

**Rapport final sur la qualité des
données françaises EU-SILC 2006**

TABLE DES MATIÈRES

Rapport final sur la qualité des données françaises EU-SILC 2006	1
Table des matiÈres.....	3
1 Indicateurs COMMUNAUTAIRES LONGITUDINAUX communs FONDES SUR LA DIMENSION LONGITUDINALE DES STATISTIQUES EU-SILC.....	5
1.1 Taux de pauvreté persistant après transferts sociaux selon le sexe (60% de la médiane)	5
1.2 Taux de pauvreté persistant après transferts sociaux selon le sexe (50% de la médiane)	5
2 PrÉcision	6
2.1 Plan d'échantillonnage.....	6
2.1.1 Type de plan d'échantillonnage	6
2.1.2 Unités d'échantillonnage.....	6
2.1.3 Critères de stratification et de sous-stratification.....	6
2.1.4 Taille de l'échantillon et critères d'attribution.....	6
2.1.5 Systèmes de sélection de l'échantillon.....	6
2.1.6 Répartition de l'échantillon au cours du temps et renouvellement de l'échantillon : groupes de rotation	7
2.1.8 Pondération.....	7
2.1.9 Substitutions	11
2.2 Erreurs d'échantillonnage	12
2.3 Erreurs non dues à l'échantillonnage.....	14
2.3.1 Erreurs dans la base de sondage et la couverture.....	14
2.3.2 Erreurs de mesure et de traitements	15
2.3.3 Erreurs de non-réponse.....	20
2.4 Mode de collecte des données	31
2.5 Les imputations	33
2.6 Loyer imputé	35
3 ComparabilitÉ.....	37
3.1 Concepts et définitions de base.....	37
3.2 Composantes du revenu.....	38
3.2.1 Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC	38
3.2.2 La source ou la procédure utilisée pour la collecte des variables du revenu.....	40
3.2.3 La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes.....	40
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).....	41
3.3 Règles de suivi.....	41
4 CohÉrence	42
4.1 Comparaison des variables cibles avec des sources extérieures	42
4.1.1 Salaire et revenu assimilé (PY010N).....	42

4.1.2	Les allocations chômage (PY090N)	43
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension de réversion (PY110N).....	43
	Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante	43
4.1.4	Revenus sociaux (HY050G, HY060G et HY070G).	43
4.1.5	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation).	43
4.1.6	Loyers	44
4.1.7	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N).....	44
4.1.8	Revenu disponible du ménage et niveau de vie des individus.....	44
	Annexe méthodologique.....	45

1 INDICATEURS COMMUNAUTAIRES LONGITUDINAUX COMMUNS FONDES SUR LA DIMENSION LONGITUDINALE DES STATISTIQUES EU-SILC

1.1 Taux de pauvreté persistant après transferts sociaux selon le sexe (60% de la médiane)

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.

Le taux de pauvreté persistante ne pourra être calculé qu'à partir de la collecte 2007.

On donne cependant la part des personnes qui sont en dessous du seuil de pauvreté en 2004 et en 2006.

Longitudinal	2004			2006			2004 & 2006		
	Total	F	H	Total	F	H	Total	F	H
Pauvres	13,9	14,7	13,0	13,9	14,8	13,0	7,2	7,7	6,6
Non pauvres	86,1	85,3	87,0	86,1	85,2	87,0	92,8	92,3	93,4

Effectifs non pondérés

Longitudinal	2004			2006			2004 & 2006		
	Total	F	H	Total	F	H	Total	F	H
Total	19 301	9 903	9 398	20 951	10 744	10 207	15 425	7 954	7 471
Pauvres	2 588	1 407	1 181	2 961	1 589	1 372	1 042	582	460
Non pauvres	16 713	8 496	8 217	17 990	9 155	8 835	14 383	7 372	7 011

1.2 Taux de pauvreté persistant après transferts sociaux selon le sexe (50% de la médiane)

Part des personnes qui sont en dessous du seuil de pauvreté en 2004 et en 2006.

Longitudinal	2004			2006			2004 & 2006		
	Total	F	H	Total	F	H	Total	F	H
Pauvres	7,5	7,9	7,0	8,2	8,6	7,7	3,2	3,4	3,1
Non pauvres	92,5	92,1	93,0	91,8	91,4	92,3	96,8	96,3	96,9

Effectifs non pondérés

Longitudinal	2004			2006			2004 & 2006		
	Total	F	H	Total	F	H	Total	F	H
Total	19 301	9 903	9 398	20 951	10 744	10 207	15 425	7 954	7 471
Pauvres	1 372	748	624	1 745	930	815	451	247	204
Non pauvres	17 929	9 155	8 774	19 206	9 814	9 392	14 974	7 707	7 267

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 présents 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 deux 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 tirées. 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 les sous-échantillons entrants en 2005 et 2006 en comprenait chacun 3 000. Les sous-échantillons entrants des années à venir devraient être également calibrés à 3 000 logements.

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 684 unités secondaires de l'échantillon de EU-SILC sur les 3 217 unités de l'échantillon-maître a également été proportionnel à la taille en nombre de résidences principales. Le tirage des logements est effectué au sein du groupe de communes par un algorithme systématique sur fichier trié.

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

Cas des strates urbaines jusqu'à 100 000 habitants

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

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

Cas du « grand urbain »

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

2.1.6 Répartition de l'échantillon au cours du temps 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 2006, 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 2005 au moins une personne répondante et n'appartenant pas à l'échantillon sortant. Il comprenait 9 207 logements.

2.1.8 Pondération

En 2006, nous avons estimé pour la deuxième fois des poids longitudinaux. Nous avons suivi la méthodologie présentée par Pascal Ardilly et Pierre Lavallée (Symposium 2003, Stat Canada).

2.1.8.1 Effet du plan d'échantillonnage (entrants)

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

2.1.8.2 Ajustements pour non-réponse

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

La probabilité de non-réponse initiale est notée PI.

La correction de la non-réponse initiale

Les modèles de non-réponse initiale 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 en 2004 cinq types de modèles de non-réponse totale selon les informations disponibles dans les bases de sondage ou recueillies par l'enquêteur en 2004 :

- les ménages d'une personne habitant des résidences principales au dernier recensement de la population (RP) et en 2004,
- pour les ménages de 2 personnes des résidences principales au RP et en 2004,
- les ménages de 3 personnes et plus des résidences principales au RP et en 2004,
- 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 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 au 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 de sous-populations. La probabilité de réponse initiale d'un ménage dans une sous-population donnée vaut donc :

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

Pour les échantillons entrants suivants, compte tenu de la plus faible taille d'échantillonnage, 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 l'année de collecte :

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

Les entrants de 2005 ont été répartis en 12 groupes et ceux de 2006 en 8 groupes.

Pour l'ensemble des ménages, PI est compris entre 0,5 et 0,9. Le poids corrigé de la non-réponse initiale (WI/PI) varie dans un rapport de 1 à 6.

La correction de la non-réponse en réinterrogation

Le modèle de non-réponse est estimé au niveau individuel. Une probabilité de réponse est attribuée à chaque individu panel qui appartient à l'échantillon en 2005. Est considéré comme répondant tout individu qui appartient à un ménage répondant (DB135='1').

Les nouveau-nés de mère panel se voient attribuer la probabilité de réponse de leur mère.

Soit $PR(2005)$ la probabilité de répondre en 2005 sachant qu'on a répondu en 1^{ère} vague,

Pour les individus hors champ en 2005, $PR(2005)=1$.

Pour estimer la probabilité de réponse en réinterrogation en 2005, nous avons estimé $PR(2005)$ pour les 21 900 individus panel en 2004 encore dans le champ en 2005.

Probabilité de répondre en 2005 sachant qu'on a répondu en 2004

Un modèle de non-réponse a été estimé sur les 17 040 adultes panel en 2004, réinterrogés et encore dans le champ en 2005. L'information utilisée est celle de 2004.

Les facteurs explicatifs de la non-réponse sont :

- le fait d'avoir quitté son ménage depuis la dernière vague,
- le fait d'avoir déménagé avec l'ensemble de son ménage depuis la dernière vague,
- la nationalité (magrébine/autre),
- le statut d'activité (actif occupé ou inactif),
- la CS (artisans/commerçants (CS=2) ou inactifs autres que retraité (CS=8) ou autres),
- l'âge (20-24 ans/plus de 65 ans/autre),
- la situation familiale (seul/non),
- la perception de minima sociaux au sein du ménage,
- le diplôme (diplômé/non),
- la localisation géographique (agglomération parisienne/autre),
- le quartile de niveau de vie (1er, 2d ou au-dessus de la médiane) du ménage.

À partir de ces résultats, nous avons formé différentes sous-populations de répondants en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif ont été regroupés pour assurer la robustesse des résultats). Au final, nous avons réparti les 17 040 individus réinterrogés en 2005 dans 21 groupes et nous avons supposé le mécanisme de réponse homogène à l'intérieur de sous-populations. La probabilité $PR(2005)$ vaut donc :

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

Pour les 4 860 enfants (moins de 16 ans en 2004), pour lesquels nous n'avons que peu d'information, nous appliquons un taux de réponse homogène.

La probabilité $PR(2005)$ est comprise entre 0,5 et 1.

En 2006, deux modèles ont été estimés selon les mêmes facteurs explicatifs de la non-réponse :

- un premier modèle portant sur la probabilité de répondre en 2006 sachant qu'on a répondu en 2005.
- un second modèle portant sur la probabilité de répondre en 2006 sachant qu'on n'a pas répondu en 2005 mais répondu en 2004.

Les individus sont ensuite répartis en sous-groupes dans lesquels nous supposons le mécanisme de réponse homogène.

Dans ces deux derniers modèles, les individus hors champ en 2006 sont considérés comme répondants en 2006 donc $PR(2006)=1$ et pour les enfants, nous appliquons un taux de réponse uniforme (probabilité de réponse moyenne) sur les deux sous-groupes constitués (enfants répondants en 2005 et encore dans le champ en 2006 d'une part et enfants de 2004 non répondants en 2005 et encore dans le champ en 2006 d'autre part).

Enfin, on affecte aux nouveaux-nés panel 2005 et 2006 le poids de leur mère.

La probabilité $PR(2006)$ est comprise entre 0,3 et 1.

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

Nous attribuons à chaque individu panel appartenant à l'échantillon de réinterrogation un poids corrigé de la non réponse. On obtient les pondérations suivantes :

Pour les échantillons 13 à 19, soit en 3^{ème} interrogation =>

$WC(2004_2006) = WI / (PI * PR(2006))$. Ce poids varie dans un rapport de 1 à 10.

$WC(2005_2006) = WI / ((7 * (anais < 2004) + 1) * PR(2005) * PR(2006) * \alpha(2006, 2005, 3))$ avec $\alpha(2006, 2005, 3) = 0.9491$.

Pour l'échantillon 20, soit en 2^{ème} interrogation (2005 et 2006) =>

$WC(2005_2006) = WI / ((7 * (anais < 2004) + 1) * PR(2005) * PR(2006) * 1)$.

Ce poids varie dans un rapport de 1 à 20.

Nous ne disposons pas de marges pour la population de référence de l'échantillon longitudinal, qui est l'ensemble des personnes ayant vécu en 2004, 2005 et 2006 en ménage ordinaire en France métropolitaine. Nous avons effectué un calage sur les marges de 2004 en prenant en compte l'ensemble des personnes sorties du champ entre les enquêtes. Nous avons ensuite mis le poids des personnes hors champ à 0. On effectue plusieurs calages correspondants aux différentes pondérations longitudinales calculées : soit $WC04_06$, $WC05_06$.

L'enquête est, comme toutes les enquêtes ménages de l'Insee réalisées en 2004, calée sur les marges issues de l'enquête Emploi de 2003. 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.

Le calage s'est effectué avec le logiciel Calmar. Nous avons utilisé la fonction de distance « logit ».

Les variables et les modalités introduites dans le calage sont les suivantes :

- nombre d'individus 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 d'individus par type de ménage :
 - couple sans enfant,
 - couple avec 1 enfant
 - couple avec 2 enfants ou plus
 - famille monoparentale
 - personne seule, autre configuration ;
- nombre d'hommes par tranches d'âge (6 modalités, des moins de 15 ans aux 76 ans et plus). L'âge est mesuré au 31/12/2003 ;
- nombre de femmes par tranches d'âge (6 modalités) ;
- nombre d'individus selon le diplôme :

- moins de 16 ans, sans diplôme, non déclaré, diplôme inférieur baccalauréat (CAP, BEPC),
- Baccalauréat, bac+2,
- diplôme supérieur ;
- nombre d'individus 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 et moins de 16 ans,
- nombre de ménages par région :
 - Ile de France et bassin parisien,
 - Nord, Est et Ouest,
 - Sud-Ouest, Centre-Est et Méditerranée ;
 - nombre d'individus selon la nationalité :
 - o française,
 - o étrangère.

Dans le cas des pondérations longitudinales 2004-2006, les rapports de poids avant/après calage évoluent entre 0,5 et 2. Les pondérations finales des individus varient quant à elles dans un rapport de 1 à 21. 32 individus ont un poids plus de 10 fois supérieur au poids le plus faible.

Pour les pondérations longitudinales 2005-2006, les rapports de poids avant/après calage évoluent entre 0,4 et 1,7. Les pondérations finales des individus varient quant à elles dans un rapport de 1 à 37. 371 individus ont un poids plus de 10 fois supérieur au poids le plus faible.

2.1.8.4 Pondération longitudinale finale

Pour les répondants, le poids est modifié pour corriger de la non-réponse individuelle en 2004, 2005 ou 2006. 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. Elles ne respectent 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 1,1 %, l'impact de cette correction est très limité, de 0,0 % à 4,2 %.

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.

2004	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Revenu disponible total (HY020)	27 154	7 923	7 923	183
Revenu disponible total avant prestations sociales autres que retraites (HY022)	24 329	7 923	7 923	187
Revenu disponible total avant prestations sociales (HY023)	17 340	7 923	7 923	176
Composantes du revenu net au niveau ménage				
Revenus tirés de la location d'une propriété ou de terres (HY040N)	315	502	502	79
Allocation famille/enfant (HY050N)	764	838	2 121	33
Exclusion sociale non classée ailleurs (HY060N)	160	284	401	35
Aides au logement (HY070N)	454	1477	1 673	15
Transferts interménages perçus régulièrement en espèces (HY080N)	295	496	509	81
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	689	5990	5 990	23
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	154	154	5
Intérêts (HY100N)	272	1512	1 512	24
Impôts réguliers sur la fortune (HY120N)	-38	61	61	57
Transferts interménages versés régulièrement en espèces (HY130N)	-326	769	769	48
Remboursements consécutifs à des ajustements fiscaux (HY145N)	-1 771	6418	6 418	58
Composantes du revenu net au niveau individu				
Salaire ou revenu assimilé (PY010N)	9 357	7 886	7 886	199
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	957	615	615	275
Autoconsommation (PY070N)	48	1 627	1 627	12
Allocations chômage(PY090N)	500	1 062	1 151	89
Retraites (PY100N)	3 542	3 406	3 515	154
Pension de réversion (PY110N)	125	168	175	129
Indemnités de maladie (PY120N)	103	447	447	51
Pensions d'invalidité (PY130N)	160	373	380	66
Allocation d'éducation (PY140N)	25	300	300	14

2006	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Revenu disponible total (HY020)	28 917	8 191	8 191	191
Revenu disponible total avant prestations sociales autres que retraites (HY022)	25 882	8 191	8 191	194
Revenu disponible total avant prestations sociales (HY023)	18 443	8 191	8 191	186
Revenus tirés de la location d'une propriété ou de terres (HY040N)	421	711	711	97
Allocation famille/enfant (HY050N)	788	1 986	2 336	29
Exclusion sociale non classée ailleurs (HY060N)	159	296	401	35
Aides au logement (HY070N)	432	1 438	1 791	14
Transferts interménages perçus régulièrement en espèces (HY080N)	189	495	495	49
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	769	6 468	6 468	22
Revenus perçus par les personnes de moins de 16 ans (HY110N)	3	169	169	2
Intérêts (HY100N)	403	1 776	1 776	26
Impôts réguliers sur la fortune (HY120N)	-35	91	91	93
Transferts interménages versés régulièrement en espèces (HY130N)	-300	786	786	48
Remboursements consécutifs à des ajustements fiscaux (HY145N)	-2 076	6 651	7 192	59
Composantes du revenu net au niveau individu				
Salaire ou revenu assimilé (PY010N)	9 753	8 435	8 435	204
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	1 077	704	704	362
Autoconsommation (PY070N)	35	1 808	1 808	5
Allocations chômage(PY090N)	556	1 229	1 229	113
Retraites (PY100N)	3 833	3 728	3 728	144
Pension de réversion (PY110N)	110	160	160	70
Indemnités de maladie (PY120N)	121	652	652	57
Pensions d'invalidité (PY130N)	225	467	467	83
Allocation d'éducation (PY140N)	28	346	346	15

Revenu disponible 2004	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Total	17 139	19 225	19 226	126
Ménage d'une personne	16 290	2 347	2 348	415
Ménage de 2 personnes	19 048	5 029	5 029	453
Ménage de 3 personnes	17 674	3 594	3 594	588
Ménage de 4 personnes et +	15 749	8 255	8 255	467
0-24 ans	15 401	6 434	6 435	134
25-34 ans	17 285	2 328	2 328	411
35-44 ans	17 373	2 832	2 832	339
45-54 ans	19 014	2 762	2 762	413
55-64 ans	19 519	2 024	2 024	583
65+ ans	16 858	2 845	2 845	405
Hommes	17 327	9 360	9 361	108
Femmes	16 959	9 865	9 865	100

Revenu disponible 2006	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Total	18 438	20 655	20 656	128
Ménage d'une personne	16 674	2 341	2 342	429
Ménage de 2 personnes	19 885	5 473	5 473	474
Ménage de 3 personnes	19 609	3 860	3 860	878
Ménage de 4 personnes et +	17 357	8 981	8 981	327
0-24 ans	16 975	6 984	6 985	141
25-34 ans	18 496	2 309	2 309	479
35-44 ans	18 535	3 004	3 004	324
45-54 ans	20 722	2 910	2 910	458
55-64 ans	21 388	2 444	2 444	638
65+ ans	17 030	3 004	3 004	319
Hommes	18 740	10 047	10 047	119
Femmes	18 154	10 608	10 609	104

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 bonne qualité, même s'il existe quelques omissions de logements (la qualité du recensement est certes moins bonne pour les personnes physiques - mais la base utilisée ici est une base de logements). L'ancienneté du recensement explique qu'un certain nombre de logements de la base soit hors champ et que nous soyons obligés de prendre dans l'échantillon des résidences vacantes ou secondaires au moment du recensement alors que notre champ se limite aux résidences principales.

- une base qui a mis à jour la précédente fin 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), certains ménages continuent à « répondre en francs », même si l'enquêteur demande une réponse en euros. Un certain nombre de personnes âgées raisonnent toujours en anciens francs¹. Une calculatrice intégrée permet la conversion par l'enquêteur de montants exceptionnellement fournis en francs et les cartes codes 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 zéro 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

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

Structure du questionnaire

L'enquête est composée d'un questionnaire « Ménage » (destiné à l'ensemble du ménage) et d'un questionnaire « Individu » posé à toutes les personnes du ménage âgées de 16 ans ou plus (au 1er 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...),
- de certaines prestations (prestations maladie, invalidité ou dépendance, bourses d'études, etc..)
- de minima sociaux (allocation de parent isolé, revenu minimum d'insertion et minimum vieillesse) qui dépendent néanmoins des ressources du foyer : la logique est ici plutôt celle d'un « foyer » allocataire bien qu'elle soit versée à un individu allocataire.

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

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

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

L'architecture du questionnaire est la suivante :

- Tronc commun des enquêtes ménages (TCM) :
 - Identification du logement, contact ;
 - Tableau des habitants du logement et contour des unités de vie (THL).
- Questionnaire ménage (renseigné par un adulte quelconque du ménage « unité de vie ») :
 - Ressources et charges en période courante ;
 - Changements récents et jeunes enfants ;
 - Conditions de logement ;
 - 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, avant de collecter les montants correspondants,
- vérifier que le montant n'a pas été inclu 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 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 il a effectué pour l'année de référence, et s'il consent à s'y reporter. Pour les revenus d'activité et de remplacement, le questionnement prend deux formes différentes selon la réponse :

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

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

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

Codification de la profession

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

Questionnement longitudinal

Le questionnement longitudinal a commencé en 2005. Les personnes réinterrogées ont répondu à un questionnaire un peu différent de celui des entrants si elles avaient accepté 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 en général entre -20 % et + 30 %),
- sur l'évolution de montants agrégés de revenus (impôts, salaires, retraites).

Tests du questionnaire 2004

Une première version du questionnaire (questionnaire sous CAPI) est testée au printemps 2003 par cinq directions régionales de l'Insee (Picardie, Alsace, Bourgogne, Nord, Ile-de-France) et porte sur environ 400 ménages. La formation dure une journée. La formation est précédée d'une demi-journée de formation au Tronc Commun des enquêtes Ménages.

Cette formation se compose à la fois d'une présentation générale de l'enquête (contexte européen, structure du questionnaire, présentation de ses différents modules) et d'exercices sous CAPI.

Trois directions régionales de l'Insee (Franche Comté, Centre, Provence-Alpes-Côte-D'azur, Languedoc-Roussillon) testent une nouvelle version du questionnaire à l'automne 2003 auprès d'environ 400 ménages. La formation dure de nouveau une journée (encore une fois précédée d'une demi-journée de formation au Tronc Commun des enquêtes Ménages).

Des tests en bureau ont également été menés au 1^{er} trimestre 2004. Ce type de tests permet de se placer dans des configurations particulières qui n'ont pas été rencontrées lors des tests sur le terrain.

Tests du questionnaire 2005

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 ré-interrogation 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.

Tests du questionnaire 2006

Le questionnaire transversal de 2006 a bénéficié du bilan de la collecte de 2005. 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 Bretagne, en Franche-Comté et en Provence-Alpes-Côte-D'azur. Il a notamment conduit à demander une confirmation du revenu global, qu'il soit donné en clair ou en tranche.

Le questionnaire de 2006 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2005 dans les directions régionales de Franche-Comté et d'Île-de-France. Il a été mené auprès de 146 ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2005 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2006. Ce type de test permet de se placer dans des configurations particulières qui n'ont pas été rencontrées lors des tests sur le terrain.

À l'issue de chaque test et suite aux bilans de collecte et aux différentes remarques des enquêteurs, des modifications du questionnaire (structure, formulation) sont proposées au maître d'ouvrage par

l'équipe de conception. Ces modifications du questionnaire sont validées à l'occasion d'un comité de pilotage du projet.

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

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

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

Formation des enquêteurs

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

En revanche, pour la collecte en grandeur réelle (mai-juin 2004/2006), 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 envoyée aux enquêtés pour les prévenir de la venue d'un enquêteur,
- d'une plaquette de présentation du dispositif envoyée avec la lettre avis,
- d'une instruction aux enquêteurs,
- d'un cahier des cartes (les cartes sont utilisées dans le cas où l'énumération des modalités d'une question par l'enquêteur serait trop longue ou fastidieuse),
- d'un fascicule « revenus », compléments d'information pour les enquêteurs sur certaines aides et prestations existantes et sur les conditions à remplir pour en bénéficier, réalisé afin d'aider les enquêteurs à retrouver un type de revenus lorsque le ménage ne sait pas exactement ce qu'il perçoit,
- d'une architecture du questionnaire,
- de fiches de prise de contact et de suivi. Les fiches de suivi comportent quelques informations (numéros de téléphone fixe ou portable, adresse internet, personne-relais) pour anticiper les déménagements,
- d'un document présentant les résultats de l'enquête de l'année précédente, donné à tous les ménages répondant.

2.3.3 Erreurs de non-réponse

2.3.3.1 Taille de l'échantillon obtenu

Les deux tableaux ci-dessous fournissent le nombre de ménages, puis le nombre d'individus correspondants, pour lesquels l'entretien a été accepté suite aux collectes 2004, 2005 et 2006.

Nombre de ménages pour lesquels un entretien est accepté pour la base de données (DB135=1)			
	(DB010=2004)	(DB010=2005)	(DB010=2006)
Nombre de ménages	7 923	8 712	8 191

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. (DB135=1 & RB250=11)			
	(DB010=2004)	(DB010=2005)	(DB010=2006)
Nombre de personnes	14 950	16 634	15 719
Nombre d'individus panels (RB100=1)	14 950	16 369	15 287
Nombre de co-résidents (RB100=2)	0	265	432

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.

Vague 2004 (sous échantillons 3 à 9)

Adresses contactées avec succès DB120=11	7 923
Total DB120	9 493
Adresse inexistante DB120=23	1 483
Adresses validées sélectionnées	8 010
Taux de contact à l'adresse Ra	98,9%
Entretiens de ménages achevés et acceptés pour la base de données DB135=1	7 923
Ménages éligibles aux adresses contactées Total DB130	9 493
Proportion d'entretiens de ménages achevés et acceptés pour la base de données Rh	83,5%
Taux de non-réponse des ménages NRh	17,4%

Taux de réponse des ménages : comparaisons des codes résultat des vagues 2004 et 2005

2004		2005									Total
		DB130=11 Questionnaire rempli									
		DB135=1 Entretien accepté pour la base de données	DB135=2 Entretien rejeté pour la base de données	DB130=22 Ensemble du ménage absent	DB130=23 Ménage incapable de répondre	DB130=24 Autres raisons	DB130=21 Refus de coopérer	DB120=21 Impossible de localiser l'adresse	NC DB110=3-7 Ménage a quitté le panel	DB110=10 Fusion de ménages	
DB130=11 Questionnaire rempli	DB135=1 Entretien accepté	6 757	11	18	159	91	391	9	316	64	7 923
	DB135=2 Entretien rejeté	0	0	0	0	0	0	0	0	0	0
DB120=21 Impossible de localiser l'adresse											0
DB120=23 Adresse inexistante											0
DB130=21 Refus de coopérer											0
DB130=22 Ensemble du ménage absent											0
DB130=23 Ménage incapable de répondre											0
DB130=24 Autres raisons											0
Total		6 757	11	18	159	91	391	9	316	64	7 923

2005		2005								Total	
		DB130=11									
		DB135=1	DB135=2	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	NC DB110=3-7		DB110=10
DB110=8 Éclatement de ménage		140	1	21	4	13	17	4	0	0	209
DB110=9 Nouvelle adresse ou 1e vague		1 815	0	0	0	0	0	19	0	0	2 244

8 712	12	39	163	104	408	32	316	64	10 376
-------	----	----	-----	-----	-----	----	-----	----	--------

Nombre total de ménages	10 376	100%
Taux de réponse de la vague 2005 (DB135=1)	8 712	84,0%
Refus (DB130=21)	408	3,9%
Non contactés et autres	567	5,5%

Taux de suivi longitudinal	90,0%
Ratio de suivi (*100)	115
Ratio taille finale échantillon (*100)	110

Taux de réponse des ménages : comparaisons des codes résultat des vagues 2005 et 2006

2005		2006									Total
		DB130=11 Questionnaire rempli									
		DB135=1 Entretien accepté pour la base de données	DB135=2 Entretien rejeté pour la base de données	DB130=22 Ensemble du ménage absent	DB130=23 Ménage incapable de répondre	DB130=24 Autres raisons	DB130=21 Refus de coopérer	DB120=21 Impossible de localiser l'adresse	NC DB110=3-7 Ménage a quitté le panel	DB110=10 Fusion de ménages	
DB130=11 Questionnaire rempli	DB135=1 Entretien accepté	7 827	11	24	140	71	236	1	276	11	8 712
	DB135=2 Entretien rejeté	8	1	0	0	0	1	0	2	0	12
DB120=22 Adresse inaccessible	NH (db110=9 1e vague)										0
	No NH	0	0	0	0	0	0	0	0	0	0
DB130=22 Ensemble du ménage absent	NH (db110=9 1e vague)										0
	No NH	13	0	5	1	0	8	0	8	1	39
DB130=23 Ménage incapable de répondre	NH (db110=9 1e vague)										0
	No NH	54	1	0	36	3	36	0	26	1	163
DB130=24 Autres raisons	NH (db110=9 1e vague)										0
	No NH	39	0	0	7	8	25	0	14	2	104

2005		2006									Total
		DB130=11									
		DB135=1	DB135=2	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	NC DB110=3-7	DB110=10	
Total		8 024	13	30	196	84	336	1	406	16	10 376

2006											Total
DB110=8 Éclatement de ménage		167	0	36	4	26	24	9	0	0	273
DB110=9 Nouvelle adresse ou 1e vague		0	0	0	0	0	0	0	0	0	0

7 941	13	29	184	82	306	1	326	15	9 030
-------	----	----	-----	----	-----	---	-----	----	-------

Nombre total de ménages	9 030	100%
Taux de réponse de la vague 2006 (DB135=1)	7 941	87,9%
Refus (DB130=21)	306	3,4%
Non contactés et autres	466	5,2%

Taux de suivi longitudinal	92%
Ratio de suivi (*100)	94,6
Ratio taille finale échantillon (*100)	91,2

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

Vague 2004 (sous échantillons 3 à 9)

Nombre entretiens individuels achevés RB250 = 11	14 950
Nombre d'individus éligibles dans les ménages dont les entretiens ont été achevés et acceptés pour la base de données RB245 = 1	15 060
Rp	99,3 %
Taux de non-réponse individuelle NRp	0,7 %
Taux global de non-réponse individuelle *NRp	18,0 %

Taux de réponse individuelle

		2005									Total
		RB250=11 Entretien individuel achevé	Non complété						HH nc	Pn RB110=6 ou RB120=2,3	
			RB250=21-22 Entretien n'a pas pu être réalisé	RB250=23 Refus de coopérer	RB250=31 ou 33 Non contacté	RB250=32 Non contacté	RB250=33 Non contacté				
Individus panel issus de la vague précédente (RB100 = 1 et RB245=1)											
1	RB110=1-2 Membre courant du ménage	12 841	0	52	0	38	5				12 936
2	RB110=6 décédé										52
3	RB110=-1										0
4	RB120=2 Parti en institution										11
5	RB120=3 Parti à l'étranger										15
6	RB120=4 Perdu										64
7	DB135=2 ou -1, ou DB110=7, ou DB120=21-23 ou -1, ou DB130=21-24 ou -1										146
8	DB110=3-6										0
Nouveaux individus panel											
9	Ont atteint 16 ans	239	0	2	0	0	0	0	0	0	241
10	Individus panel supplémentaires	0	0	0	0	0	0	0	0	0	0
Individus non panel de plus de 16 ans											
11	Vague 2005	Présents en 2004	0	0	0	0	0	0	0	0	0
11		Non présents en 2004	265	0	17	0	10	0	0	0	0
12	Vague 2004	Présents en 2004									0
12		Non présents en 2004									0
Individus panel 2004 non présents en 2005											
13	Individus panel 2004 non présents en 2005										0
SOMMES											
1 + 3 + 6 + 7 + 9 + 10		13 080	0	54	0	38	5	0	0	0	13 387
1 + 3 + 6 + 7 + 9 + 10 + 13		13 080	0	54	0	38	5	0	0	0	13 387
1 + 3 + 6 + 7 + 9 + 10 + 11		13 080	0	54	0	38	5	0	0	0	13 387

2006										
RB250=11 Entretien individuel achevé	Non complété									Total
	RB250=21-22 Entretien n'a pas pu être réalisé	RB250=23 Refus de coopérer	RB250=31 ou 33 Non contacté	RB250 =32 Non contacté	RB250 =33 Non contacté	HH nc	Pn RB110=6 ou RB120=2,3	PI RB110=4 or -1		

Individus panel issus de la vague précédente (RB100 = 1 et RB245=1)

1	RB110=1-2 Membre courant du ménage	11 708	0	43	0	57	1				11 810
2	RB110=6 décédé										47
3	RB110=-1										0
4	RB120=2 Parti en institution										16
5	RB120=3 Parti à l'étranger										22
6	RB120=4 Perdu										54
7	DB135=2 ou -1, ou DB110=7, ou DB120=21-23 ou -1, ou DB130=21-24 ou -1										195
8	DB110=3-6										0

Nouveaux individus panel

9	Ont atteint 16 ans	239	0	2	0	1	0	0	0	0	242
10	Individus panel supplémentaires	0	0	0	0	0	0	0	0	0	0

Individus non panel de plus de 16 ans

11	Vague 2006	Présents en 2004	0	0	0	0	0	0	0	0	0
11		Non présents en 2004	436	0	22	1	14	0	0	0	0
12	Vague 2005	Présents en 2004									0
12		Non présents en 2004									

Individus panel 2005 non présents en 2006

13	Individus panel 2005 non présents en 2006										0
----	--	--	--	--	--	--	--	--	--	--	---

SOMMES

	1 + 3 + 6 + 7 + 9 + 10	11 947	0	45	0	58	1	0	0	0	12 301
	1 + 3 + 6 + 7 + 9 + 10 + 13	11 947	0	45	0	58	1	0	0	0	12 301
	1 + 3 + 6 + 7 + 9 + 10 + 11	12 383	0	67	1	72	1	0	0	0	12 774

Taux de réponse des individus panel de la vague 2006	97,1%
Taux de réponse des co-résidents (sélectionnés en 2004) de la vague 2006	-
Taux de suivi longitudinal (RB250=11)	97,1%
Refus (RB250=23)	0,4%
Non contactés et autres (RB250=32)	0,5%
Ratio (taille finale 2006 / taille finale 2005) individus panel	91,3%
Ratio (taille finale 2006 / taille finale 2005) individus panel et co-résidents	94,7%
Ratio (taille finale 2006 / taille finale 2005) co-résidents (sélectionnés en 2004)	-
Taux de réponse des personnes non-panel	92,2%

2.3.3.3 Ventilation des ménages selon le « statut du ménage » (DB110), l'« enregistrement du contact à l'adresse » (DB120), le « résultat du questionnaire ménage » (DB130), et l'« acceptation de l'entretien » (DB135)

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2004

2004	Total	DB110=1 Même adresse en 2004 & 2005	2 Tout le ménage a déménagé (reste dans le champ)	3 Tout le ménage est parti en institution	4 Tout le ménage est parti à l'étranger	5 Tout le ménage est décédé	6 Le ménage ne contient pas d'individus panels	7 Adresse non contactée	8 Éclatement de ménage	9 Nouvelle adresse	10 Fusion
Total	9 493	0	0	0	0	0	0	0	0	9 493	0
%	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	0,0

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2005

2005	Total	DB110=1 Même adresse en 2004 & 2005	2 Tout le ménage a déménagé (reste dans le champ)	3 Tout le ménage est parti en institution	4 Tout le ménage est parti à l'étranger	5 Tout le ménage est décédé	6 Le ménage ne contient pas d'individus panels	7 Adresse non contactée	8 Éclatement de ménage	9 Nouvelle adresse	10 Fusion
Total	10 376	6 958	585	25	35	32	0	224	209	2 244	64
%	100	67,1	5,6	0,2	0,3	0,3	0,0	2,2	2,0	21,6	0,6

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2006

2006	Total	DB110=1 Même adresse en 2004 & 2005	2 Tout le ménage a déménagé (reste dans le champ)	3 Tout le ménage est parti en institution	4 Tout le ménage est parti à l'étranger	5 Tout le ménage est décédé	6 Le ménage ne contient pas d'individus panels	7 Adresse non contactée	8 Éclatement de ménage	9 Nouvelle adresse	10 Fusion
Total	9 527	8 092	740	30	28	59	0	289	273	0	16
%	100	84,9	7,8	0,3	0,3	0,6	0,0	3,0	2,9	0,0	0,2

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2004

2004	Total	DB120=11 Adresse contactée	21 Adresse non localisée	22 Adresse inaccessible	23 Adresse inexistante	Manquant

Total	9 493	7 923	87	0	1 483	0
%	100	83,5	0,9	0,0	15,6	0,0

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2005

2005	Total	DB120=11 Adresse contactée	21 Adresse non localisée	22 Adresse inaccessible	23 Adresse inexistante	Manquant
Total	10 376	2 480	32	0	410	7 454
%	100	23,9	0,3	0,0	4,0	71,8

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2006

2006	Total	DB120=11 Adresse contactée	21 Adresse non localisée	22 Adresse inaccessible	23 Adresse inexistante	Manquant
Total	9 527	851	10	0	0	8 666
%	100	8,9	0,1	0,0	0,0	91,0

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2004

2004	Total	DB130=11 complété	21 refus de coopérer	22 tout le ménage est temporairement absent	23 tout le ménage est incapable de répondre	24 autre motif de non réponse	Manquant
Total	9 493	7 923	0	0	0	0	1 570
%	100	83,5	0,0	0,0	0,0	0,0	16,5

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2005

2005	Total	DB130=11 complété	21 refus de coopérer	22 tout le ménage est temporairement absent	23 tout le ménage est incapable de répondre	24 autre motif de non réponse	Manquant
Total	10 376	8 724	408	39	163	104	938
%	100	84,1	3,9	0,4	1,6	1,0	9,0

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2006

2006	Total	DB130=11 complété	21 refus de coopérer	22 tout le ménage est temporairement absent	23 tout le ménage est incapable de répondre	24 autre motif de non réponse	Manquant
Total	9 527	8 204	360	66	200	110	587
%	100	86,1	3,8	0,7	2,1	1,1	6,2

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2004

2004	Total	DB135=1 accepté	DB135=2 refusé	Manquant
Total	9 493	7 923	0	1 570
%	100	83,5	0,0	16,5

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2005

2005	Total	DB135=1 accepté	DB135=2 refusé	Manquant
Total	10 376	8 712	12	16 52
%	100	84,0	0,1	15,9

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2006

2006	Total	DB135=1 accepté	DB135=2 refusé	Manquant
Total	9 527	8 191	13	1 323
%	100	86,0	0,1	13,9

2.3.3.4 Distribution des personnes en fonction de leur statut (RB110)

Distribution des personnes en fonction de leur statut (RB110)

2005	Total	Membres habituels du ménage				Membres non habituels du ménage			Manquant
		RB110=1 Était présent à la vague précédente	RB110=2 Venu d'un autre ménage panel	RB110=3 Venu d'un ménage non panel	RB110=4 Naissance	RB110=5 A quitté le ménage	RB110=6 Décédé	RB110=7 Dans le ménage au moins 3 mois pendant période de référence	
Total	22 015	20 872	238	362	150	340	53	0	0
%	100	94,8	1,1	1,6	0,7	1,5	0,2	0,0	0,0

2006	Total	Membres habituels du ménage				Membres non habituels du ménage			Manquant
		RB110=1 Était présent à la vague précédente	RB110=2 Venu d'un autre ménage panel	RB110=3 Venu d'un ménage non panel	RB110=4 Naissance	RB110=5 A quitté le ménage	RB110=6 Décédé	RB110=7 Dans le ménage au moins 3 mois pendant période de référence	
Total	20 951	19 587	221	368	223	494	58	0	0
%	100	93,5	1,1	1,8	1,1	2,4	0,3	0,0	0,0

Distribution des personnes ayant déménagé par destination (RB120)

RB110=5						
2005	Total	RB120=1 : ménage ordinaire		RB120=2 Collectivité	RB120=3 Etranger	RB120=4 Perdu
		Membre courant d'un ménage	Membre non courant d'un ménage			
Total	340	161	74	12	18	75
%	100	47,3	21,8	3,5	5,3	22,1

RB110=5						
2006	Total	RB120=1 : ménage ordinaire		RB120=2 Collectivité	RB120=3 Etranger	RB120=4 Perdu
		Membre courant d'un ménage	Membre non courant d'un ménage			
Total	494	191	167	19	25	92
%	100	38,7	33,8	3,9	5,1	18,6

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 valeurs manquantes et le taux d'information partielle.

Ventilation des non-réponses à certaines questions en 2004

Ménages	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Revenu disponible total (HY020)	100,0	3,0	49,6
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et pension de réversion (HY022)	98,6	3,7	28,6
Revenu disponible des ménages avant transferts sociaux (HY023)	95,0	3,0	15,6
Revenus tirés de la location d'une propriété ou de terres (HY040N)	6,3	0,0	0,0
Allocation famille/enfant (HY050N)	26,8	16,2	3,5
Exclusion sociale non classée ailleurs (HY060N)	5,1	1,5	1,1
Aides au logement (HY070N)	21,1	2,5	6,3
Transferts interménages perçus régulièrement en espèces (HY080N)	6,4	0,2	0,1
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	75,6	-	-
Revenus perçus par les personnes de moins de 16 ans (HY110N)	1,9	-	-
Impôts réguliers sur la fortune (HY120N)	0,8	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	9,7	-	0,1
Remboursements consécutifs à des ajustements fiscaux (HY145N)	81,0	-	2,1

	% 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)	52,8	4,0	5,8
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,1	0,1	0,4
Allocations chômage (PY090N)	7,7	0,6	0,8
Retraites (PY100N)	23,5	0,7	8,4
Pensions de réversion (PY110N)	1,2	0,1	0,5
Indemnités de maladie (PY120N)	3,0	-	0,0
Pensions d'invalidité (PY130N)	2,5	0,1	0,1
Allocations d'éducation (PY140N)	2,0	-	-

Ventilation des non-réponses à certaines questions en 2005

Ménages	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Revenu disponible total (HY020)	100,0	1,8	42,2
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et pension de réversion (HY022)	98,4	2,5	15,5
Revenu disponible des ménages avant transferts sociaux (HY023)	97,3	2,7	10,1
Revenus tirés de la location d'une propriété ou de terres (HY040N)	8,3	-	0,0
Allocation famille/enfant (HY050N)	29,5	6,8	19,9
Exclusion sociale non classée ailleurs (HY060N)	7,8	1,0	0,6
Aides au logement (HY070N)	21,0	2,8	3,1
Transferts interménages perçus régulièrement en espèces (HY080N)	5,8	-	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)	0,9	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	9,1	-	-
Remboursements consécutifs à des ajustements fiscaux (HY145N)	82,9	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)	53,2	3,7	2,8
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,5	-	0,3
Allocations chômage(PY090N)	7,8	-	0,8
Retraites (PY100N)	23,8	0,0	2,9
Pensions de réversion (PY110N)	1,1	-	0,1
Indemnités de maladie (PY120N)	2,9	-	0,0
Pensions d'invalidité (PY130N)	2,8	-	0,1
Allocations d'éducation (PY140N)	2,1	-	-

Ventilation des non-réponses à certaines questions en 2006

Ménages	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Revenu disponible total (HY020)	100,0	1,8	51,3
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et pension de réversion (HY022)	99,0	2,3	36,6
Revenu disponible des ménages avant transferts sociaux (HY023)	97,0	4,7	23,0
Revenus tirés de la location d'une propriété ou de terres (HY040N)	8,7	-	-
Allocation famille/enfant (HY050N)	28,5	4,3	9,6
Exclusion sociale non classée ailleurs (HY060N)	4,9	1,3	0,6
Aides au logement (HY070N)	21,9	4,3	1,4
Transferts interménages perçus régulièrement en espèces (HY080N)	6,0	-	-
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	79,0	-	-
Revenus perçus par les personnes de moins de 16 ans (HY110N)	2,1	-	-
Impôts réguliers sur la fortune (HY120N)	1,1	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	9,6	-	-
Remboursements consécutifs à des ajustements fiscaux (HY145N)	87,8	6,6	15,1

	% de personnes de plus de 16 ans...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
Salaire ou revenu assimilé (PY010N)	53,4	3,2	0,9
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,5	-	0,4
Allocations chômage(PY090N)	7,8	-	0,6
Retraites (PY100N)	23,6	-	7,4
Pensions de réversion (PY110N)	1,0	-	0,1
Indemnités de maladie (PY120N)	4,1	-	0,0
Pensions d'invalidité (PY130N)	3,0	-	0,1
Allocations d'éducation (PY140N)	2,2	-	-

2.4 Mode de collecte des données

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,3% pour 2004, 99,1% pour 2005 et 98,8% pour 2006. En effet, les ménages répondants ont tous rempli au moins un questionnaire individuel et les proxys sont autorisés.

Ventilation des membres des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3) selon le statut des données (RB250), vague 2004

2004	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	15 060	14 950	48	57	5
%	100	99,3	0,3	0,4	0,0

Ventilation des membres non panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=2) selon le statut des données (RB250), vague 2004

Sans objet.

En 2004, 1ère vague de l'enquête en France, tous les individus sont panels.

Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3) selon le statut des données (RB250), vague 2005

2005	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	16 906	16 750	87	64	5
%	100	99,1	0,5	0,4	0,0

Ventilation des membres panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=1) selon le statut des données (RB250), vague 2005

2005	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	16 614	16 485	70	54	5
%	100	99,2	0,4	0,3	0,0

Ventilation des membres non panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=2) selon le statut des données (RB250), vague 2005

2005	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	292	265	17	10	0
%	100	90,8	5,8	3,4	0,0

Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3) selon le statut des données (RB250), vague 2006

2006	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	15 982	15 794	88	98	2
%	100	98,8	0,6	0,6	0,0

Ventilation des membres panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=1) selon le statut des données (RB250), vague 2006

2006	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	15 509	15 358	66	84	1
%	100	99,0	0,4	0,5	0,0

Ventilation des membres non panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=2) selon le statut des données (RB250), vague 2006

2006	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	473	436	22	14	1
%	100	92,2	4,6	3,0	0,2

Seuls les membres du ménage ont le droit de répondre à la place d'un individu. Le tableau suivant détaille les taux de recours au proxy, qui s'élèvent globalement à 24,8% en 2004, 26,9% en 2005 et 27,3% en 2006.

Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB250=11) selon le type d'entretien (RB260), vague 2004

2004	Total	RB260=2 CAPI	RB260=5 Proxi
Total	14 950	11 243	3 707
%	100	75,2	24,8

Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB250=11) selon le type d'entretien (RB260), vague 2005

2005	Total	RB260=2 CAPI	RB260=5 Proxi
Total	16 750	12 251	4 499
%	100	73,1	26,9

Ventilation des membres panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=1 et RB250=11) selon le type d'entretien (RB260), vague 2005

2005	Total	RB260=2 CAPI	RB260=5 Proxi
Total	16 485	12 131	4 354
%	100	73,6	26,4

Ventilation des membres non panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=2 et RB250=11) selon le type d'entretien (RB260), vague 2005

2005	Total	RB260=2 CAPI	RB260=5 Proxi
Total	265	120	145
%	100	45,3	54,7

Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB250=11) selon le type d'entretien (RB260), vague 2006

2006	Total	RB260=2 CAPI	RB260=5 Proxi
Total	15 794	11479	4 315
%	100	72,7	27,3

Ventilation des membres panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=1 et RB250=11) selon le type d'entretien (RB260), vague 2006

2006	Total	RB260=2 CAPI	RB260=5 Proxi
Total	15 358	11 269	4 089
%	100	73,4	26,6

Ventilation des membres non panels des ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3 et RB100=2 et RB250=11) selon le type d'entretien (RB260), vague 2006

2006	Total	RB260=2 CAPI	RB260=5 Proxi
Total	436	210	226
%	100	48,2	51,8

2.5 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 N sont relatifs à plusieurs dates. L'impôt payé a pour assiette les revenus imposables perçus au cours de l'année N-1. Les revenus collectés sont relatifs à l'année N-1. Enfin certains revenus concernent la période courante de l'enquête, soit mai N, 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 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 de 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 contentés de lui attribuer le montant moyen de préretraites d'individus ayant des caractéristiques similaires.

Allocations de vieillesse (PY100N) ou pension de réversion (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 N-1 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 N-1 et la précédente : toujours sous la même hypothèse nous comparons l'impôt sur le revenu N-1 déclaré et l'impôt sur les revenus N-2 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 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 N-1. 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.

2.6 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.
- 4) Les huit équations estimées sont exportées vers l'enquête SILC pour y imputer soit un loyer fictif soit un loyer manquant. Lors de l'imputation on rajoute à la valeur prédite un résidu tiré selon une procédure de hot-deck stratifié.
- 5) Les résultats avant prise en compte de l'augmentation des loyers sont présentés ici.

Loyers mensuels réels ou fictifs en appartement

Cf. rapport intermédiaire 2006.

Loyers mensuels réels ou fictifs en maison individuelle

Cf. rapport intermédiaire 2006.

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.

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.

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

Appartenance au ménage

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

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

L'enquête est conduite en mai-juin N. L'année de référence est N-1. On collecte au niveau individuel et ménage les revenus perçus au cours de l'année N-1. On collecte au niveau du ménage les impôts payés en N-1 au titre des revenus perçus au cours de l'année N-2. Les cotisations sociales sont relatives aux revenus perçus en N-1.

Période de référence pour les impôts sur la fortune

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

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 N, ce qui rajoute une troisième période de référence après l'année N-1 (revenus déclarés) et l'année N-2 (assiette des impôts sur le revenu).

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

La collecte SILC 2004 s'est déroulée du 10 mai au 25 juin 2004.

La collecte SILC 2005 s'est déroulée du 9 mai au 30 juin 2005.

La collecte SILC 2006 s'est déroulée du 17 mai au 1er juillet 2006.

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

On constitue dans CAPI un tableau d'activité afin de cerner de façon détaillée l'activité des enquêtés au cours de l'année N-1. On commence par collecter l'activité principale de l'individu au 1^{er} janvier N-1. Par la suite, on recense tous les changements d'activités qui ont pu avoir lieu depuis cette date. Cette partie du questionnaire est parfois longue (et laborieuse) à renseigner, en particulier dans le cas de personnes intérimaire ou occupant des emplois saisonniers.

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 catégorie socioprofessionnelle des actifs ou des anciens actifs est codée automatiquement et en cours de collecte par le système SICORE, intégré à CAPI. SICORE (Système Informatique de COdage des Réponses aux Enquêtes) est un système de chiffrement automatique développé par l'INSEE. 95% des professions sont codées automatiquement, les autres professions étant « reprises » à la main par une équipe spécialisée à l'INSEE.

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

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 (deux 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 (trois 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 journalière de présence parentale (AJPP) 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 remplacées par la prestation d'accueil du jeune enfant (PAJE) et ne peuvent donc plus être distinguées des autres prestations familiales. La mise en place de la PAJE se passe de la façon suivante : les ménages qui ont un nouveau-né après le 1^{er} janvier 2004 passent entièrement sous le régime de la PAJE, les autres continuent de bénéficier des anciennes allocations.

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

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

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

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 sont assises sur les salaires, les pensions, les revenus d'indépendants, les allocations chômage, les retraites, les prestations familiales et les allocations logement.

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

Salaires ou revenu assimilé (PY010N)

Les salaires sont nets des cotisations sociales salariées. 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, 60 ans en cas d'invalidité ou d'invalidité.

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

Pension de réversion (PY110N)

Ce sont les pensions reçues par les survivants quel que soit leur âge.

Indemnités de maladie (PY120N)

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

Pensions d'invalidité (PY130N)

Les allocations suivantes sont comptées dans cette composante du revenu :

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

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

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

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

Allocations d'études (PY140N)

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

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

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

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

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

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

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

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

Salaire ou revenu assimilé (PY010 et PY030G)

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

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

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

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

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

Retraites de régimes privés, allocations chômage (PY090), de vieillesse (PY100) et pension de réversion (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.

3.3 Règles de suivi

Les règles de suivi pour la France sont conformes au règlement européen.

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 du Ministère de la santé, de la jeunesse, des sports et de la vie associative.

4.1.1 Salaire et revenu assimilé (PY010N)

Année de revenus	EU-SILC Longitudinal			EU-SILC Transversal			ERF		
	2003	2004	2005	2003	2004	2005	2003	2004	2005
Salaires (milliards d'€)	-	-	-	432	438	449	464	475	489
Bénéficiaires (millions)	-	-	-	24,5	25,2	25,1	25,9	26,2	26,2
Moyenne	17 552	17 740	18 395	17 605	17 361	17 861	17 885	18 164	18 617
Minimum	10	12	10	10	12	10	1	1	1
D10	3 771	3 398	3 360	3 981	3 800	3 744	3 151	3 162	3 151
D20	8 125	7 545	8 134	8 776	8 000	8 350	7 694	7 460	7 694
D30	11 455	11 400	11 998	11 595	11 595	11 911	11 882	11 467	11 882
D40	13 507	13 513	14 202	13 551	13 542	13 996	14 430	13 939	14 430
D50	15 600	15 568	16 264	15 477	15 477	15 888	16 492	15 954	16 492
D60	17 681	17 865	18 484	17 427	17 574	18 011	18 665	18 195	18 665
D70	20 430	20 500	21 207	19 943	20 068	20 467	21 497	21 020	21 497
D80	24 384	24 455	25 239	23 698	23 837	24 156	25 377	24 903	25 377
D90	31 272	31 560	32 418	30 286	30 267	30 974	32 640	31 904	32 640
Maximum	454 224	225 791	338 928	1 253 904	225 791	338 928	847 620	1 247 812	847 620

4.1.2 Les allocations chômage (PY090N)

Année de revenus	EU-SILC Longitudinal			EU-SILC Transversal			ERF		
	2003	2004	2005	2003	2004	2005	2003	2004	2005
Chômage (milliards d'€)	-	-	-	23,4	23,0	26,0	24,8	27,8	27,4
Bénéficiaires (millions)	-	-	-	3,78	3,85	3,82	4,26	4,50	4,75
Moyenne	6 266	6 060	7 046	6 185	6 076	6 827	5 816	6 180	5764
Minimum	2	5	27	2	5	27	1	1	1
D10	869	897	828	890	992	833	693	703	680
D20	1 533	1 735	1 508	1 600	1 800	1 522	1 489	1 465	1 344
D30	2 600	2 836	2 500	2 700	2 816	2 480	2 344	2 389	2 157
D40	3 708	3 770	3 565	3 768	3 760	3 648	3 285	3 400	3 141
D50	4 673	4 810	4 738	4 680	4 810	4 800	4 362	4 585	4 226
D60	5 619	5 800	5 913	5 680	5 800	5 889	5 315	5 720	5 255
D70	7 274	7 203	7 651	7 259	7 317	7 599	7 035	7 443	6 902
D80	9 361	9 180	9 762	9 144	9 180	9 600	8 964	9 505	9 128
D90	12 000	11 978	14 000	11 775	11 776	13 488	12 270	12 898	12 227
Maximum	132 480	60 840	153 643	132 480	60 840	153 643	62 203	108 034	61 737

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

Cf. rapport intermédiaire 2006.

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

Cf. rapport intermédiaire 2006.

4.1.4 Revenus sociaux (HY050G, HY060G et HY070G).

Cf. rapport intermédiaire 2006.

4.1.4.1 Prestations familiales (HY050N).

Cf. rapport intermédiaire 2006.

4.1.4.2 Aides au logement (HY070N)

Cf. rapport intermédiaire 2006.

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

Cf. rapport intermédiaire 2006.

4.1.6 Loyers

Cf. rapport intermédiaire 2006.

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

Cf. rapport intermédiaire 2006.

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

Cf. rapport intermédiaire 2006.

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 entre 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.