unité de consommation)	18 199	104	0,6	1,20	24 245	20 204

Composantes du revenu disponible	Moyenne par ménage (en euros)	Écart-type	Coefficient de variation	Effet de sondage	Effectif atteint	Taille effective de l'échantillon
Revenu disponible total (HY020)	28 625	162	0,6	1,19	9 754	8 197
Revenus tirés de la location d'une propriété ou de terres (HY040N)	441	34	7,7	1,07	9 754	8 197
Allocation famille/enfant (HY050N)	770	13	1,7	1,04	9 754	8 197
Exclusion sociale non classée ailleurs (HY060N)	253	13	5,1	1,17	9 754	8 197
Aides au logement (HY070N)	437	9	2,1	1,08	9 754	8 197
Transferts interménages perçus régulièrement en espèces (HY080N)	180	11	6,1	1,07	9 754	8 197
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	737	18	2,4	1,17	9 754	8 197
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	1	22,6	1,05	9 754	8 197
Impôts réguliers sur la fortune (HY120N)	-35	12	-34,1	1,09	9 754	8 197
Transferts interménages versés régulièrement en espèces (HY130N)	-254	14	-5,5	1,16	9 754	8 197
Remboursements consécutifs à des ajustements fiscaux (HY145N)	-1 817	57	-3,1	1,28	9 754	8 197
Salaire ou revenu assimilé (PY010N)	17 123	148	0,9	1,26	9 754	8 197
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	1 927	105	5,4	1,14	9 754	8 197
Allocations chômage(PY090N)	906	31	3,4	1,06	9 754	8 197
Retraites (PY100N)	7 154	90	1,3	1,02	9 754	8 197
Allocations de survie (PY110N)	184	32	17,4	1,05	9 754	8 197
Indemnités de maladie (PY120N)	235	20	8,5	1,05	9 754	8 197
Pensions d'invalidité (PY130N)	347	23	6,6	1,10	9 754	8 197

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 très rares 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 soient hors champ et que nous soyons obligé 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 2004, dite « base de sondage de logements neufs » (BSLN). Pour la constituer, on part de l'ensemble des logements correspondant aux permis de construire délivrés par les mairies. Ces permis sont gérés dans un fichier administratif appelé SITADEL, qui comprend environ 300 000 logements chaque année. Un extrait de SITADEL est obtenu par sondage, ce qui donne lieu à un échantillon de logements initialement « fictifs » dont la construction est suivie sur le terrain par des enquêteurs jusqu'à ce qu'il y ait achèvement du logement. À la constatation de l'achèvement, le logement entre dans la BSLN. La qualité de SITADEL et du processus de suivi sur le terrain n'est pas

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

2.3.2 Erreurs de mesure et de traitements

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

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

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

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

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

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

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

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

Structure du questionnaire

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

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

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

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

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

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

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

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

L'architecture du questionnaire a été aménagée suite aux bilans de l'enquête de 2004. Il est finalement structuré de la façon 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 ;
 - Revenus non individualisables sur la période de référence ;
 - Endettement et confort financier.
- Questionnaire individuel (renseigné par chaque adulte du ménage, le recours à un proxy étant toléré) :
 - Biographie et ressources culturelles ;
 - Activité, emploi, profession ;
 - Revenus individualisables sur la période de référence, ;
 - Santé ;
 - Module secondaire.

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

Exhaustivité et absence de doubles comptes dans la collecte des revenus

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

- recenser les différents types de revenus perçus par le ménage au cours de l'année de référence (année civile 2004), 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 des documents utilisés. Parmi ces documents, la ou les déclarations des revenus de l'année de référence 2004 (remplies en mars ou avril 2005) sont privilégiées pour les revenus d'activité et de remplacement des adultes, qui sont en quasi-totalité imposables. La déclaration de revenus présente l'avantage d'un mode de recueil des revenus annuels à la fois plus facile et plus fiable. D'une part, les récapitulatifs annuels sont déjà faits. D'autre part, le concept mesuré (revenu déclaré) permet une étude de qualité par comparaison de la distribution des revenus avec celle de l'enquête Revenus fiscaux (enquête de référence à l'Insee pour la distribution des revenus).

Dans le questionnaire individuel, l'enquêté indique combien de déclarations de revenus 2004 il a effectuées, 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 chiffrage automatique développé par l'Insee. Les enquêteurs sont avertis lorsque le libellé renseigné est inconnu de la base des professions. Ils peuvent alors saisir d'autres libellés qui sont tous enregistrés. Ce ci permet de corriger les erreurs de saisie et de préciser les libellés.

Questionnement longitudinal

Le questionnement longitudinal a commencé en 2005. Les personnes réinterrogées ont répondu à un questionnaire un peu différent de celui des entrants si elles avaient accepté en 2004 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 2005 a bénéficié du bilan de la collecte de 2004. Ce bilan a été effectué grâce aux questionnaires papiers remplis par l'ensemble des enquêteurs, des synthèses rédigées par les directions régionales et des comptes-rendus des réunions de bilans organisées en Aquitaine, Île-de-France et Lorraine. Il a notamment conduit à abandonner les questions sur les cotisations sociales et à modifier l'architecture du questionnaire.

Le questionnaire de 2005 a donné lieu à deux tests sur le terrain.

Le premier test en dimension longitudinale s'est déroulé au printemps 2004 dans les trois directions régionales d'Alsace, de Bourgogne, et d'Île-de-France. Il a été mené auprès des 374 ménages qui avaient déjà été interrogés lors du test du questionnaire de 2004 en juin 2003. Il a permis de tester le suivi des individus dans le TCM longitudinal, la remontée de données antérieures en réinterrogation et les contrôles de cohérence d'une année à l'autre.

Le second test s'est déroulé en novembre 2004 dans les directions régionales du Centre, de Franche-Comté, de Languedoc-Roussillon et de Provence-Alpes-Côte-D'azur. Il a été mené auprès de 168 ménages qui avaient répondu au test de novembre 2003. Il avait pour objectif de tester le questionnaire longitudinal modifié suite au premier test panel ainsi que le questionnaire transversal amélioré suite à la collecte de 2004 et le module secondaire.

Enfin, un test en bureau a été mené en février 2005. 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 2005), 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.

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 envoyé 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 2005. 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 2005. 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 2005. L'échantillon pour lequel cette variable vaut 9 correspond à l'échantillon entrant en 2005.

Nombre de ménages pour lesquels un entretien est accepté pour la base de données	
Ventilation par groupe de rotation	
Nombre d'années dans le panel	Nombre de ménages
1	1 037
2	970
3	1 011
4	1 009
5	969
6	1 014
7	969
8	955
9	1 820
Total	9 754

Nombre de personnes âgées de 16 ans et plus, membres des ménages dont l'entretien est accepté pour la base de données, pour lesquels un entretien personnel est achevé. Ventilation par groupe de rotation	
Nombre d'années dans le panel	Nombre de personnes
1	1 967
2	1 890
3	1 965
4	1 948
5	1 878
6	1 928
7	1 876
8	1 860
9	3 457
Total	18 769

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 16,2 %. Il est de 30,3 % pour les entrants et de 13,3 % pour les réinterrogés.

Adresses contactées avec succès DB120=11	11 601
Total DB120	12 178
Adresse inexistante DB120=23	410
Adresses validées sélectionnées	11 768
Taux de contact à l'adresse Ra	98,6 %
Entretiens de ménages achevés et acceptés pour la base de données DB135=1	9 754
Ménages éligibles aux adresses contactées	
Total DB130	11 601
Proportion d'entretiens de ménages achevés et acceptés pour la base de données Rh	84,1 %
Taux de non-réponse des ménages NRh	16,2 %

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	18 769
Nombre d'individus éligibles dans les ménages dont les entretiens ont été achevés et acceptés pour la base de données RB245 = 1	18 948
Rp	99,1 %
Taux de non-réponse individuelle NRp	0,9 %
Taux global de non-réponse individuelle *NRp	17,0 %

On obtient ainsi un taux global de non-réponse individuelle de 17,0 (31,1% pour les entrants et 14,1 % pour les réinterrogés), très proche du taux de non-réponse des ménages à 16,2 %.

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

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

DB075	Enregistrement du contact à l'adresse			Total :
	11 : Adresse contactée	21 : Impossible de localiser l'adresse	23 : Adresse inexistante	
1	1 171 100,0 %	0 0,0 %	0 0,0 %	1 171
2	1 115 99,7 %	3 0,3 %	0 0,0 %	1 118
3	1 136 99,7 %	4 0,3 %	0 0,0 %	1 140
4	1 145 99,7 %	3 0,3 %	0 0,0 %	1 148
5	1 093 99,9 %	1 0,1 %	0 0,0 %	1 094
6	1 141 99,7 %	3 0,3 %	0 0,0 %	1 144
7	1 111 99,7 %	3 0,3 %	0 0,0 %	1 114
8	1 096 99,9 %	1 0,1 %	0 0,0 %	1 097
9	2 593 85,8 %	19 0,6 %	410 13,6 %	3 022
Total	11 601 96,3 %	37 0,3 %	410 3,4 %	12 048

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

DB130 Résultat du questionnaire "ménage"						
Nombre d'années dans le panel	11 :	21 :	22 :	23 :	24 :	Total
	Questionnaire rempli	Refus de coopérer	Ensemble du ménage absent	Ménage incapable de répondre	Autres raisons	
1	1 041 88,9 %	57 4,9 %	26 2,2 %	21 1,8 %	26 2,2 %	1 171
2	974 87,3 %	78 7,0 %	34 3,0 %	15 1,4 %	14 1,3 %	1 115
3	1 012 89,1 %	56 4,9 %	32 2,8 %	17 1,5 %	19 1,7 %	1 136
4	1 009 88,1 %	62 5,4 %	18 1,6 %	38 3,3 %	18 1,6 %	1 145
5	969 88,7 %	68 6,2 %	24 2,2 %	20 1,8 %	12 1,1 %	1 093
6	1 017 89,1 %	55 4,8 %	26 2,3 %	22 1,9 %	21 1,8 %	1 141
7	972 87,5 %	65 5,9 %	30 2,7 %	24 2,2 %	20 1,8 %	1 111
8	956 87,2 %	74 6,7 %	24 2,2 %	26 2,4 %	16 1,5 %	1 096
9	1 825 70,4 %	397 15,3 %	208 8,0 %	117 4,5 %	46 1,8 %	2593
Total	9 775 84,3 %	912 7,9 %	422 3,6 %	300 2,6 %	192 1,7 %	11 601

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

DB135	Acceptation de l'entretien		Total
	1 :	2 :	
Nombre d'années dans le panel	Entretien accepté pour la base de données	Entretien rejeté pour la base de données	
1	1037 99,6 %	4 0,4 %	1041
2	970 99,6 %	4 0,4 %	974
3	1011 99,9 %	1 0,1 %	1012
4	1009 100,0 %	0 0,0 %	1009
5	969 100,0 %	0 0,0 %	969
6	1014 99,7 %	3 0,3 %	1017
7	969 99,7 %	3 0,3 %	972
8	955 99,9	1 0,1 %	956
9	1820 99,7 %	5 0,3 %	1825
Total	9754 99,8 %	21 0,2 %	9775

2.3.3.4 Ventilation des unités remplacées

Sans objet dans le cas de la France.

2.3.3.5 Non-réponse à certaines questions

Les tableaux suivants fournissent par composante du revenu net disponible la part d'unités percevant cette composante, le pourcentage de valeur manquante et le taux d'information partielle. Ce dernier s'élève à 36,4 % pour le revenu disponible total. Les prestations sociales hors retraites sont les variables qui sont les plus sous-déclarées ou mal renseignées par les ménages.

Ventilation des non-réponses à certaines questions

Ménages	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Revenu disponible total (HY020)	100	1,7	36,4
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et de survie (HY022)	98,9	2,4	8,3
Revenu disponible des ménages avant transferts sociaux (HY023)	95,7	2,7	6,8
Revenus tirés de la location d'une propriété ou de terres (HY040N)	8,2	0,0	0,0
Allocation famille/enfant (HY050N)	29,7	6,8	19,9
Exclusion sociale non classée ailleurs (HY060N)	7,6	1,0	0,6
Aides au logement (HY070N)	21,0	3,1	2,8
Transferts interménages perçus régulièrement en espèces (HY080N)	5,9		0,0
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	76,5		
Revenus perçus par les personnes de moins de 16 ans (HY110N)	2,1		
Impôts réguliers sur la fortune (HY120N)	1,0		
Transferts interménages versés régulièrement en espèces (HY130N)	9,0		
Remboursements consécutifs à des ajustements fiscaux (HY145N)	82,8	0,0	0,7

	% de personnes de plus de 16 ans...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Salaire ou revenu assimilé (PY010N)	53,4	3,8	0,8
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,5		0,3
Allocations chômage (PY090N)	7,7		0,8
Retraites (PY100N)	23,7	0,0	0,6
Allocations de survie (PY110N)	1,1		0,0
Indemnités de maladie (PY120N)	3,0		0,0
Pensions d'invalidité (PY130N)	2,7		0,1
Allocations d'éducation (PY140N)	2,1		

2.3.3.6 Nombre total des non-réponses

La non-réponse totale 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 2 014 ménages, soit 16,2 % 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	24 152	0	93 ¹	0	0,0 %
Taux de pauvreté après transferts selon le statut professionnel	18 510	0	5 297 ²	179 ³	1,0 %
Taux de pauvreté après transferts selon le type de ménage	24 094	0	0	121 ⁴	0,5 %
Taux de pauvreté après transferts selon le statut d'occupation du logement	24 245	0	0	0	0,0 %
Taux de pauvreté après transferts selon l'intensité d'emploi	20 851	0	2 933 ⁵	461 ⁶	2,2 %
Médiane du revenu disponible par unité de consommation - après transferts sociaux	24 245	0	0	0	0,0 %
Seuil de pauvreté - après transferts sociaux - personnes seules	24 245	0	0	0	0,0 %
Seuil de pauvreté - après transferts sociaux - 2 adultes, 2 enfants	24 245	0	0	0	0,0 %
Inégalité de la distribution de revenu - Rapport du partage de revenus S80/S20	24 245	0	0	0	0,0 %
Écart moyen relatif du taux de pauvreté	3 271	0	20 974 ⁷	0	0,0 %
Dispersion autour du seuil de pauvreté	24 245	0	0	0	0,0 %
Taux de pauvreté avant transferts sociaux autres que retraites	24 152	0	93 ¹	0	0,0 %
Taux de pauvreté avant tout transfert social	24 152	0	93 ¹	0	0,0 %
Coefficient de Gini	24 245	0	0	0	0,0 %
Revenu disponible par uc moyen	24 245	0	0	0	0,0 %

¹ 93 enfants nés après le 31 décembre 2004.

² dont 5297 enfants de moins de 16 ans et 259 individus qui n'ont pas gardé le même statut professionnel plus de la moitié de l'année.

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

⁴ Nous n'avons pu déterminer le type de ménage pour 151 individus à cause de la non-réponse individuelle de 41 personnes âgées de 16 à 24 ans.

⁵ dont 99 personnes dans un ménage d'étudiants, 2 834 dans un ménage avec seulement des enfants dépendants ou des personnes âgées de plus de 65 ans.

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

⁷ 20 974 non-pauvres.

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

2.4 Mode de collecte des données

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,1 %. En effet, les ménages répondants ont tous rempli au moins un questionnaire individuel et les proxys sont autorisés.

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

RB250		Statut des données			
Nombre d'années dans le panel		11 Information collectée par enquête	23 Refus	32 pas de contact pour une autre raison	Total
1	Nombre Pourcentage	1 967 98,84	9 0,45	14 0,70	1 990
2	Nombre Pourcentage	1 890 99,11	8 0,42	9 0,47	1 907
3	Nombre Pourcentage	1 965 99,09	12 0,61	6 0,30	1 983
4	Nombre Pourcentage	1 948 98,68	21 1,06	5 0,25	1 974
5	Nombre Pourcentage	1 878 99,37	8 0,42	4 0,21	1 890
6	Nombre Pourcentage	1 928 99,48	4 0,21	6 0,31	1 938
7	Nombre Pourcentage	1 876 99,00	7 0,37	12 0,63	1 895
8	Nombre Pourcentage	1 860 99,09	11 0,59	6 0,32	1 877
9	Nombre Pourcentage	3457 98,94	19 0,54	18 0,52	3 494
Total	Nombre Pourcentage	18 769 99,1	99 0,5	80 0,4	18 948

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 26,9 %.

RB260		Type d'entretien	
Nombre d'années dans le panel		2	5
		Face à face CAPI	Proxy
1	Nombre Pourcentage	1 444 73,4 %	523 26,6 %
2	Nombre Pourcentage	1 398 74,0 %	492 26,0 %
3	Nombre Pourcentage	1 451 73,8 %	514 26,2 %
4	Nombre Pourcentage	1 407 72,2 %	541 27,8 %
5	Nombre Pourcentage	1 356 72,2 %	522 27,8 %
6	Nombre Pourcentage	1 412 73,2 %	516 26,8 %
7	Nombre Pourcentage	1 370 73,0 %	506 27,0 %
8	Nombre Pourcentage	1 306 70,2 %	554 29,8 %
9	Nombre Pourcentage	2 527 74,4 %	885 25,6 %
Total	Nombre Pourcentage	13 716 73,1 %	5 053 26,9 %

2.5 Durée de l'entretien

La durée moyenne de l'entretien pour la collecte 2005 est de 54 minutes (57 minutes pour les entrants et 53 pour les réinterrogés).

2.6 Les imputations

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

Avant toute chose, il convient de présenter certaines difficultés rencontrées. Les revenus dans l'EU-SILC 2005 sont relatifs à plusieurs dates. L'impôt payé a pour assiette les revenus imposables perçus au cours de l'année 2003. Les revenus collectés sont relatifs à l'année 2004. Enfin certains revenus concernent la période courante de l'enquête, soit mai 2005, ou bien une moyenne estimée sur une période relativement récente. Nous devons faire l'hypothèse d'une certaine continuité, pour pouvoir comparer ces différents revenus et les contrôler.

Certains revenus tirés de la collecte 2005 ont été modifiés. En effet, des comparaisons faites avec des fichiers externes tels l'enquête Revenus fiscaux (ERF), l'enquête Logement, ou les statistiques de la Caisse nationale des allocations familiales (Cnaf) font apparaître des écarts. De nombreuses raisons expliquent ces écarts. Nous avons décelé des erreurs francs/euros, des revenus déclarés comme étant annuels alors qu'ils sont mensuels, trimestriels, voire quadrimestriels. Pour les seules retraites,

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

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

Nous décrivons maintenant les opérations concernant les revenus principaux.

Les deux méthodes retenues pour les imputations

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

Salaires ou revenu assimilé (PY010N)

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

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

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

Les salaires des non-salariés

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

Préretraites

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

Allocations de vieillesse (PY100N) ou de survie (PY100N)

Les retraites sont traitées différemment des autres revenus. L'hypothèse majeure justifiant ce choix est que les retraites sont quasi stables entre l'année 2004 et la suivante. Comme nous ne disposons pas d'une retraite courante, nous utilisons le revenu courant comme élément de contrôle et nous le comparons à la somme des revenus courants 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 l'année 2004 et la précédente : toujours sous la même hypothèse nous comparons l'impôt sur le revenu 2004 déclaré et l'impôt sur les revenus 2003 que nous calculons. Deux contrôles sont donc possibles avant de prendre une décision.

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

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

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

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

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

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

Prestations familiales

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

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

Aides au logement

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

Minima sociaux

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

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

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

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

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

Loyer imputé

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

La méthode retenue comprend quatre étapes, dont deux régressions :

- 1) Estimation d'une équation de loyer à partir des données de l'enquête logement 2002. L'estimation porte sur les logements locatifs du parc privé hors loi de 1948². Les variables explicatives sont les caractéristiques du logement (surface, confort, sanitaires, équipement, état du logement ...) et de localisation (tranche de taille d'agglomération, zone climatique, typologie socio-économique de Nicole Tabard ...). Deux équations distinctes sont estimées, l'une pour les appartements (variable expliquée : le logarithme du loyer au m²) et l'autre pour les maisons (variable expliquée : le logarithme du loyer, la surface figurant parmi les variables explicatives).
- 2) Les équations précédemment estimées sont utilisées pour imputer un loyer fictif aux propriétaires occupants ainsi qu'aux ménages logés gratuitement et un loyer de marché aux locataires du parc social ou en Loi de 1948 de l'enquête logement. On a rajouté à la valeur issue de l'équation un résidu tiré selon une procédure de hot-deck stratifié.
- 3) Ce loyer imputé est régressé sur deux types de variables : des variables du tronc commun des enquêtes ménages de l'Insee d'une part, et des variables géographiques d'autre part. À caractéristiques sociodémographiques et de localisation identiques, les logements occupés par les accédants à la propriété sont d'une qualité moyenne supérieure à ceux des propriétaires sans charge de remboursement, qui sont eux-mêmes de meilleure qualité que ceux du parc social. Estimer une seule équation aurait pu biaiser les estimations. Aussi huit régressions distinctes ont-elles été estimées sur des segments relativement homogènes du parc :
 - appartements, propriétaires sans charge de remboursement et ménages logés gratuitement ;
 - appartements, accédants à la propriété ;
 - appartements, locataires du parc social ou loi de 1948 ;
 - appartements, locataires du parc libre louant vide ;
 - maisons, propriétaires sans charge de remboursement et ménages logés gratuitement ;
 - maisons, accédants à la propriété ;
 - maisons, locataires du parc social ou Loi de 1948 ;
 - maisons, locataires du parc libre louant vide.

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

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

Locataires mensuels réels ou fictifs en appartement

	Propriétaires		Accédants		Locataires parc social		Locataires parc libre	
	Enquête Logement	EU-SILC	Enquête Logement	EU-SILC	Enquête Logement	EU-SILC	Enquête Logement	EU-SILC
Effectif	2 137 397	2 240 295	1 004 068	1 085 626	3 719 367	3 956 889	3 399 559	2 985 510
Moyenne	557	580	566	565	421	415	430	434
Minimum	24	42	47	114	14	53	24	27
Q5	192	173	247	222	176	165	171	167
Q10	245	223	296	275	216	204	210	208
Q25	341	341	390	380	290	284	283	284
Q50	483	498	507	502	389	389	387	395
Q75	674	708	681	683	512	511	517	521
Q90	917	1 004	902	950	658	652	687	709
Q95	1 189	1 236	1 079	1 101	777	767	815	866
Maximum	3 911	4 269	2 565	2 350	1 987	1 837	3 811	3 190

Locataires mensuels réels ou fictifs en maison individuelle

	Propriétaires		Accédants		Locataires parc social		Locataires parc libre	
	Logement	EU-SILC	Logement	EU-SILC	Logement	EU-SILC	Logement	EU-SILC
Effectif	7 468 125	8 319 884	4 140 726	4 545 356	757 224	961 505	1 430 050	1 548 979
Moyenne	377	385	558	554	404	372	439	413
Minimum	16	19	54	72	43	39	45	29
Q5	98	107	197	222	134	121	149	180
Q10	133	140	261	280	178	158	200	215
Q25	211	220	384	384	278	248	289	279
Q50	328	333	523	518	388	343	402	381
Q75	489	499	691	673	507	463	561	503
Q90	679	686	885	842	644	605	731	640
Q95	816	838	1 049	1 007	733	723	845	754
Maximum	2 277	1 746	2 014	2 186	1 360	1 235	1 847	1 972

- 6) L'enquête Logement ayant été réalisée en 2002, les loyers obtenus doivent être actualisés pour 2004. Nous utilisons des coefficients d'actualisation issus de l'enquête trimestrielle Loyers et charges de l'Insee. Ces coefficients distinguent le secteur privé du secteur social.

3 COMPARABILITÉ

3.1 Concepts et définitions de base

Population de référence

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

Définition du ménage privé

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

1,5 % 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 2005. L'année de référence est 2004. On collecte au niveau individuel et ménage les revenus perçus au cours de l'année 2004. On collecte au niveau du ménage les impôts payés en 2004 au titre des revenus perçus au cours de l'année 2003. Les cotisations sociales sont relatives aux revenus perçus en 2004.

Période de référence pour les impôts sur la fortune

On considère le montant de l'impôt de solidarité sur la fortune payé en 2004. Il est relatif au patrimoine détenu au 1^{er} janvier 2004.

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 2005, ce qui rajoute une troisième période de référence après l'année 2004 (revenus déclarés) et l'année 2003 (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 30 juin 2005.

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

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

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

La profession des parents, collectée dans le cadre du module secondaire, a été codée de la même façon dans la nomenclature française. L'information collectée était cependant moins précise.

Le codage en ISCO se fait à partir du code des professions et catégories socioprofessionnelles (PCS) à quatre chiffres, de la nomenclature d'activité française en 712 postes et du nombre de salariés de l'établissement. Dans 95 % des cas, la matrice de passage élaborée pour l'enquête Emploi permet de coder la profession de l'enquêté en un code ISCO à deux positions. Sinon, la codification en ISCO a dû se faire par imputation statistique. Le code ISCO est alors affecté aléatoirement en fonction de la répartition du code ISCO par PCS à deux chiffres dans l'enquête Emploi du deuxième trimestre de 2005.

Dans 1,9 % 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.

En 2005, nous devions dans le cadre du module secondaire coder la profession des parents avec le même niveau de détail que celle des enquêtés. Or la réponse des enquêtés sur la profession de leurs parents est relativement imprécise parce qu'ils la connaissent mal ou ne s'en souviennent plus très bien. Le codage dans la nomenclature française est donc plus imprécis : il se fait plus souvent sur deux positions et est sans doute moins fiable. De ce fait, le passage au code ISCO à deux positions a dû être réalisé par imputation dans 15 % des cas. Ces imputations ont pris pour répartition de référence le code ISCO à 2 positions par catégorie socioprofessionnelle à 2 chiffres obtenue dans l'enquête Formation et qualification professionnelle de 2003.

3.2 Composantes du revenu

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

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

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

Ces trois composantes du revenu ne sont obligatoires qu'à partir de 2007, c'est pourquoi ils ne sont pas inclus dans le **revenu du ménage** brut disponible après ou avant transferts. Par contre, les avantages en nature liés à la voiture de fonction sont inclus dans les **rémunérations en espèces** (PY010N).

Revenus de la propriété (HY040N et HY090N)

Pas de classification particulière.

Allocations famille/enfants (HY050)

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

- l'allocation d'adoption (ADA) versée pendant 21 mois après l'adoption sous condition de ressources ;
- l'allocation parentale d'éducation (APE) versée aux parents qui arrêtent ou ont arrêté totalement ou partiellement de travailler pour élever leurs enfants. L'APE de rang 2 (2 enfants à charge) est versée à condition d'avoir travaillé deux ans (consécutifs ou non) dans les 5 ans précédant l'arrivée du dernier enfant. L'APE de rang 3 (3 enfants ou plus à charge) est versée à condition d'avoir travaillé deux ans au cours des dix ans précédant l'arrivée du dernier-né ;
- les allocations familiales qui sont versées à toutes les familles d'au moins deux enfants à charge ;
- le complément familial versé aux familles de trois enfants ou plus dont le benjamin est âgé d'au moins trois ans. Cette allocation est versée sous condition de ressources ;

- l'allocation pour jeune enfant (APJE) versée aux familles ayant un jeune enfant (de moins de trois ans), elle est aussi versée sous condition de ressources ;
- l'allocation parent isolé (API) est un minimum social versé aux personnes seules élevant seules un enfant. Cette allocation est versée durant douze mois ou bien jusqu'au mois précédant le troisième anniversaire du dernier-né ;
- l'allocation de soutien familial (ASF) est versée quand un parent ne concourt pas aux besoins de l'enfant ;
- l'allocation de rentrée scolaire (ARS) est versée en début d'année scolaire sous condition de ressources ;
- l'allocation d'éducation spéciale (AES) pour enfants handicapés ;
- l'allocation de présence parentale (APP) pour enfants handicapés ou malades en cas d'arrêt total ou partiel de travail des parents.

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

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

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

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

Aides au logement (HY070)

Pas de classification particulière.

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

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

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

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

Impôts réguliers sur la fortune (HY120)

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

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

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

L'agrégat HY145N comprend l'impôt sur le revenu des personnes physiques (IRPP), la taxe d'habitation et les impôts payés à l'étranger. La prime pour l'emploi est comptée négativement. L'IRPP assis sur les revenus de l'année N est payé l'année N+1 sous forme d'acomptes et d'un solde. Il n'est donc pas prélevé à la source et porte sur plusieurs types de revenus tous perçus l'année précédente,

c'est pourquoi nous l'avons intégralement enregistré comme ajustement d'impôt. La taxe d'habitation est payée par les personnes occupant un logement au premier janvier, elle est assise sur la valeur locative du logement mais son montant dépend aussi des revenus du ménage. La prime pour l'emploi est versée aux actifs aux revenus faibles, elle est déduite de l'IRPP.

Salaire ou revenu assimilé (PY010N)

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

Cotisations sociales des employeurs (PY030G)

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

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

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

Valeur des biens produits pour la consommation propre (PY070N)

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

Allocations de chômage (PY090N)

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

Allocations de vieillesse (PY100N)

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

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

Allocations de survie (PY110N)

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

Indemnités de maladie (PY120N)

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

Pensions d'invalidité (PY130N)

Les allocations suivantes sont comptées dans cette composante du revenu :

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

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

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

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

Allocations d'études (PY140N)

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

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

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

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

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

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

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

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

Salaire ou revenu assimilé (PY010 et PY030G)

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

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

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

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

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

Retraites de régimes privés, allocations chômage (PY090), de vieillesse (PY100) et de survie (PY110)

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

Loyers imputés (HY030)

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

Prestations familiales (HY050) et allocations logement (HY070)

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

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

La valeur brute est égale à la valeur nette.

4 COHÉRENCE

4.1 Comparaison des variables cibles avec des sources extérieures

Les données externes ayant servi à contrôler les éléments constitutifs du revenu sont diverses. Nous avons utilisé l'enquête Revenus fiscaux et l'enquête Logement. L'enquête Revenus fiscaux 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 sont imputées sur barèmes.

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 2003 publié par la Direction de la recherche, des études, de l'évaluation et des statistiques.

4.1.1 Salaire et revenu assimilé (PY010N)

Année de revenus	EU-SILC		ERF	
	2003	2004	2003	2004
Salaires (milliards d'€)	432	438	464	475
Bénéficiaires (millions)	24,5	25,2	25,9	26,2
Moyenne		17 361		18 164
Minimum		12		1
D10		3 800		3 162
D20		8 000		7 460
D30		11 595		11 467
D40		13 542		13 939
D50		15 477		15 954
D60		17 574		18 195
D70		20 068		21 020
D80		23 837		24 903
D90		30 267		31 904
Maximum		225 791		1 247 812

Les écarts entre les deux enquêtes sont de l'ordre de 8 % en masses et de 5 % en nombre de bénéficiaires. Ces variations trouvent en partie leur origine par des différences de champ entre les deux enquêtes. Le concept fiscal est plus large que celui de l'enquête SILC. C'est ainsi que les indemnités maladie sont incluses dans le salaire fiscal et isolées dans celui du panel (6,1 milliards d'euros). Les bourses d'étudiants sont traitées de la même façon (820 millions d'euros). C'est aussi le cas des indemnités de départ en retraite ou en préretraite. C'est pourquoi les retraites sont sensiblement supérieures dans le panel. L'effet mémoire n'est pas non plus à écarter, il peut arriver dans le panel qu'un salarié oublie de mentionner certaines primes.

4.1.2 Les allocations chômage (PY090N)

	EU-SILC		ERF	
	2003	2004	2003	2004
Revenus				
Chômage (milliards d'€)	23,4	23,0	24,8	27,8
Bénéficiaires (millions)	3,78	3,85	4,26	4,50
Moyenne		6 076		6 180
Minimum		5		1
D10		992		703
D20		1 800		1 465
D30		2 816		2 389
D40		3 760		3 400
D50		4 810		4 585
D60		5 800		5 720
D70		7 317		7 443
D80		9 180		9 505
D90		11 776		12 898
Maximum		60 840		108 034

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

4.1.3 Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et de survie (PY110N)

	EU-SILC		ERF	
	2003	2004	2003	2004
Année de revenus				
Retraites (milliards d'€)	175	189	171	180
Bénéficiaires (millions)	12,3	12,9	13,5	13,9
Moyenne		14 666		12 951
Minimum		50		7
D10		3 840		2 400
D20		7 006		4 410
D30		8 660		7 099
D40		10 556		9 119
D50		12 409		11 193
D60		14 400		13 369
D70		16 716		15 678
D80		20 098		18 908
D90		25 698		24 420
Maximum		280 000		317 478

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

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

Année de revenus	EU-SILC		ERF	
	2003	2004	2003	2004
Revenus d'indépendants (milliards d'€)	4,33	5,11	5,15	5,40
Bénéficiaires (millions)	1,93	2,17	2,04	2,04
Moyenne		23 473		26 541
Minimum		2		1
D10		2 000		1 134
D20		6 000		3 395
D30		8 628		6 137
D40		11 000		9 640
D50		14 400		13 264
D60		18 000		19 253
D70		25 548		26 816
D80		34 858		38 363
D90		56 000		60 195
Maximum		360 000		702 191

Les revenus des indépendants sont abordés de façon différente dans l'ERF et le panel, aussi les comparaisons ne sont pas possibles. Le panel privilégie le revenu privé, à savoir les sommes que l'indépendant prélève sur les ressources de l'entreprise à des fins privées, pour consommer ou épargner. L'ERF enregistre un résultat fiscal ayant pour base un forfait, un bénéfice réel ou un régime micro-entreprise, résultant d'abattements, d'amortissements et de déductions.

4.1.5 Revenus sociaux (HY050G, HY060G et HY070G).

Année de revenus s	EU-SILC		ERF	
	2003	2004	2003	2004
Sociaux (milliards d'€)	36,0	38,3	39,2	39,9
Bénéficiaires (millions)	9,81	10,56	9,86	9,93

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

4.1.5.1 Prestations familiales (HY050N).

	EU-SILC	Cnaf
Année de revenus	2004	2004
Prestations familiales (milliards d'€)	19,8	22,7
Bénéficiaires (millions)	6,6	

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 2004 qui ne peut être comparé avec le nombre de personnes déclarant dans SILC avoir perçu une prestation au cours de l'année 2004.

4.1.5.2 Aides au logement (HY070N)

	EU-SILC	Cnaf
Année de revenus	2004	2004
Allocations logements (milliards d'€)	11,5	13,5
Bénéficiaires (millions)	5,52	

Les aides au logement sont de 10 % inférieures dans SILC, ceci s'explique essentiellement par l'écart de champ. Des études de la Cnaf montrent en effet qu'environ 10 % des bénéficiaires de ces allocations vivent en ménage non ordinaire.

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

Année de revenus	EU-SILC		ERF	
	2003	2004	2003	2004
Impôts (milliards d'€)	45,4	46,4	51,3	52,5

Les deux enquêtes diffèrent de 13 % sur la masse de l'impôt sur le revenu. La période de référence n'est pas la même : l'ERF retient l'impôt de l'année des revenus tandis que le panel collecte l'impôt de l'année précédant les revenus. Mais ceci n'explique qu'une partie de la différence observée.

4.1.7 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 2002.

	Enquête Logement 2002	Enquête Logement 2002 actualisée ¹	EU-SILC 2005
Effectif	9 504 209	9 504 209	9 37 9041
Moyenne	356	391	378
Q0	14	15	1
Q5	152	167	150
Q10	183	201	194
Q25	236	259	250
Q50	317	348	340
Q75	427	469	460
Q90	559	614	600
Q95	683	750	716
Q100	3 811	4 189	5 000

¹Les loyers sont actualisés à partir de l'évolution moyenne des loyers estimée à partir de l'enquête trimestrielle Loyers et charges.

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

4.1.8 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. Six postes sont distingués lors de la collecte, les livrets d'épargne exonérés, les livrets bancaires soumis à l'impôt sur le revenu, l'épargne logement, les revenus de valeurs mobilières, les produits d'assurance vie, les autres placements financiers. Un taux de rendement issu de l'enquête Patrimoine de 2002 est appliqué aux stocks collectés afin d'estimer le revenu.

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

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

	EU-SILC	ERF
Année des revenus	2004	2004
Revenus du patrimoine financier (milliards d'€)	1,86	1,38
Bénéficiaires (millions)	19,45	8,08
Moyenne	957	1 708
Minimum	2	-132
D10	25	7
D20	49	24
D30	92	57
D40	208	114
D50	273	208
D60	489	374
D70	729	668
D80	1 184	1 227
D90	2 304	2 747
Maximum	26 120	690 280

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

La comparaison se fait sur le revenu disponible après transferts sociaux tel qu'il est estimé dans l'ERF.

	EU-SILC	ERF
Année des revenus	2004	2004
Revenus disponible des ménages (milliards d'€)	734,05	720,00
Bénéficiaires (millions)	25,64	25,28
Moyenne	28 624	28 475
Minimum	-209 101	-361 826
D10	10 621	10 900
D20	13 987	14 068
D30	17 234	17 160
D40	20 734	20 614
D50	24 514	24 287
D60	28 635	28 309

D70	33 334	32 912
D80	39 591	39 094
D90	50 171	49 293
Maximum	286 475	755 511

	EU-SILC	ERF
Revenus	2004	2004
Niveau de vie des individus (milliards d'€)	464,63	448,84
Bénéficiaires (millions)	59,47	59,33
Moyenne	18 013	17871
Minimum	-209 101	-201 014
D10	8 748	8 830
D20	10 648	10 956
D30	12 394	12 556
D40	14 011	14 133
D50	15 783	15 679
D60	17 773	17 580
D70	20 093	19 791
D80	23 093	22 798
D90	29 097	28 253
Maximum	244 860	503 674

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

Annexe méthodologique

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

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

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

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

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

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

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

⁵ Dans certains cas seulement la différence peut être négligeable.

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

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