

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

TABLE DES MATIÈRES

Rapport final sur la qualité des données françaises EU-SILC 2010	1
Table des matières	3
1 Indicateurs communautaires longitudinaux communs fondés sur la dimension longitudinale des statistiques eu-silc	5
1.1 Taux de risque persistant de pauvreté selon l'âge et le sexe (à 60 % de la médiane)	5
1.2 Taux de risque persistant de pauvreté selon l'âge et le sexe (à 50 % de la médiane)	5
2 Précision.....	7
2.1 Plan d'échantillonnage.....	7
2.1.1 Type de plan d'échantillonnage	7
2.1.2 Unités d'échantillonnage.....	7
2.1.3 Critères de stratification et de sous-stratification.....	7
2.1.4 Taille de l'échantillon et critères d'attribution.....	11
2.1.5 Systèmes de sélection de l'échantillon.....	11
2.1.6 Répartition de l'échantillon au cours du temps et renouvellement de l'échantillon : groupes de rotation	11
2.1.7 Renouvellement de l'échantillon : groupes de rotation.....	12
2.1.8 Pondération.....	12
2.1.9 Substitutions	16
2.2 Erreurs d'échantillonnage	17
2.3 Erreurs non dues à l'échantillonnage.....	19
2.3.1 Erreurs dans la base de sondage et la couverture	19
2.3.2 Erreurs de mesure et de traitements	20
2.3.3 Erreurs de non-réponse	25
2.4 Mode de collecte des données	51
2.5 Les imputations	54
2.6 Loyer imputé	57
3 Comparabilité	59
3.1 Concepts et définitions de base.....	59
3.2 Composantes du revenu	60
3.2.1 Les différences entre les définitions nationales et les définitions type des statistiques EU-SILC 60	
3.2.2 La source ou la procédure utilisée pour la collecte des variables du revenu.....	63
3.2.3 La forme sous laquelle ont été obtenues les variables du revenu au niveau des composantes	63
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).....	63
3.3 Règles de suivi.....	64
4 Cohérence.....	65
4.1 Comparaison des variables cibles avec des sources extérieures	65

4.1.1	Salaire et revenu assimilé (PY010N).....	65
4.1.2	Les allocations chômage (PY090N)	65
4.1.3	Retraites de régime privé (PY080N), allocations vieillesse (PY100N) et pension de réversion (PY110N).....	66
	Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante	66
4.1.4	Revenus sociaux (HY050G, HY060G et HY070G).	66
4.1.5	Impôts (impôt sur le revenu des personnes physiques et taxe d'habitation).	66
4.1.6	Loyers	66
4.1.7	Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	66
4.1.8	Revenu disponible du ménage et niveau de vie des individus.....	67

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

1.1 Taux de risque persistant de pauvreté selon l'âge et le sexe (à 60 % de la médiane)

Le taux de risque persistant de pauvreté est le pourcentage de personnes ayant un revenu disponible équivalent inférieur à 60 % du revenu national médian équivalent durant l'année en cours et durant au moins deux des trois années précédentes.

La France ne publie temporairement plus le taux de risque persistant de la pauvreté du fait du changement de méthodologie intervenu en 2008 sur le mode de collecte des revenus et ce jusqu'à la collecte 2011.

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

Ce changement de collecte permet de disposer en particulier d'une meilleure information sur les revenus du patrimoine et foncier qui étaient mal recensés par voie de collecte et une meilleure qualité globale pour les autres revenus. Cependant selon la source, certaines personnes qui seraient en dessous du seuil de pauvreté avec les données d'enquête ne le sont plus avec les données fiscales, et vice-versa : le pourcentage d'individus concernés a été évalué à 9 % (cf. Comparability Study, Laurence Dauphin), environ la moitié de ces individus passerait en dessous du seuil de pauvreté et la moitié au-dessus. 40 % changerait de décile de niveau de vie selon les données utilisées et 18 % s'écarterait de plus d'un décile de leur décile d'origine. Si en transversal, l'impact est limité, en dynamique cela n'est donc pas sans conséquence en particulier pour les parcours de pauvreté.

Pour mémoire :

% des individus en risque de pauvreté en 2006 et dans cette situation au moins 2 années au cours des 3 années qui précèdent

Age	Sexe	%
Total	Total	6,5
	M	6,0
	F	6,9
0-17	Total	6,5
18-64	Total	5,8
	M	5,4
	F	6,2
65 et +	Total	8,5
	M	8,3
	F	8,7

1.2 Taux de risque persistant de pauvreté selon l'âge et le sexe (à 50 % de la médiane)

Pour mémoire :

% des individus en risque de pauvreté en 2006 et dans cette situation au moins 2 années au cours des 3 années qui précèdent

Age	Sexe	%
Total	Total	2,7
	M	2,6
	F	2,7
0-17	Total	2,0
18-64	Total	2,8
	M	2,9
	F	2,8
65 et +	Total	2,7
	M	2,7
	F	2,8

2 PRÉCISION

2.1 Plan d'échantillonnage

2.1.1 Type de plan d'échantillonnage

Jusqu'à l'enquête 2009, le système d'échantillonnage de l'enquête SILC en France reposait exclusivement sur l'échantillon-maître (EM) issu des données du dernier recensement exhaustif de 1999 et la base de logements neufs (BSLN). Depuis 2004, la France est passée à une nouvelle méthode : le recensement rotatif continu. Désormais, les communes de moins de 10 000 habitants ou « petites communes », sont recensées exhaustivement tous les cinq ans par roulement : pour cela, cinq groupes de rotation ont été définis aléatoirement, dans lesquels ont été réparties ces petites communes. Pour ce qui est des communes comprenant 10 000 habitants ou plus, ou « grandes communes », elles font l'objet d'une enquête de recensement plus complexe, par sondage chaque année au taux moyen de 8 %. Les adresses de ces communes sont réparties aléatoirement entre cinq groupes de rotation disjoints : chaque année, des logements appartenant à un échantillon d'adresse puisé dans le groupe de rotation « actif » sont recensés. Au total, cinq groupes de rotations ont ainsi été constitués.

Ce nouveau recensement conduit à une modification radicale des méthodes employées pour la construction des échantillons des enquêtes. Un nouvel échantillon-maître, OCTOPUSSE¹, adapté au contexte original du nouveau recensement a donc été mis au point².

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

- Le premier, tiré à partir de l'ancien échantillon-maître et de la BSLN, concerne les individus déjà présents dans l'échantillon en 2009. L'ancien échantillon-maître était constitué par une réserve de logements regroupant, d'une part des logements recensés en mars 1999, d'autre part des logements achevés après cette date, appelés « logements neufs ».
- Le second, tiré depuis le nouvel échantillon-maître, concerne seulement le sous échantillon entrant. L'échantillon-maître OCTOPUSSE est constitué à partir de deux listes de logements issues de la dernière enquête annuelle de recensement disponible. La première liste est celle relative aux logements situés dans une grande commune, la seconde celle relative aux logements localisés dans une petite commune.

2.1.2 Unités d'échantillonnage

L'unité d'échantillonnage gérée dans ces bases de sondage est le logement. Lorsque l'échantillon est entrant, on interroge l'ensemble des individus de chaque ménage présent dans le logement. L'échantillonnage de logements est stratifié, à plusieurs degrés : selon les strates, il y a deux ou trois degrés de tirage.

2.1.3 Critères de stratification et de sous-stratification

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

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

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

¹ Organisation Coordinée de Tirages Optimisés Pour une Utilisation Statistique des Échantillons.

² M. Christine et S. Faivre « OCTOPUSSE : un système d'Échantillon-Maître pour le tirage des échantillons dans la dernière Enquête Annuelle de Recensement » *Journées de méthodologie statistique 2009*.

- unités urbaines de plus de 100 000 habitants (« grand urbain »),
- unités urbaines de Paris

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

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

Ensuite, il n'y a pas de stratification préalable au second degré en zone rurale et pour les unités urbaines de moins de 20 000 habitants. En revanche, dans les strates urbaines, il y a une sous-stratification constituée par des groupes de communes. Il n'y a pas de critères sociodémographiques utilisés pour définir ces groupes de communes, on peut seulement dire qu'il s'agit de communes contiguës regroupées afin d'atteindre une certaine taille - mais cela permet au moins de distinguer ville-centre et banlieue. Le tirage est stratifié par groupes de communes pour garantir la dispersion géographique des districts. Il s'agit d'un tirage à probabilités égales des districts dans l'UU et équilibré par la méthode du Cube sur les critères de revenu et d'âge.

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

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

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

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

Le partage en régions administratives constitue la première stratification. Dans le cas de l'île de France, une sous-stratification permet de séparer la grande de la petite couronne. Les unités primaires sont les zones d'action enquêteurs (ZAE)³. Le tirage est également équilibré sur des *totaux régionaux*. Il est nécessaire d'équilibrer non seulement au niveau de l'ensemble de la ZAE mais aussi de chacun des 5 groupes de rotation, de manière à avoir chaque année une base de sondage « représentative ».

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

³ Ces nouvelles unités primaires diffèrent donc des unités primaires constituées à partir de l'ancien recensement.

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

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

Synthèse sur la constitution de l' Echantillon-Maître1999 (EM99) et le tirage d'un échantillon pour une enquête ménage

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

Vocabulaire

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

2.1.4 Taille de l'échantillon et critères d'attribution

L'échantillon 2004 comprenait 16 000 logements (ventilés en 9 sous-échantillons), et les sous-échantillons entrant entre 2005 et 2010 en comprenait chacun 3 000.

2.1.5 Systèmes de sélection de l'échantillon

L'échantillon de logements de SILC est tiré à probabilités égales (taux f) à l'exception des logements recensés secondaires qui sont tirés avec la probabilité $f/4$ et des logements recensés vacants en strate rurale qui sont tirés au taux $f/2$ jusqu'en 2005 (à partir de 2006 taux f pour les logements vacants urbains et $f/18$ pour les logements vacants ruraux).

Cas de la strate rurale

Le tirage des 128 unités primaires de l'échantillon-maître parmi les 2 227 recensées a été effectué proportionnellement à la taille définie en nombre de résidences principales, selon un algorithme de tirage équilibré (l'équilibrage a porté sur des structures par âge et sur le revenu déclaré aux services fiscaux). Le tirage de 719 unités secondaires de l'échantillon de EU-SILC sur les 3 217 unités de l'échantillon-maître a également été proportionnel à la taille en nombre de résidences principales. Le tirage des logements est effectué au sein du groupe de communes par un algorithme systématique sur fichier trié.

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

Cas des strates urbaines jusqu'à 100 000 habitants

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

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

Cas du « grand urbain »

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

Pour l'échantillon entrant en 2010 :

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

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

2.1.7 Renouvellement de l'échantillon : groupes de rotation

En 2010, l'échantillon entrant était constitué de 3 000 logements tirés de la façon décrite plus haut. L'échantillon des réinterrogés était constitué des logements où habitaient en 2009 au moins un individu panel n'appartenant pas à l'échantillon sortant et ayant répondu à l'enquête en 2009 ou 2008 (on conserve dans l'échantillon les non répondants de l'année précédente sauf les refus et les non répondants deux années de suite).

2.1.8 Pondération

En 2010, nous avons estimé pour la cinquième 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 W_i , varient dans un rapport de 1 à 4 pour les logements non vacants 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 P_i .

La correction de la non-réponse initiale

La correction de la non-réponse initiale pour les entrants en 2010

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 critères ou variables affectant significativement 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).

La probabilité de non-réponse initiale a été estimée de la même façon sur les échantillons entrants en 2005, 2006, 2007, 2008 et 2009 : le nombre de groupe et les motifs principaux de non réponse variant à la marge.

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 2010. 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.

▪ En 2010, trois modèles ont été estimés :

- un premier modèle portant sur la probabilité de répondre en 2010 sachant qu'on a répondu en 2009.
- un second modèle portant sur la probabilité de répondre en 2010 sachant qu'on a répondu en 2008.
- un troisième modèle portant sur la probabilité de répondre en 2010 sachant qu'on a répondu en 2007.

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

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 (maghrébine/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.

Les individus hors champ en 2010 sont considérés comme répondants en 2010 donc $PR(2010)=1$ et pour les enfants, pour lesquels nous n'avons que peu d'information, nous appliquons un taux de réponse uniforme (probabilité de réponse moyenne).

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

- Est attribué ensuite à chaque individu panel appartenant à l'échantillon de réinterrogation un poids corrigé de la non réponse selon la formule suivante :

$$WC_i(t',t) = \frac{W_i(t,k)}{L_i(t',t) \times \Phi_i(t,k) \times \theta_i(t,k) \times \alpha(t,t',k)}$$

où $W_i(t,k)$ est le poids de sondage « brut » : il s'agit du poids de sondage du logement dans lequel se trouve i à la date de son tirage en tant qu'individu-panel, donc lors du tirage annuel

$\Phi_i(t,k)$ est la probabilité de répondre en t en k ème interrogation⁴

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

$L_i(t',t)$ est le nombre de sous-échantillon dans lequel l'individu i aurait pu être tiré

$\alpha(t,t',k)$ égal à la proportion de répondants en t' parmi les répondants en t de l'échantillon $a_{t,k}$.

Trois jeux de pondérations pour Eurostat sont calculés :

$$WC(2007_2010) = W_i(2007,k) / ((L_i(2007,2010) * P_i(2007,k) * PR(2010/2007,k) * \alpha(2010,2007,k))$$

$$WC(2008_2010) = W_i(2008,k) / ((L_i(2008,2010)) * P_i(2008,k) * PR(2010/2008,k) * \alpha(2010,2008,k))$$

$$WC(2009_2010) = W_i(2009,k) / ((L_i(2009,2010)) * P_i(2009,k) * PR(2010/2009,k) * \alpha(2010,2009,k))$$

On affecte aux nouveau-nés panel le poids de leur mère.

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

Nous ne disposons pas de marges pour la population de référence de l'échantillon longitudinal, qui est l'ensemble des personnes ayant vécu de 2007 à 2010 en ménage ordinaire en France métropolitaine. Nous avons effectué un calage sur les marges de 2007 en prenant en compte l'ensemble des personnes sortis du champ entre les enquêtes (ceux ayant répondu en 2010 et ceux sortis définitivement du champ en 2010). Le poids des personnes hors champ (c'est-à-dire sortis définitivement) est ensuite mis à zéro. On effectue plusieurs calages correspondants aux différentes pondérations longitudinales calculées : soit WC07_10, WC08_10 et WC09_10.

⁴ $\Phi_i(t,k)$ et $\theta_i(t,k)$ sont les probabilités calculées pour estimer les poids transversaux.

Les marges utilisées sont celles constituées à partir des 4 enquêtes emploi trimestrielles de l'année n - 1. Ces marges reflètent une situation « moyenne » 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
 - 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 :
 - Française,
 - Étrangère.

2.1.8.4 Pondération longitudinale finale

Pour les répondants, le poids est modifié pour corriger de la non-réponse individuelle en 2007, 2008, 2009 ou 2010. 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 %, l'impact de cette correction est très limité.

2.1.9 Substitutions

Sans objet dans le cas de la France.

2.2 Erreurs d'échantillonnage

2007	Moyenne (en euros)	Nombre d'observations		Écart- type ⁵
		Avant imputation	Après imputation	
Revenu disponible brut (HY010)	36327	7736	7736	
Revenu disponible total (HY020)	28739	7736	7736	156
Revenu disponible total avant prestations sociales autres que retraites (HY022)	25398	7669	7669	161
Revenu disponible total avant prestations sociales (HY023)	18102	7496	7496	155
Revenus tirés de la location d'une propriété ou de terres (HY040N)	425	657	657	31
Allocation famille/enfant (HY050N)	964	936	2188	11
Exclusion sociale non classée ailleurs (HY060N)	184	233	383	4
Aides au logement (HY070N)	607	1171	1813	5
Transferts internénages perçus régulièrement en espèces (HY080N)	237	457	457	6
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	741	5998	5998	19
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	155	155	0
Intérêts (HY100N)	582	1663	1663	10
Impôts réguliers sur la fortune (HY120N)	-74	120	120	10
Transferts internénages versés régulièrement en espèces (HY130N)	-228	666	666	9
Remboursements consécutifs à des ajustements fiscaux (HY145N)	-1989	6777	6777	46
Composantes du revenu net au niveau individu				
Salaire ou revenu assimilé (PY010N)	10310	7463	8032	118
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	956	624	624	63
Autoconsommation (PY070N)	40	1628	1628	2
Allocations chômage(PY090N)	487	1092	1092	22
Retraites (PY100N)	4226	3736	3736	66
Pension de réversion (PY110N)	85	149	149	6
Indemnités de maladie (PY120N)	150	645	645	13
Pensions d'invalidité (PY130N)	227	449	449	13
Allocation d'éducation (PY140N)	21	318	318	2

⁵ Calcul sur l'échantillon 2007 (cf Rapport transversal 2007)

2010	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Revenu disponible brut (HY010)	44852	9050	9064	
Revenu disponible total (HY020)	36799	9060	9065	
Revenu disponible total avant prestations sociales autres que retraites (HY022)	33271	8919	9022	
Revenu disponible total avant prestations sociales (HY023)	24069	8913	8945	
Revenus tirés de la location d'une propriété ou de terres (HY040N)	958	1153	1293	
Allocation famille/enfant (HY050N)	966	2309	2566	
Exclusion sociale non classée ailleurs (HY060N)	275	698	724	
Aides au logement (HY070N)	476	1082	1850	
Transferts interménages perçus régulièrement en espèces (HY080N)	164	422	423	
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	3546	8331	8336	
Revenus perçus par les personnes de moins de 16 ans (HY110N)	4	164	164	
Intérêts (HY100N)	660	1927	1927	
Impôts réguliers sur la fortune (HY120N)	-47	150	150	
Transferts interménages versés régulièrement en espèces (HY130N)	-266	200	562	
Remboursements consécutifs à des ajustements fiscaux (HY145N)	-1797	8257	8257	
Composantes du revenu net au niveau individu				
Salaire ou revenu assimilé (PY010N)	10878	9246	10034	
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	953	815	863	
Autoconsommation (PY070N)	431	1850	1850	
Allocations chômage(PY090N)	536	1469	1627	
Retraites (PY100N)	5555	5205	5205	
Pension de réversion (PY110N)	28	69	69	
Indemnités de maladie (PY120N)	133	805	805	
Pensions d'invalidité (PY130N)	252	635	635	
Allocation d'éducation (PY140N)	21	289	289	

Revenu disponible 2007	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Total	18189	19388	19388	139
Ménage d'une personne	16532	2227	2227	187
Ménage de 2 personnes	19942	5289	5289	234
Ménage de 3 personnes	19471	3630	3630	258
Ménage de 4 personnes et +	16946	8242	8242	187
0-24 ans	16606	6512	6512	95
25-34 ans	18119	2167	2167	173
35-44 ans	18474	2728	2728	171
45-54 ans	19827	2702	2702	192
55-64 ans	21332	2370	2370	228
65+ ans	17673	2907	2907	168
Hommes	18453	9356	9356	116
Femmes	17946	10030	10030	110

Revenu disponible 2010	Moyenne (en euros)	Nombre d'observations		Écart- type
		Avant imputation	Après imputation	
Total	23529	22147	22147	
Ménage d'une personne	21398	2745	2745	
Ménage de 2 personnes	26143	6405	6405	
Ménage de 3 personnes	24094	3912	3912	
Ménage de 4 personnes et +	22077	9085	9085	
0-24 ans	20924	7076	7077	
25-34 ans	21937	2413	2413	
35-44 ans	23057	3018	3018	
45-54 ans	25601	3046	3046	
55-64 ans	27651	3010	3010	
65+ ans	25054	3583	3583	
Hommes	23959	10664	10664	
Femmes	23126	11482	11483	

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.

/* Rien sur la qualité d'Octopusse ? */

2.3.2 Erreurs de mesure et de traitements

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

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

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

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

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

Structure du questionnaire

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

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

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

L'architecture du questionnaire est la suivante :

- Tronc commun des enquêtes ménages (TCM) :

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

Pour limiter les erreurs de collecte, l'équipe de conception a d'une part introduit dans CAPI des filtres et des contrôles, d'autre part autorisé une collecte en francs pour 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

L'appariement avec les sources fiscales et sociales permet de garantir une bonne exhaustivité et de limiter les doubles comptes. Pour les revenus non imposables et certains individus non appariés (jeunes adultes), la stratégie de questionnement sur les revenus reste la même que les années précédentes. Comme l'enquêteur n'est pas en mesure de connaître à priori toutes les sources de revenus du ménage et que ce dernier peut oublier d'en mentionner une, on met en œuvre une stratégie générale de « balayage » des revenus. Cette stratégie consiste à :

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

Utilisation des documents administratifs et fiscaux

Le questionnement est fonction des documents utilisés. Parmi ces documents, la ou les déclarations des revenus de l'année de référence 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

La catégorie socioprofessionnelle des actifs ou des anciens actifs est codée automatiquement après la collecte par un programme proche du système de codification automatique SICORE développé à l'INSEE. 85 % des professions sont codées automatiquement, les autres professions étant « reprises » à la main par une équipe spécialisée à l'Insee.

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 l'année précédente 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 2007

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

Le questionnaire de 2007 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2006 dans les directions régionales de Franche-Comté et de PACA. Il a été mené auprès de 116 ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2006 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2007.

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 mesure.

Tests du questionnaire 2008

Le questionnaire transversal de 2008 a bénéficié des bilans des collectes 2004 à 2007. Le bilan de la collecte 2008 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 Basse-Normandie, Bourgogne et Lorraine.

Le questionnaire de 2008 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2007 dans les directions régionales de Franche-Comté et de PACA. Il a été mené auprès de 116 ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2007 ainsi que le module secondaire.

Enfin, un test en bureau a été mené en février 2008.

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 mesure.

Tests du questionnaire 2009

Le questionnaire transversal de 2009 a bénéficié du bilan de la collecte de 2008. 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 Alsace, Haute Normandie et Nord.

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

Enfin, un test en bureau a été mené en février 2009. Ce type de test permet de se placer dans des configurations particulières qui n'ont pas été rencontrées lors des tests sur le terrain.

Tests du questionnaire 2010

Le questionnaire transversal de 2010 a bénéficié du bilan de la collecte de 2009. 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 Franche Comté et Île-de-France.

Le questionnaire de 2010 a donné lieu à un test sur le terrain qui s'est déroulé en novembre 2009 dans les directions régionales du Nord-Pas de Calais et Languedoc-Roussillon. Il a été mené auprès de **148** ménages (répondants). Il avait pour objectif de tester le questionnaire transversal amélioré suite à la collecte de 2009 ainsi que le module secondaire et pour la première fois cette année, un questionnaire auto-administré papier contenant des questions subjectives portant sur le « bien-être de l'individu ».

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 2007/2010), l'équipe de conception ne forme pas directement les enquêteurs mais les gestionnaires responsables de l'enquête dans les directions régionales. Les gestionnaires des directions régionales forment ensuite les enquêteurs SILC. Cependant, la formation des enquêteurs reproduit à l'identique la formation des gestionnaires. L'équipe de conception SILC fournit aux gestionnaires des directions régionales (en plus des documents de collecte) les transparents et les exercices CAPI. La formation dispensée aux gestionnaires et aux enquêteurs a duré trois jours.

Lors de la collecte, chaque nouvel enquêteur et la majorité des anciens doivent être accompagnés 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 2007 à 2010.

	Nombre de ménages pour lesquels un entretien est accepté pour la base de données (DB135=1)			
	(DB010=2007)	(DB010=2008)	(DB010=2009)	(DB010=2010)
Nombre de ménages	7736	8861	9923	9066
	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=2007)	(DB010=2008)	(DB010=2009)	(DB010=2010)
Nombre de personnes	14794	16893	18663	17100
Nombre d'individus panels (RB100=1)	14418	16361	17991	16303
Nombre de co-résidents (RB100=2)	376	532	672	797

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 2007 (sous échantillons 7 à 9 et 1 à 3)

Adresses contactées avec succès DB120=11	4404
Total DB120	4803
Adresse inexistante DB120=23	399
Adresses validées sélectionnées	4404
Taux de contact à l'adresse Ra	100%
Entretiens de ménages achevés et acceptés pour la base de données DB135=1	7736
Ménages éligibles aux adresses contactées	8220
Total DB130	
Proportion d'entretiens de ménages achevés et acceptés pour la base de données Rh	94%
Taux de non-réponse des ménages NRh ($= [1 - (Ra * Rh)] * 100$)	5,9%

Taux de réponse des ménages : comparaisons des codes résultat des vagues 2007 et 2008

		2008											
		DB130=11											
		Questionnaire rempli											
2007		DB135=1	DB135=2	DB120=22	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	DB120=23	NC DB110 in (3 à 7) Ménage a quitté le panel	DB110=10	Total
		Entretien accepté pour la base de données	Entretien rejeté pour la base de données	Impossible d'accéder à l'adresse	Ens. Du ménage absent	Ménage incapable de répondre	Autres raisons	Refus de coopérer	Impossible de localiser l'adresse	Adresse inexistante		Fusion de ménages	
DB130=11 Questionnaire rempli	DB135=1 Entretien accepté	6712	9	61	15	112	53	536	0	0	68	0	7566
	DB135=2 Entretien rejeté	3	1	1	0	0	0	0	0	0	0	0	5
DB120=21													0
Impossible de localiser l'adresse													
DB120=22 Impossible d'accéder à l'adresse													
DB120=23													0
Adresse inexistante													
DB130=21													0
Refus de coopérer													
DB130=22													0
Ensemble du ménage absent													
DB130=23													0
Ménage incapable de répondre													
DB130=24													0
Autres raisons													
Total		6715	10	62	15	112	53	536	0	0	68	0	7571
		2008											
		DB130=11											
		DB135=1	DB135=2	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	DB120=22	DB120=23	NC DB110 in (3 à 7)	DB110=10	Total
2007													
DB110=8 Eclatement de ménage		108	0	6	2	5	12	0	8	0			141
DB110=9 Nouvelle adresse ou 1ère vague		1630	2	3	27	17	66	0	20	0			1765
Total		8453	12	24	141	75	614	0	90	0	68	0	9477

Nombre total de ménages	9477	100%
Taux de réponse de la vague 2008 (DB135=1)	8453	89%
Refus (DB130=21)	614	6%
Taux de suivi longitudinal	92%	
Ratio de suivi	115%	
Ratio taille finale échantillon	112%	

Le taux de suivi longitudinal correspond au pourcentage de ménages passés à la vague $t+1$ à des fins de suivi parmi les ménages passés de la vague $t-1$ à la vague t , à l'exclusion des ménages hors champ (selon les règles de suivi) ou inexistants.

Le ratio de suivi est le rapport du nombre de ménages passés de la vague t à la vague $t+1$ avec le nombre de ménages passés à des fins de suivi de la vague $t-1$ à la vague t .

Le **ratio taille finale d'échantillon** est le rapport entre le nombre de ménages acceptés pour la base de données (DB135=1) dans la vague t et le nombre de ménage acceptés pour la base de données (DB135=1) dans la vague $t-1$.

Taux de réponse des ménages : comparaisons des codes résultat des vagues 2008 et 2009

		2009											
		DB130=11											
		Questionnaire rempli											
2008		DB135=1	DB135=2	DB120=22	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	DB120=23	NC DB110 in (3 à 7) Ménage a quitté le panel	DB110=10	Total
		Entretien accepté pour la base de données	Entretien rejeté pour la base de données	Impossible d'accéder à l'adresse	Ens. Du ménage absent	Ménage incapable de répondre	Autres raisons	Refus de coopérer	Impossible de localiser l'adresse	Adresse inexistante		Fusion de ménages	
DB130=11 Questionnaire rempli	DB135=1 Entretien accepté	7746	17	0	8	120	68	558	4	0	79	11	8611
	DB135=2 Entretien rejeté	0	0	0	0	0	0	0	0	0	0	0	0
DB120=21 Impossible de localiser l'adresse													0
DB120=22 Impossible d'accéder à l'adresse													
DB120=23													0
Adresse inexistante													
DB130=21													0
Refus de coopérer													
DB130=22													0
Ensemble du ménage absent													
DB130=23													0
Ménage incapable de répondre													
DB130=24													0
Autres raisons													
Total		7746	17	0	8	120	68	558	4	0	79	11	8611
		2009											
		DB130=11											
		DB135=1	DB135=2	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	DB120=22	DB120=23	NC DB110 in (3 à 7)	DB110=10	Total
2009													
DB110=8 Eclatement de ménage		136	1	1	2	4	9	0	0	0			153
DB110=9 Nouvelle adresse ou 1ère vague		1695	5	1	28	14	70	1	0	0			1814
Total		9577	23	10	150	86	637	5	0	0	79	11	10578

Nombre total de ménages			10578	100%
Taux de réponse de la vague 2007 (DB135=1)			9577	91%
Refus (DB130=21)			637	6%
Taux de suivi longitudinal		92%		
Ratio de suivi		114%		
Ratio taille finale échantillon		111%		

Taux de réponse des ménages : comparaisons des codes résultat des vagues 2009 et 2010

		2010											
		DB130=11 Questionnaire rempli											
2009		DB135=1 Entretien accepté pour la base de données	DB135=2 Entretien rejeté pour la base de données	DB120=22 Impossible d'accéder à l'adresse	DB130=22 Ens. 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	DB120=23 Adresse inexistante	NC DB110 in (3 à 7) Ménage a quitté le panel	DB110=10 Fusion de ménages	Total
DB130=11 Questionnaire rempli	DB135=1 Entretien accepté	8768	11	0	7	168	63	548	85	671	87	23	10431
	DB135=2 Entretien rejeté	0	0	0	0	0	0	0	0	0	0	0	0
DB120=21 Impossible de localiser l'adresse													0
DB120=22 Impossible d'accéder à l'adresse													
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		8768	11	0	7	168	63	548	85	671	87	23	10431
		2010											
		DB130=11											
2009		DB135=1	DB135=2	DB130=22	DB130=23	DB130=24	DB130=21	DB120=21	DB120=22	DB120=23	NC DB110 in (3 à 7)	DB110=10	Total
DB110=8 Eclatement de ménage		147	0	1	3	5	11	5	0	0			172
DB110=9 Nouvelle adresse ou 1ère vague		1660	2	2	38	14	58	14	0	0			1788
Total		10575	13	10	209	82	617	104	0	671	87	23	12391

Nombre total de ménages	12391	100%
Taux de réponse de la vague 2007 (DB135=1)	10575	85%
Refus (DB130=21)	617	5%
Taux de suivi longitudinal	86%	
Ratio de suivi	104%	
Ratio taille finale échantillon	101%	

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

Taux de réponse individuelle 2008

		2008								Total
		RB250=11	Non complété							
		Entretien individuel achevé	RB250=21-22 Entretien n'a pas pu être réalisé	RB250=23 Refus de coopérer	RB250=31 Non contacté	RB250=32 Non contacté	RB250=33 Non contacté	HH nc	Pn RB110=6 ou RB120=2,3	
Individus panel issus de la vague précédente (RB100 = 1 et RB245=1)										
RB110=1-2										
Membre courant du ménage		12 652	0	35	0	17	0			12 704
RB110=6										
décédé										45
RB110=-1										0
RB120=2										
Parti en institution										17
RB120=3										
Parti à l'étranger										15
RB120=4										
Perdu										97
DB135=2 ou -1, ou DB110=7, ou DB120=21-23 ou -1, ou DB130=21-24 ou -1										79
DB110=3-6										2
Nouveaux individus panel										
Ont atteint 16 ans		242	0	3	0	0	0	0	0	245
Individus panel supplémentaires (entrant 2008)		3 779	0	6	0	9	0			3 794
Individus non panel de plus de 16 ans										
Vague 2008	Présents en 2007	302	0	11	0	2	1	0	0	316
	Non présents en 2007	249	0	9	0	5	0	0	0	263
Vague 2008	Présents en 2004									0
	Non présents en 2004									0
Individus panel 2007 non présents en 2008										
Individus panel 2007 non présents en 2008		164		0		0				164
SOMMES										
1+3+6+7+9+10		16 673	0	44	0	26	0	0	0	16 919
1+3+6+7+9+10+13		16 837	0	44	0	26	0	0	0	17 083
1+3+6+7+9+10+11		17 224	0	64	0	33	1	0	0	17 498

Taux de réponse des individus panel de la vague 2008	98,5%
Taux de réponse des co-résidents (sélectionnés en 2007) de la vague 2008	95,6%
Taux de suivi longitudinal (RB250=11)	98,6%
Refus (RB250=23)	0,26%
Non contactés et autres (RB250=32)	0,15%
Ratio (taille finale 2008 / taille finale 2007) individus panel	101,61%
Ratio (taille finale 2008 / taille finale 2007) individus panel et co-résidents	102,16%
Ratio (taille finale 2008 / taille finale 2007) co-résidents (sélectionnés en 2007)	148,8%
Taux de réponse des personnes non-panel	95,16%

Taux de réponse individuelle 2009

	2009									Total
	RB250=11 Entretien individuel achevé	Non complété								
		RB250=21-22	RB250=23	RB250=31	RB250=32	RB250=33	HH nc	Pn	PI	
		Entretien n'a pas pu être réalisé	Refus de coopérer	Non contacté	Non contacté	Non contacté		RB110=6 ou RB120=2,3	RB110=4 or -1	
Individus panel issus de la vague précédente (RB100 = 1 et RB245=1)										
RB110=1-2										
Membre courant du ménage	14 374	0	28	0	17	11				14 430
RB110=6										
décédé										55
RB110=-1										0
RB120=2										
Parti en institution										20
RB120=3										
Parti à l'étranger										29
RB120=4										
Perdu										149
DB135=2 ou -1, ou DB110=7, ou DB120=21-23 ou -1, ou DB130=21-24 ou -1										59
DB110=3-6										0
Nouveaux individus panel										
Ont atteint 16 ans	224	0	1	0	0	0	0	0	0	225
Individus panel supplémentaires (entrant 2009)	3 707	0	7	0	16	3				3 733
Individus non panel de plus de 16 ans										
Vague 2009	Présents en 2008	411	0	9	0	3	0	0	0	423
	Non présents en 2008	274	0	12	0	2	2	0	0	290
Vague 2009	Présents en 2004									0
	Non présents en 2004									0
Individus panel 2008 non présents en 2009										
Individus panel 2008 non présents en 2009	194		8		0					202
SOMMES										
1+3+6+7+9+10	18 305	0	36	0	33	14	0	0	0	18 596
1+3+6+7+9+10+13	18 499	0	44	0	33	14	0	0	0	18 798
1+3+6+7+9+10+11	18 990	0	57	0	38	16	0	0	0	19 309

Taux de réponse des individus panel de la vague 2009	98,4%
Taux de réponse des co-résidents (sélectionnés en 2008) de la vague 2009	97,2%
Taux de suivi longitudinal (RB250=11)	98,4%
Refus (RB250=23)	0,23%
Non contactés et autres (RB250=32)	0,18%
Ratio (taille finale 2009 / taille finale 2008) individus panel	109,79%
Ratio (taille finale 2009 / taille finale 2008) individus panel et co-résidents	110,25%
Ratio (taille finale 2009 / taille finale 2008) co-résidents (sélectionnés en 2008)	136,1%
Taux de réponse des personnes non-panel	96,07%

Taux de réponse individuelle 2010

	2010									Total
	RB250=11	Non complété								
	Entretien individuel achevé	RB250=21-22 Entretien n'a pas pu être réalisé	RB250=23 Refus de coopérer	RB250=31 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)										
RB110=1-2										
Membre courant du ménage	16 037	0	1	0	41	27				16 106
RB110=6										
décédé										54
RB110=-1										0
RB120=2										
Parti en institution										17
RB120=3										
Parti à l'étranger										23
RB120=4										
Perdu										176
DB135=2 ou -1, ou DB110=7, ou DB120=21-23 ou -1, ou DB130=21-24 ou -1										55
DB110=3-6										0
Nouveaux individus panel										
Ont atteint 16 ans	271	0	2	0	0	0	0	0	0	273
Individus panel supplémentaires (entrant 2010)	268	0	4	0	272	0				544
Individus non panel de plus de 16 ans										
Vague 2010	Présents en 2009	542	0	8	0	7	0	0	0	557
	Non présents en 2009	277	0	16	0	9	302	0	0	604
Vague 2010	Présents en 2004									0
	Non présents en 2004									0
Individus panel 2009 non présents en 2010										
Individus panel 2009 non présents en 2010	184		184		0					368
SOMMES										
1+3+6+7+9+10	16 576	0	7	0	313	27	0	0	0	17 154
1+3+6+7+9+10+13	16 760	0	191	0	313	27	0	0	0	17 522
1+3+6+7+9+10+11	17 395	0	31	0	329	329	0	0	0	18 315

Taux de réponse des individus panel de la vague 2010	96,6%
Taux de réponse des co-résidents (sélectionnés en 2009) de la vague 2010	97,3%
Taux de suivi longitudinal (RB250=11)	95,7%
Refus (RB250=23)	1,09%
Non contactés et autres (RB250=32)	1,79%
Ratio (taille finale 2010 / taille finale 2009) individus panel	90,55%
Ratio (taille finale 2010 / taille finale 2009) individus panel et co-résidents	91,60%
Ratio (taille finale 2010 / taille finale 2009) co-résidents (sélectionnés en 2009)	131,9%
Taux de réponse des personnes non-panel	70,54%

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 2007

		DB110=1	2	3	4	5	6	7	8	9	10
2007	Total	Même adresse en 2004 & 2005	Tout le ménage a déménagé (reste dans le champ)	Tout le ménage est parti en institution	Tout le ménage est parti à l'étranger	Tout le ménage est décédé	Le ménage ne contient pas d'individus panels	Adresse non contactée	Éclatement de ménage	Nouvelle adresse	Fusion
Total	8 966	5 804	490	18	28	26	0	149	186	2 250	15
%	100	64,7	5,5	0,2	0,3	0,3	0,0	1,7	2,1	25,1	0,2

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2008

		DB110=1	2	3	4	5	6	7	8	9	10
2008	Total	Même adresse en 2004 & 2005	Tout le ménage a déménagé (reste dans le champ)	Tout le ménage est parti en institution	Tout le ménage est parti à l'étranger	Tout le ménage est décédé	Le ménage ne contient pas d'individus panels	Adresse non contactée	Éclatement de ménage	Nouvelle adresse	Fusion
Total	11 288	7 142	597	20	28	32	1	201	215	3 037	15
%	100	63,3	5,3	0,2	0,2	0,3	0,0	1,8	1,9	26,9	0,1

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2009

		DB110=1	2	3	4	5	6	7	8	9	10
2009	Total	Même adresse en 2004 & 2005	Tout le ménage a déménagé (reste dans le champ)	Tout le ménage est parti en institution	Tout le ménage est parti à l'étranger	Tout le ménage est décédé	Le ménage ne contient pas d'individus panels	Adresse non contactée	Éclatement de ménage	Nouvelle adresse	Fusion
Total	12 536	8 227	707	20	27	46	1	224	240	3 028	16
%	100	65,6	5,6	0,2	0,2	0,4	0,0	1,8	1,9	24,2	0,1

Ventilation des ménages selon le « statut du ménage » (DB110) à la vague 2010

		DB110=1	2	3	4	5	6	7	8	9	10
2010	Total	Même adresse en 2004 & 2005	Tout le ménage a déménagé (reste dans le champ)	Tout le ménage est parti en institution	Tout le ménage est parti à l'étranger	Tout le ménage est décédé	Le ménage ne contient pas d'individus panels	Adresse non contactée	Éclatement de ménage	Nouvelle adresse	Fusion
Total	10 543	9 247	733	35	23	49	1	221	208	0	26
%	100	87,7	7,0	0,3	0,2	0,5	0,0	2,1	2,0	0,0	0,2

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2007

2007	Total	DB120=11	21	22	23	Manquant
		Adresse contactée	Adresse non localisée	Adresse inaccessible	Adresse inexistante	
Total (DB110=2, 8, 9)	3 605	3 121	135	0	349	0
%	100	86,6	3,7	0,0	9,7	0,0

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2008

2008	Total	DB120=11	21	22	23	Manquant
		Adresse contactée	Adresse non localisée	Adresse inaccessible	Adresse inexistante	
Total (DB110=2, 8, 9)	3 849	3 246	30	95	409	69
%	100	84,3	0,8	2,5	10,6	1,8

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2009

2009	Total	DB120=11	21	22	23	Manquant
		Adresse contactée	Adresse non localisée	Adresse inaccessible	Adresse inexistante	
Total (DB110=2, 8, 9)	3 975	3 456	57	0	384	78
%	100	86,9	1,4	0,0	9,7	2,0

Ventilation des ménages selon le « enregistrement du contact à l'adresse » (DB120) à la vague 2010

2010	Total	DB120=11	21	22	23	Manquant
		Adresse contactée	Adresse non localisée	Adresse inaccessible	Adresse inexistante	
Total (DB110=2, 8, 9)	941	820	119	0	0	2
%	100	87,1	12,6	0,0	0,0	0,2

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2007

2007	Total	DB130=11	21	22	23	24	Manquant
		Complété	Refus de coopérer	Tout le ménage est temporairement absent	Tout le ménage est incapable de répondre	Autre motif de non réponse	
Total (DB120=11 ou DB110=1)	8220	7742	243	36	123	76	0
%	100	94,2	3,0	0,4	1,5	0,9	0,0

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2008

2008	Total	DB130=11	21	22	23	24	Manquant
		Complété	Refus de coopérer	Tout le ménage est temporairement absent	Tout le ménage est incapable de répondre	Autre motif de non réponse	
Total	10388	8872	887	248	251	130	0
DB120=11 ou DB110=1)							
%	100	85,4	8,5	2,4	2,4	1,3	0,0

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2009

2009	Total	DB130=11	21	22	23	24	Manquant
		Complété	Refus de coopérer	Tout le ménage est temporairement absent	Tout le ménage est incapable de répondre	Autre motif de non réponse	
Total	11683	9943	992	242	254	135	117
DB120=11 ou DB110=1)							
%	100	85,1	8,5	2,1	2,2	1,2	1,0

Ventilation des ménages selon le « résultat du questionnaire ménage » (DB130) à la vague 2010

2010	Total	DB130=11	21	22	23	24	Manquant
		Complété	Refus de coopérer	Tout le ménage est temporairement absent	Tout le ménage est incapable de répondre	Autre motif de non réponse	
Total	10067	9080	650	32	213	92	0
DB120=11 ou DB110=1)							
%	100	90,2	6,5	0,3	2,1	0,9	0,0

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2007

2007	Total	DB135=1	2	3
		Accepté	Refusé	Manquant
Total	7742	7736	6	0
(DB130=11)				
%	100	99,9	0,1	0,0

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2008

2008	Total	DB135=1	2	3
		Accepté	Refusé	Manquant
Total	8872	8861	11	0
(DB130=11)				
%	100	99,9	0,1	0,0

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2009

2009	Total	DB135=1	2	3
		Accepté	Refusé	Manquant
Total	9943	9923	20	0
(DB130=11)				
%	100	99,8	0,2	0,0

Ventilation des ménages selon l'« acceptation de l'entretien » (DB135) à la vague 2010

2010	Total	DB135=1	2	3
		Accepté	Refusé	Manquant
Total (DB130=11)	9080	9066	14	0
%	100	99,8	0,2	0,0

2.3.3.4 Distribution des personnes en fonction de leur statut (RB110) et distribution des personnes ayant déménagé par destination (RB120)

Distribution des personnes en fonction de leur statut (RB110)

		Membres habituels du ménage				Membres non habituels du ménage				
2008	Total	RB110=1	RB110=2	RB110=3	RB110=4	RB110=5	RB110=6	RB110=7	Manquant	
		Etait présent à la vague précédente	Venu d'un autre ménage panel	Venu d'un ménage non panel	Naissance	A quitté le ménage	Décédé	Dans le ménage au moins 3 mois pendant la période de référence		
Total	22147	21037	178	347	160	380	45	0		0
%	100	95,0	0,8	1,6	0,7	1,7	0,2	0,0		0,0

		Membres habituels du ménage				Membres non habituels du ménage				
2009	Total	RB110=1	RB110=2	RB110=3	RB110=4	RB110=5	RB110=6	RB110=7	Manquant	
		Etait présent à la vague précédente	Venu d'un autre ménage panel	Venu d'un ménage non panel	Naissance	A quitté le ménage	Décédé	Dans le ménage au moins 3 mois pendant la période de référence		
Total	24496	23192	214	382	186	467	55	0		0
%	100	94,7	0,9	1,6	0,8	1,9	0,2	0,0		0,0

		Membres habituels du ménage				Membres non habituels du ménage				
2010	Total	RB110=1	RB110=2	RB110=3	RB110=4	RB110=5	RB110=6	RB110=7	Manquant	
		Etait présent à la vague précédente	Venu d'un autre ménage panel	Venu d'un ménage non panel	Naissance	A quitté le ménage	Décédé	Dans le ménage au moins 3 mois pendant la période de référence		
Total	22393	21070	186	405	200	478	54	0		0
%	100	94,1	0,8	1,8	0,9	2,1	0,2	0,0		0,0

Distribution des personnes ayant déménagé par destination (RB120)

RB110=5						
		RB120=1 Ménage ordinaire				
2008	Total	Membre courant d'un ménage	Membre non courant d'un ménage	RB120=2	RB120=3	RB120=4
				Collectivité	Etranger	Perdu
Total	380	244	7	17	15	97
%	100	64,2	1,8	4,5	3,9	25,5
RB110=5						
		RB120=1 Ménage ordinaire				
2009	Total	Membre courant d'un ménage	Membre non courant d'un ménage	RB120=2	RB120=3	RB120=4
				Collectivité	Etranger	Perdu
Total	467	248	23	20	29	147
%	100	53,1	4,9	4,3	6,2	31,5
RB110=5						
		RB120=1 Ménage ordinaire				
2010	Total	Membre courant d'un ménage	Membre non courant d'un ménage	RB120=2	RB120=3	RB120=4
				Collectivité	Etranger	Perdu
Total	478	242	22	17	23	174
%	100	50,6	4,6	3,6	4,8	36,4

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 2007

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	1,6	53.1
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et pension de réversion (HY022)	99,3	2,2	37.8
Revenu disponible des ménages avant transferts sociaux (HY023)	97.3	4.6	23.5
Revenus tirés de la location d'une propriété ou de terres (HY040N)	8,5	-	-
Allocation famille/enfant (HY050N)	28,1	14.0	0.7
Exclusion sociale non classée ailleurs (HY060N)	4,5	1,3	0,3
Aides au logement (HY070N)	22.6	7.7	0.0
Transferts interménages perçus régulièrement en espèces (HY080N)	5.8	-	-
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	79,3	-	-
Revenus perçus par les personnes de moins de 16 ans (HY110N)	2,0	-	-
Impôts réguliers sur la fortune (HY120N)	1,5	0.0	1.5
Transferts interménages versés régulièrement en espèces (HY130N)	9,2	-	-
Remboursements consécutifs à des ajustements fiscaux (HY145N)	88.3	6,2	15,4

	% 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)	54.3	3,8	0,6
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4,3	0.0	0,1
Allocations chômage(PY090N)	7,4	0.3	0.2
Retraites (PY100N)	25.0	5.3	0.7
Pensions de réversion (PY110N)	1,0	0.1	0,0
Indemnités de maladie (PY120N)	4,1	-	0,0
Pensions d'invalidité (PY130N)	3.0	0.0	0,1
Allocations d'éducation (PY140N)	2,1	-	-

Ventilation des non-réponses à certaines questions en 2008

	% de ménages...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Ménages</i>			
Revenu disponible total (HY020)	100.0	0.4	97.9
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et des pensions au survivant (HY022)	99.2	1.2	92.4
Revenu disponible des ménages avant transferts sociaux (HY023)	97.1	9.8	81.3
Revenus tirés de la location d'une propriété ou de terres (HY040N)	12.4	-	-
Allocation famille/enfant (HY050N)	27.4	27.4	-
Exclusion sociale non classée ailleurs (HY060N)	6.8	3.1	0.6
Aides au logement (HY070N)	23.7	2.3	12.3
Transferts interménages perçus régulièrement en espèces (HY080N)	5.2	-	-
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	79.1	0.1	69.1
Revenus perçus par les personnes de moins de 16 ans (HY110N)	1.8	-	-
Impôts réguliers sur la fortune (HY120N)	1.5	-	-
Transferts interménages versés régulièrement en espèces (HY130N)	6.4	5.0	0.8
Remboursements consécutifs à des ajustements fiscaux (HY145N)	87.1	12.3	58.6

	% de personnes de plus de 16 ans...		
	... ayant perçu un montant (après imputation)	... avec valeur manquante (avant imputation)	... avec information partielle (avant imputation)
<i>Individus</i>			
Salaire ou revenu assimilé (PY010N)	54.7	5.0	0.6
Avantages en nature (PY020N)	11.8	-	-
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	4.6	-	-
Allocations chômage(PY090N)	6.8	-	5.7
Retraites (PY100N)	28.3	-	14.0
Allocations au survivant (PY110N)	0.5	-	0.2
Indemnités de maladie (PY120N)	4.6	-	4.6
Pensions d'invalidité (PY130N)	1.9	-	-
Allocations d'éducation (PY140N)	1.9	-	1.9

Ventilation des non-réponses à certaines questions en 2009

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

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

Ventilation des non-réponses à certaines questions en 2010

Ménages	% de ménages...		
	...ayant perçu un montant (après imputation)	...avec valeur manquante (avant imputation)	...avec information partielle (avant imputation)
Revenu disponible total (HY020)	99,98*	0,8	97,3
Revenu disponible des ménages avant transferts sociaux à l'exception des allocations de vieillesse et des pensions au survivant (HY022)	99,6*	9,0	83,7
Revenu disponible des ménages avant transferts sociaux (HY023)	98,8*	5,7	79,4
Revenus tirés de la location d'une propriété ou de terres (HY040N)	14,2	2,0	2,4
Allocation famille/enfant (HY050N)	27,5	3,6	2,4
Exclusion sociale non classée ailleurs (HY060N)	8,1	0,4	2,6
Aides au logement (HY070N)	21,6	5,6	0,6
Transferts interménages perçus régulièrement en espèces (HY080N)	4,6	0,0	0,7
Intérêts, dividendes, bénéfices tirés d'investissements en capital dans les opérations incorporelles (HY090N)	91,3	0,2	1,9
Revenus perçus par les personnes de moins de 16 ans (HY110N)	1,7	0,0	0,0
Impôts réguliers sur la fortune (HY120N)	1,6	0,0	1,6
Transferts interménages versés régulièrement en espèces (HY130N)	6,3	4,1	1,6
Remboursements consécutifs à des ajustements fiscaux (HY145N)	92,9	0,0	77,5

*Les écarts à 100% concernent des ménages avec un revenu disponible calculé à zéro.

Individus	% 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)	57,9	4,4	0,3
Avantages en nature (PY020N)	13,1	0,0	0,0
Bénéfices en espèces ou perte de trésorerie en rapport avec une activité indépendante (PY050N)	5,0	0,3	0,3
Allocations chômage(PY090N)	9,3	1,0	8,3
Retraites (PY100N)	29,9	0,0	9,7
Allocations au survivant (PY110N)	0,4	0,0	0,1
Indemnités de maladie (PY120N)	4,6	0,0	4,6
Pensions d'invalidité (PY130N)	3,5	0,0	0,3
Allocations d'éducation (PY140N)	1,7	0,0	1,7

2.4 Mode de collecte des données

Le taux d'acceptation des individus appartenant aux ménages répondants est très élevé : 99,5 % pour 2007, 99,4 % pour 2008 et 2009 et 99,5 % pour 2010. 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 2007

2007	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	14967	14888	1	49	29
%	100	99,5	0,0	0,3	0,2

- 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 2007

2007	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	14559	14498	1	33	27
%	100	99,6	0,0	0,2	0,2

- 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 2007

2007	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	408	390	16	2	0
%	100	95,6	3,9	0,5	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 2008

2008	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	17073	16979	0	61	33
%	100	99,4	0,0	0,4	0,2

- 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 2008

2008	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	16499	16432	0	41	26
%	100	99,6	0,0	0,2	0,2

- 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 2008

		RB250=11	23	32	33
2008	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	575	547	20	7	1
%	100	95,1	3,5	1,2	0,2

- Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3) selon le statut des données (RB250), vague 2009

		RB250=11	23	32	33
2009	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	18901	18807	0	56	38
%	100	99,5	0,0	0,3	0,2

- 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 2009

		RB250=11	23	32	33
2009	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	18192	18124	0	35	33
%	100	99,6	0,0	0,2	0,2

- 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 2009

		RB250=11	23	32	33
2009	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	711	683	21	5	2
%	100	96,1	3,0	0,7	0,3

- Ventilation des membres de ménages répondants âgés de plus de 16 ans (RB245 = 1 à 3) selon le statut des données (RB250), vague 2010

		RB250=11	23	32	33
2010	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	17340	17223	1	67	49
%	100	99,3	0,0	0,4	0,3

- 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 2010

		RB250=11	23	32	33
2010	Total	Entretien individuel achevé	Refus de coopérer	Non contacté	Non contacté
Total	16486	16409	1	43	33
%	100	99,5	0,0	0,3	0,2

- 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 2010

2010	Total	RB250=11 Entretien individuel achevé	23 Refus de coopérer	32 Non contacté	33 Non contacté
Total	854	814	24	16	0
%	100	95,3	2,8	1,9	0,0

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 28,1 % en 2007, 28,0 % en 2008, 28,1 % en 2009 et 28,0 % en 2010. Le protocole de collecte n'autorisant pas une deuxième visite, ceci explique pour partie le recours élevé au proxy afin que les questionnaires individuels soient renseignés.

- 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 2007

2007	Total	RB260=2 CAPI	5 Proxy
Total	14888	10712	4176
%	100	72,0	28,0

- 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 2007

2007	Total	RB260=2 CAPI	5 Proxy
Total	14498	10508	3990
%	100	72,5	27,5

- 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 2007

2007	Total	RB260=2 CAPI	5 Proxy
Total	390	204	186
%	100	52,3	47,7

- 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 2008

2008	Total	RB260=2 CAPI	5 Proxy
Total	16979	12204	4775
%	100	71,9	28,1

- 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 2008

2008	Total	RB260=2 CAPI	5 Proxy
Total	16432	11943	4489
%	100	72,7	27,3

- 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 2008

2008	Total	RB260=2 CAPI	5 Proxy
Total	547	261	286
%	100	47,7	52,3

- 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 2009

2009	Total	RB260=2 CAPI	5 Proxy
Total	18807	13633	5174
%	100	72,5	27,5

- 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 2009

2009	Total	RB260=2 CAPI	5 Proxy
Total	18124	13317	4807
%	100	73,5	26,5

- 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 2009

2009	Total	RB260=2 CAPI	5 Proxy
Total	683	316	367
%	100	46,3	53,7

- 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 2010

2010	Total	RB260=2 CAPI	5 Proxy
Total	17223	12333	4890
%	100	71,6	28,4

- 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 2010

2010	Total	RB260=2 CAPI	5 Proxy
Total	16409	11953	4456
%	100	72,8	27,2

- 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 2010

2010	Total	RB260=2 CAPI	5 Proxy
Total	814	380	434
%	100	46,7	53,3

2.5 Les imputations

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

Les prestations familiales et les minima sociaux sont obtenus auprès des organismes les détenant, c'est toujours l'adresse du ménage qui permet la recherche et les caractéristiques démographiques de

l'allocataire (sexe, date de naissance) : pour diverses raisons certaines adresses ne correspondent pas, ces revenus sont alors imputés.

L'imputation est d'abord nécessaire parce qu'il existe des données manquantes ou en tranches. Les données manquantes proviennent des individus dont la déclaration fiscale n'a pu être retrouvée. Les données en tranches concernent les enfants rattachés fiscalement au foyer fiscal de leurs parents et ne résidant pas avec ceux-ci. Les revenus dans l'EU-SILC 2010 sont relatifs à plusieurs dates. L'impôt payé a pour assiette les revenus imposables perçus au cours de l'année 2008. Les revenus appariés et collectés sont relatifs à l'année 2009. 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.

Salaire ou revenu assimilé (PY010N)

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

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

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

Les salaires des non-salariés

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

Préretraites

Un petit nombre de préretraites sont à imputer : nous leur attribuons le montant moyen de préretraites d'individus ayant des caractéristiques similaires.

Allocations de vieillesse (PY100N) ou pension de réversion (PY100N)

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

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

Les retraites étant quasi stables, il est possible par ailleurs de contrôler le montant imputé. Comme nous ne disposons pas d'une retraite courante (relative au mois de l'enquête), nous utilisons le revenu courant comme élément de contrôle et nous le comparons à la somme des revenus de l'année de référence du ménage. Pour les ménages concernés, la retraite est un élément prépondérant du revenu total, ce qui justifie la comparaison. Nous ajoutons un autre contrôle, cette fois entre les années de revenu antérieures, toujours sous la même hypothèse nous comparons les impôts sur le revenu déclarés que nous calculons. Deux contrôles sont donc possibles avant de prendre une décision.

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

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

Ces revenus sont collectés sous deux formes : d'une part la forme fiscale, comprenant les amortissements et autres abattements, et d'autre part la forme privée, correspondant au revenu net déterminé par le ménage. Le revenu fiscal est jugé peu réaliste, c'est pourquoi le revenu privé collecté par voie d'enquête est privilégié (les prélèvements privés que la personne a effectué sur les ressources de son activité d'indépendant pour ses besoins de consommation ou d'épargne). Ainsi si les deux revenus sont renseignés et si le revenu privé est vraisemblable, le revenu privé est le seul pris en compte. À défaut de revenu privé, le revenu fiscal est retenu. Si les deux types de revenus sont manquants ou peu crédibles, l'imputation se fait par hot-deck.

Prestations familiales

Les prestations familiales sont, depuis la collecte 2008 sur les prestations 2007, récupérées par appariement auprès des organismes de prestations sociales. Dans le cas où l'allocataire n'a pu être apparié, les prestations sont calculées sur barème. La principale difficulté est la période de référence des revenus pour les prestations sous conditions de ressources. Pour les Caisses d'allocations familiales (Caf), jusque juillet d'une année N, les revenus retenus pour le calcul des aides est celui de l'année N-2 ; à partir de juillet N, les revenus retenus sont ceux de N-1, Nous utilisons uniquement les revenus de l'année N pour imputer les prestations de l'année N.

Les résultats sont conformes aux données des CAF. Les valeurs calculées sont imputées si l'individu n'a pas été retrouvé dans les fichiers sociaux.

Aides au logement

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

Minima sociaux

Au 01/06/2009 le RMI est remplacé par le Revenu de Solidarité Active (RSA). Le RSA assure un revenu minimum à des personnes qui ne travaillent pas (RSA socle) ou à des personnes qui travaillent déjà et dont les revenus sont faibles (RSA socle ou/et RSA activité). Le RSA remplace le RMI, l'API et certaines aides forfaitaires comme la prime de retour à l'emploi. En cas de non-

appariement, le montant est imputé sur barème si la personne a déclaré être bénéficiaire. Des contrôles de cohérence avec la source Cnaf sont réalisés.

Pour les allocataires non appariés, trois minima sociaux sont imputés dans SILC : le revenu minimum d'insertion (RMI) puis RSA, l'APSA ou ex 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 puis RSA est un revenu trimestriel que nous ne connaissons pas. Le RMI puis RSA imputé est égal à la différence entre le plafond du RMI puis RSA 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.

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 1948⁶. 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 ...) et le quintile de niveau de vie. 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 ;

⁶ 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.

- 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 2010.

Loyers mensuels réels ou fictifs en maison individuelle

Cf. rapport intermédiaire 2010.

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 fixe⁷. Les habitants des DOM-TOM représentaient alors 2,8 % des personnes vivant en ménage ordinaire.

Définition du ménage privé

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

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

Appartenance au ménage

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

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

L'enquête est conduite en mai-juin 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 2007 s'est déroulée du 14 mai au 30 juin 2007.

La collecte SILC 2008 s'est déroulée du 10 mai au 25 juin 2008.

La collecte SILC 2009 s'est déroulée du 9 mai au 30 juin 2009.

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

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.

⁷ 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*.

Cette partie du questionnaire est parfois longue (et laborieuse) à renseigner, en particulier dans le cas de personnes intérimaire ou occupant des emplois saisonniers.

La catégorie socioprofessionnelle des actifs ou des anciens actifs est codée automatiquement après la collecte par un programme proche du système de codification automatique SICORE développé à l'INSEE. 85 % des professions sont codées automatiquement, les autres professions étant « reprises » à la main par une équipe spécialisée à l'Insee.

Le codage en ISCO se fait à partir du code des professions et catégories socioprofessionnelles (PCS) à quatre chiffres, de la nomenclature d'activité française en 712 postes et du nombre de salariés de 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 sont obligatoires à partir de 2008. Il n'est pas possible d'isoler des salaires récupérés par appariement avec les fichiers administratifs l'avantage en nature que représente la voiture de fonction. La variable PY021 n'est donc pas renseignée.

Revenus de la propriété (HY040N et HY090N)

Pas de classification particulière.

Allocations famille/enfants (HY050)

Elles incluent l'ensemble des allocations versées par les Caisses d'allocations familiales hormis les allocations logement et le RMI puis RSA. 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 puis RSA et les aides financières sociales versées par les collectivités locales. Ne disposant pas des sources de calage extérieures sur les aides locales, ce poste est sans doute sous-estimé par rapport à la réalité.

Aides au logement (HY070)

Pas de classification particulière.

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

Ces transferts excluent les versements exceptionnels mais incluent le paiement de loyer par un tiers. 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. Depuis 2008, la majorité des salaires sont récupérées par appariement, ils comprennent les avantages en nature déclarés aux impôts et les allocations de préretraite imposables (ceux-ci ne peuvent pas être distingués des salaires). Mais nous disposons aussi de salaires non imposables disponibles dans l'enquête mais non présents dans les fichiers fiscaux.

Cotisations sociales des employeurs (PY030G)

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

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

Nous avons privilégié les prélèvements privés que la personne a effectués sur les ressources de son activité d'indépendant pour ses besoins de consommation ou d'épargne. Le revenu fiscal est utilisé en cas de non-disponibilité de la première information.

Valeur des biens produits pour la consommation propre (PY070N)

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

Allocations de chômage (PY090N)

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

Allocations de vieillesse (PY100N)

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

Pension de réversion (PY110N)

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

Retraites non obligatoires (PY080N)

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

Indemnités de maladie (PY120N)

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

Pensions d'invalidité (PY130N)

Les allocations suivantes sont comptées dans cette composante du revenu :

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

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

- l'aide personnalisée à l'autonomie (APA) qui est versée aux personnes âgées dépendantes qui ont recours à une aide à domicile,

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

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

Allocations d'études (PY140N)

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

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

Depuis la collecte 2008, les revenus sont récupérés par appariement avec les fichiers fiscaux et sociaux principalement et collectés par enquête pour ceux non imposables ou pour les personnes pour lesquelles nous savons que l'appariement ne sera pas possible (cf. la description de la procédure utilisée pour recueillir les revenus, partie 2.3.2). Jusqu'en 2007, les revenus étaient collectés par enquête.

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

Les variables de revenu sont déclarées nettes des cotisations sociales. 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)

Depuis 2008, les impôts sont récupérés par appariement. Jusqu'en 2007, ils étaient collectés. 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 le salaire déclaré à l'administration fiscale.

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 (familiales, logement, minima sociaux) sont également récupérées par appariement.

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)

	EU-SILC Longitudinal (année de revenus)				EU-SILC Transversal (année de rvenus)				ERFS			
Année de revenus	2006	2007	2008	2009	2006	2007	2008	2009	2006	2007	2008	2009
Salaires (milliards d'€)	451	503	512	523	444	505	527	533	510	527	553	554
Bénéficiaires (millions)	25.7	28.7	27.8	27.8	25.6	28.4	27.8	28.6	26.6	26.8	27.1	26.9
Moyenne	17590	17513	18425	18830	17365	17793	18957	18649	19179	19680	20412	20558
Minimum	10	3	10	4	10	3	10	4	1	1	1	1
D10	3800	2506	3330	2456	3760	2664	3408	2481	3257	3660	3927	3450
D20	7608	6574	7557	6512	7738	7007	7904	6525	7774	8260	8846	8291
D30	11308	10437	11670	10736	11366	11137	12160	10862	12327	12676	13210	12909
D40	13850	13618	14472	14469	13744	14048	14923	14463	14854	15372	15908	15935
D50	15857	15681	16563	16951	15743	15993	17000	16919	16965	17532	18179	18275
D60	17883	17837	18744	19326	17780	18217	19306	19196	19234	19849	20453	20697
D70	20495	20692	21598	22370	20313	21009	22190	22063	22012	22653	23365	23762
D80	24117	24400	25330	26554	23699	24628	26063	26107	26082	26630	27600	28030
D90	30954	31643	32572	34334	30397	31968	33321	33058	33758	34161	35465	36330
Maximum	159038	544194	360554	559214	1590389	594622	360554	1199980	765766	2282103	875204	831288

4.1.2 Les allocations chômage (PY090N)

	EU-SILC Longitudinal				EU-SILC Transversal				ERFS			
Année de revenus	2006	2007	2008	2009	2006	2007	2008	2009	2006	2007	2008	2009
Chômage (milliards d'€)	24.6	25.4	24.4	27.9	25.2	25.9	24.2	28.7	25.2	24.3	23.1	27.6
Bénéficiaires	3.7	4.7	4.4	4.8	3.8	4.6	4.4	5.0	4.5	4.2	4.0	4.5

(millions)												
Moyenne	6590	5408	5570	5853	6545	5586	5535	5739	6051	5760	5432	6142
Minimum	28	1	22	3	28	1	22	3	8	1	1	1
D10	775	644	604	656	749	651	583	630	610	634	651	748
D20	1478	1180	1165	1367	1561	1191	1170	1358	1326	1337	1363	1645
D30	2500	1927	1927	2098	2447	1922	2020	2162	2144	2243	2110	2599
D40	3600	2802	2705	3102	3560	2806	2833	3169	3062	3112	3062	3550
D50	4766	4060	3816	4398	4726	3790	3970	4325	4078	4145	4136	4676
D60	5853	5225	5012	5565	5640	5094	5191	5437	5346	5328	5282	5626
D70	7700	6614	6252	7365	7378	6157	6382	7038	6536	6455	6502	7282
D80	9782	8511	8426	9248	9722	8367	8475	9094	8845	8736	8671	9643
D90	13100	11144	11228	11744	13093	11144	11206	11555	11749	12178	11857	12320
Maximum	106936	69231	240864	105980	157154	433555	240864	105980	74702	113039	330013	114124

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

Cf. rapport intermédiaire 2010.

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

Cf. rapport intermédiaire 2010.

4.1.4 *Revenus sociaux (HY050G, HY060G et HY070G).*

Cf. rapport intermédiaire 2010.

4.1.4.1 *Prestations familiales (HY050N).*

Cf. rapport intermédiaire 2010.

4.1.4.2 *Aides au logement (HY070N)*

Cf. rapport intermédiaire 2010.

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

Cf. rapport intermédiaire 2010.

4.1.6 *Loyers*

Cf. rapport intermédiaire 2010.

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

Cf. rapport intermédiaire 2010.

4.1.8 *Revenu disponible du ménage et niveau de vie des individus.*
Cf. rapport intermédiaire 2010.

Annexe méthodologique

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

On peut justifier une telle approche lorsque l'on dit que l'on souhaite effectuer une prédiction de l'estimateur de sondage non calculable. Fournir cette prédiction est optimale au sens que nous obtenons in fine l'estimateur le moins risqué⁸. 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'intérêt⁹. L'estimateur de sondage d'une médiane, d'un quantile ou d'un indicateur d'inégalité sont des quantités non linéaires en les variables d'intérêt. Le choix d'une imputation aléatoire, autrement dit par simulation, a donc été retenu. En effet, produire l'estimateur qui correspondrait à la moyenne d'estimateurs de sondage calculés sur plusieurs jeux indépendants de données simulées approcherait la prédiction de l'estimateur de sondage. Afin de satisfaire la demande d'Eurostat nous avons uniquement procédé à une imputation aléatoire simple. Lorsque le taux de valeurs simulées est très inférieur à celui des valeurs renseignées et conservées, la différence est négligeable. La différence est également d'autant plus faible que nous utilisons des informations auxiliaires par conditionnement (informations en fourchettes et jusqu'à un certain point l'information fournie par des covariables).

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

Nous avons modélisé les montants comme issus de tirages dans une loi log-normale dont la moyenne du logarithme est linéaire en des covariables et avec un facteur d'échelle constant. Les montants sont en général observés soit en clair soit en tranches. Une estimation est réalisée sur le groupe des répondants et nous faisons implicitement l'hypothèse, qu'au sein des deux groupes (répondants/non répondants), la loi (conditionnelle à l'observation des covariables) du montant est la même. Il est possible d'imaginer la présence de biais de sélection, mais nous avons pris le parti de les négliger. Les biais de sélection sont d'autant plus faibles que nous avons utilisé le conditionnement par des observations de covariables.

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

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

Si aucune information n'est disponible sur la valeur du montant manquant, la simulation est faite dans la loi log-normale. Lorsqu'une plage de valeurs est disponible, ou reconstituée à partir de la connaissance d'autres variables, la simulation se fait dans la loi précédente conditionnelle au fait que

⁸ Pour le risque quadratique calculé avec l'aléa de modèle.

⁹ Dans certains cas seulement la différence peut être négligeable.

la variable est dans la plage de valeurs prescrite. Dans ce second cas la simulation revient à la simulation de lois normales tronquées. La simulation dans la loi normale tronquée est effectuée par inversion de la fonction de répartition dès que cela est possible. Il est arrivé parfois que des problèmes numériques (plage loin de la valeur prédite par le modèle...) nous fassent utiliser une méthode d'acceptation/rejet avec optimisation de la probabilité d'acceptation. Cette méthode est une variante de la méthode de la macro « résidus simulés » mais est beaucoup plus efficace¹⁰.

¹⁰ 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.