

L'attrition dans SRCV : Déterminants et effets de l'attrition sur la mesure des variables

Carine Burricand (), Jean-Paul Lorgnet (*)*

() Insee, Direction des Statistiques Démographiques et Sociales*

Introduction

L'enquête SRCV (statistiques sur les revenus et les conditions de vie des ménages) est la partie française de l'enquête européenne *EU-SILC* (Statistics on Income and Living Conditions). Le système statistique EU-SILC porte sur la mesure du revenu et des conditions de vie des personnes vivant en ménage ordinaire. Il a pour vocation de permettre la production d'indicateurs structurels sur la répartition des revenus, de la pauvreté et de l'exclusion comparables pour les pays de l'Union Européenne. Il s'inscrit dans le programme d'action communautaire de lutte contre l'exclusion sociale et fournit le matériau statistique au rapport de synthèse annuel de la Commission Européenne sur ces questions. Le règlement européen établit un cadre commun pour la production systématique de statistiques communautaires sur le revenu et les conditions de vie, englobant des données transversales et longitudinales.

Pour remplir ces objectifs, sous les recommandations d'Eurostat, les pays ont mis en place un panel rotatif sur quatre années. Bien que l'enquête relève d'un règlement européen et sous la tutelle d'Eurostat, elle est menée de façon indépendante par chaque états membres de l'union européenne qui peuvent adapter le plan de sondage proposé par Eurostat afin de répondre à des besoins nationaux. La Norvège et la France ont ainsi mis en place un panel d'une plus longue durée : huit et neuf années respectivement, afin d'étudier sur une plus grande période les mécanismes d'entrées et sorties de pauvreté. Si l'utilisation d'un panel comporte de nombreux avantages (réduction des coûts en réinterrogation, suivi dans le temps des phénomènes), un des risques résulte dans l'attrition, cause de possibles biais et de réduction de la taille d'échantillonnage, même si la mise en place d'un panel rotatif permet de limiter ces effets au moins en transversal en faisant entrer de nouveaux individus dans l'échantillon.

L'objectif de cette étude est de mesurer et d'analyser les déterminants de l'attrition dans l'enquête SRCV et de vérifier la qualité de la correction de l'attrition par la pondération.

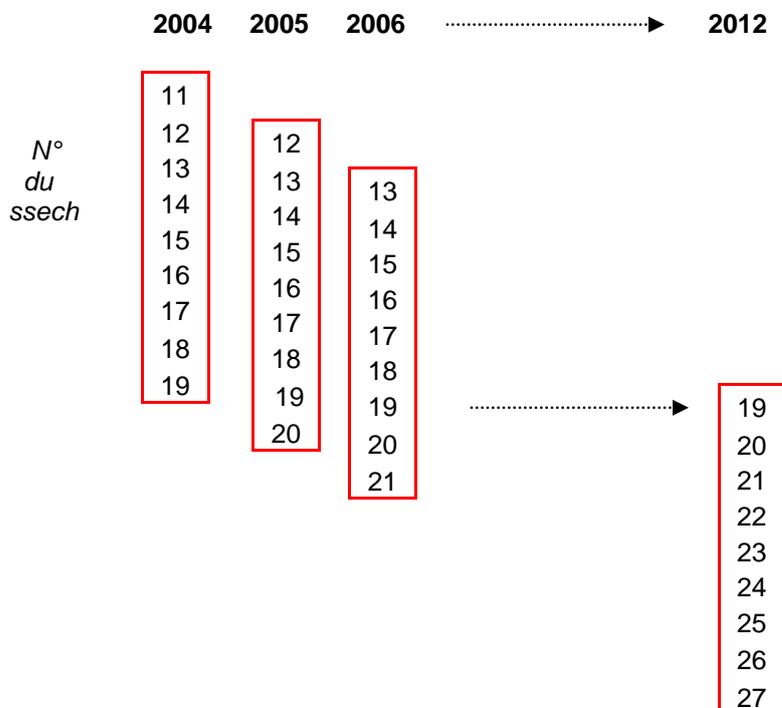
1. Taux de réponse dans le panel SRCV

1.1. Le panel SRCV

L'échantillon de l'enquête SRCV est constitué d'un panel rotatif sur neuf années ; le panel est donc constitué par la réunion de neufs sous-échantillons panels, tirés dans des conditions identiques, pour partie dans l'échantillon maître du RP1999, pour partie dans la base de sondage des logements neufs (BSLN) et à partir de 2010 dans la base de sondage Octopusse (« organisation coordonnée de tirages optimisés pour une utilisation statistique des échantillons) tirée dans les enquêtes de recensement.

L'initialisation, en 2004, a conduit à un échantillon de 16 000 logements, divisé en neuf sous-échantillons de taille égale. Chaque année, un sous-échantillon sort et un nouvel sous-échantillon entre, composé à partir de 2005 de 3000 logements. L'un des sous-échantillons constitué en 2004 a été interrogé une seule fois (en 2004), un autre deux fois (2004 et 2005), un autre trois fois (2004,

2005 et 2006), etc. En régime de croisière, un panel donné sera interrogé neuf années de suite. Durant la phase d'initialisation, qui s'achèvera en 2012 avec la sortie du neuvième et dernier sous-échantillon issu du tirage de 2004, les sous-échantillons seront de fait sollicités à moins de neuf reprises. A partir de 2012, chaque sous-échantillon sera interrogé pour une vague d'interrogation différente, de la première à la neuvième.



Chaque panel entrant dans SILC réunit tous les individus résidant dans l'ensemble des logements tirés et appartenant à un ménage répondant. L'interrogation de l'ensemble des individus des ménages appartenant aux logements sélectionnés permet de produire des estimations à la fois au niveau des individus et des ménages de la population. Il n'y a aucune sur-représentation de catégories d'individus.

Le protocole de collecte permet de considérer chaque sous-échantillon comme un panel d'individus. Les individus sont définis comme « *individus panels* » s'ils sont présents en 1^{ère} vague dans un ménage répondant. Les interviews sont aussi réalisés auprès des personnes qui emménagent avec un individu panel (ces personnes sont appelées *cohabitants*), et tant que ces nouveaux membres vivent avec celui-ci ou un autre individu panel. Par définition européenne, les enfants nés d'une mère panel sont considérés comme panel : par construction française, c'est le cas aussi des enfants adoptés, la distinction entre ces derniers et les enfants « naturels » n'étant pas possible. Ces enfants ne sont suivis que le nombre d'années restant de leur mère dans le panel ; ils ne sont donc pas considérés comme des entrants de 1^{ère} année.

On cherchera à interroger les individus panels le temps de la durée du panel, ce qui implique de les suivre même en cas de déménagement : les différentes directions régionales de l'INSEE se transmettent pour cela en cours de collecte les dossiers des individus du panel qui déménagent. En cas de déménagement, un individu non panel ne sera pas recherché s'il ne vit plus avec un individu panel. Néanmoins, leur présence dans l'échantillon une année donnée participe à une meilleure représentativité transversale : cela permet en effet de tenir compte d'une partie des immigrants arrivés depuis le début du tirage de l'échantillon initial. En terme de représentativité longitudinale par contre

et, du fait des règles de suivi, seuls les individus panels nous intéressent. On limitera donc le champ d'étude aux individus panels¹.

1.2. Taux de réponse « transversaux »

Dans le cas d'une enquête par panel, un des risques pesant sur la qualité de l'échantillon porte sur la perte de représentativité au fil du temps résultant de l'attrition si il y a sélectivité et de la diminution des effectifs diminuant la précision des indicateurs. Ce n'est donc pas la non réponse en 1^{ère} vague qui nous intéresse ici mais la non réponse en réinterrogation : cette dernière peut par ailleurs être analysée plus finement qu'en première vague car nous disposons de davantage d'informations sur les individus du fait des précédents interviews.

La non-réponse peut résulter de raisons multiples résultant du comportement de l'enquêté (refus, sorties du champ de l'enquête²) ou/et de l'échec de l'institution dans le suivi des ménages (adresse non retrouvée en cas de déménagement, impossible à joindre).

Cas de non réponse	
Sorties du champ	Décès / départ en collectivités /départ à l'étranger
Non localisable	Adresse non retrouvée en cas de déménagement
Non joignable	Impossible à joindre, absent de longue durée
Non joint	FA non traitée avant la fin de l'enquête
Refus	Refus du ménage de répondre à une nième vague et les suivantes

Ces différents processus de non réponse amènent un traitement différent dans le suivi. En effet, dans le cas d'un refus de réponse au niveau du ménage, le ménage n'est plus échantillonné l'année suivante et on perd l'ensemble des individus panels du ménage définitivement. Par contre dans le cas d'un ménage impossible à joindre ou absent pour longue durée (impossibilité de localiser l'individu), le ménage fera toujours parti de l'échantillon l'année suivante. Au bout de deux années de non réponse, le ménage est retiré de l'échantillon.

Le taux de réponse constitue une mesure clé de la qualité des données. Dans les enquêtes ménages, deux types de taux de réponse transversaux sont calculés : le taux de réponse brut défini comme le rapport du nombre de répondants sur le nombre de cas échantillonnés dans l'enquête et le taux spécifique de réussite (TSR), ne tenant pas compte des fiches adresses hors champ (logement détruit, résidence secondaire, collectivités ...). On s'intéressera ici aux taux de réponse brut³ calculé pour chaque vague d'interrogation N et mesuré par rapport aux individus appartenant aux ménages à enquêter l'année N, soit les ménages répondants de l'année N-1 et les ménages n'ayant pu être contacté (IAJ,ALD, FA non traitée avant la fin de la période de collecte) l'année N-1. En effet, la disparition de l'individu du panel ne se réalise que lorsque le ménage auquel il appartient n'est plus joignable, ne répond plus etc...mais cela ne signifie en rien que l'individu lui-même ait répondu à l'enquête. En effet, le protocole de collecte autorise le recours au proxy dans l'enquête : ainsi autour de 28 % des questionnaires individuels (QI) des personnes âgées de 16 ans ou plus chaque année sont renseignés par un autre membre du ménage que la personne elle-même. Par ailleurs, un individu peut refuser que son QI soit renseigné (c'est le cas de moins de 1 % des individus âgés de 16 ans ou plus) mais dans les deux cas ces individus peuvent appartenir à un ménage répondant (défini comme un ménage ayant renseigné un questionnaire ménage et au moins un QI) et être enquêté de nouveau l'année suivante.

¹ Les enfants définis comme panels parce-qu'ils sont nés de mère « panel » après l'année d'entrée de leur mère dans l'échantillon ne seront pas suivis.

² Le champ de l'enquête portant sur les seuls ménages ordinaires résidant en métropole, toute sortie du champ de l'enquête contribue à l'érosion de l'échantillon.

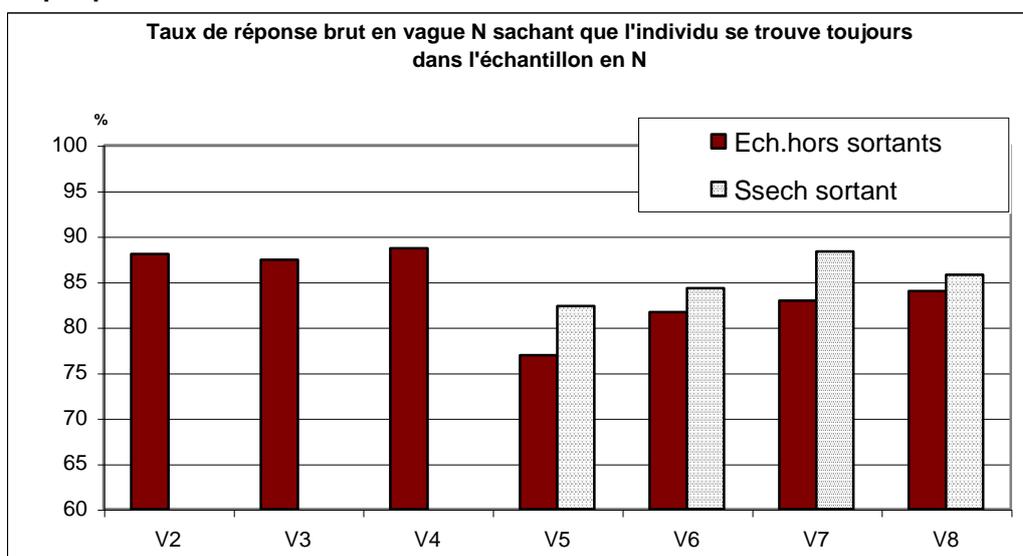
³ Les sorties hors du champ de l'enquête participant à l'érosion du panel, ils nous intéressent ici d'en tenir compte.

Parmi les individus panels définis comme tels en première vague, 88 % en moyenne appartiennent à un ménage répondant la seconde année⁴. Environ 4 % refusent et 5 % n'ont pas pu être contactés (impossibles à joindre, absents de longue durée). Ces derniers ne sont pas nécessairement perdus car remis dans l'échantillon l'année suivante. Seuls 1 % sortent du champ des ménages ordinaires ou décèdent et pour le reste, leur adresse n'a pas pu être retrouvé en cas de déménagement.

Le taux de réponse en troisième et quatrième interrogation parmi les personnes de nouveau présentes dans l'échantillon l'année suivante s'élève à 88 % également : ce taux relativement élevé pour une enquête ménage peut s'expliquer par l'adhésion obtenue lors du premier entretien à la participation à l'enquête et le caractère obligatoire de l'enquête sur les quatre premières années.

A partir de la cinquième année, l'enquête n'étant plus obligatoire⁵, le taux de réponse se contracte de 10 points (Cf. graphique 1). Les taux remontent ensuite progressivement à chaque vague ; une hypothèse probable est que les restants sont aussi les plus « intéressés », « motivés » ou ayant un plus grand sens civique.

Graphique 1 :



Source : INSEE-SRCV 2004-2011

Mode de calcul : il s'agit d'une moyenne arithmétique des taux observés par ssech à chaque vague d'interrogation identique.

On n'observe pas de fortes différences entre les sous-échantillons, exceptés pour les échantillons sortants en interrogation non obligatoire dont les taux de réponse sont systématiquement plus élevés, jusqu'à trois à cinq points au-dessus pour la 5^{ème} à la 7^{ème} vague (ceci apparaît moins marqué pour la 8^{ème} vague observée pour la première fois en 2011 avec + 1 point seulement). C'est l'effet « positif » de la dernière année, effet souligné par les enquêteurs lors des réunions de bilan : il leur est plus facile d'argumenter et d'obtenir une réponse positive de participation à l'enquête la dernière année de présence dans l'échantillon.

Les taux de réponse portent ici sur l'ensemble des individus panels, y compris sur les enfants âgés de moins de 16 ans qui ne remplissent pas de questionnaire individuel et sont donc « passifs » dans le cadre de l'enquête : les taux de réponse ou de présence encore observés dans le panel pour les enfants âgés de moins de 16 ans sont cependant identiques à ceux observés pour les personnes âgées de 16 ans ou plus.

⁴ Ces résultats ont été établis à partir de l'observation des données SRCV 2005 à 2011.

⁵ Du fait du principe de la « collecte loyale », la non obligation de réponse à partir de la cinquième année est rappelée aux enquêtés dans la lettre avis envoyée avant le passage de l'enquêteur.

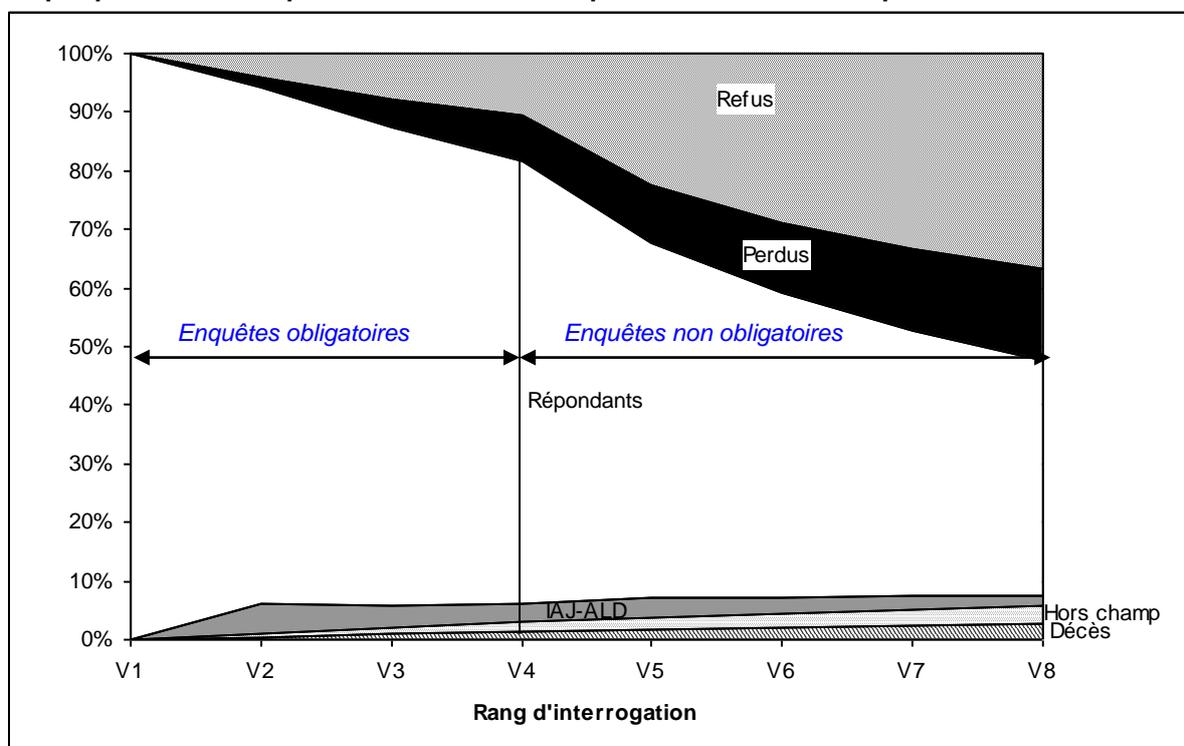
1.3. Taux de réponse « longitudinal »

Pour mesurer l'érosion de l'échantillon initial, il est possible de calculer un taux de réponse cumulé ou longitudinal établi sur le champ des individus panels. Le taux de réponse longitudinal est un rapport qui exprime le nombre d'individus panels de chacune des vagues successives en proportion du nombre de personnes ayant participé à la première interview.

Si l'on cumule les différents taux de réponse (en excluant les échantillons interrogés pour la dernière fois avant d'avoir atteint la durée complète du panel et qui ne représentent pas la réalité en régime permanent du panel), au bout de huit années d'interrogation, 4 individus panels sur 10 font encore partie d'un ménage ayant répondu. 37 % des individus ont disparu du panel suite à un refus et 15 % parce qu'on n'a pas réussi à retrouver leur nouvelle adresse. Enfin, les sorties du panel par hors champ ou décès s'élèvent à 6 % au bout de huit années.

On ne tient pas compte dans ces taux de réponse, de retour d'individus étant sortis du panel les années précédentes. Ainsi, parmi les individus « perdus », c'est à dire partis en ménage ordinaire à une adresse inconnue, à l'étranger ou en collectivité, 1 % réapparaissent l'année suivante dans le ménage⁶.

Graphique 2 : Taux de présence des individus panels sur la durée du panel



Source : INSEE-SRCV 2004-2011

Mode de calcul : il s'agit d'une moyenne arithmétique des taux cumulatifs observés par ssech à chaque vague d'interrogation identique.

Conformément aux règles de traçabilité, les ménages impossibles à joindre ou absent pour longue durée sont rééchantillonnés l'année suivante (s'ils étaient répondant l'année précédente): les taux de réponse varient selon les années d'interrogation mais entre 35 et 45 % sont répondants l'année suivante. La difficulté de les joindre est un signe précurseur de l'absence de réponse l'année suivante. Les taux de réponse pour les années au-delà de N+1 sont également systématiquement plus bas que pour les personnes ayant toujours répondu : ainsi 75 % des personnes impossibles à

⁶ Dans les traitements aval, une recherche est effectuée sur les individus entrants dans le ménage à partir des dates de naissances et du prénom pour vérifier s'il n'y avait pas quelqu'un ayant les mêmes caractéristiques les années précédentes. Si tel est le cas, on vérifie ensuite la correspondance sur d'autres variables personnelles afin de confirmer qu'il s'agit bien du même individu et lui réaffecter son identifiant individuel.

joindre en 2^{ème} vague et répondant en 3^{ème} vague sont répondants en 4^{ème} vague contre 91 % pour les répondants aux trois premières vagues.

2. Les facteurs de la non réponse

Les graphiques précédents permettent de prendre la mesure de l'effet de l'attrition sur la taille de l'échantillon des individus panels au fur et à mesure de l'augmentation du nombre de réinterrogation. La précision des indicateurs peut en être affectée si cette attrition est sélective et mal prise en compte dans le calcul de pondération. Pour étudier le caractère sélectif de l'échantillon, on cherche à analyser les effets des caractéristiques individuelles sur la probabilité de non réponse à l'aide d'un simple modèle binaire (logistique) sur données empilées et qui permet de déterminer les facteurs principaux de la non-réponse.

La non-réponse est modélisée par une variable y_{it} qui vaut 1 lorsque l'individu i ne fait pas partie d'un ménage répondant à la date t et 0 sinon. L'avantage du panel est de pouvoir intégrer dans l'étude de la non réponse, des caractéristiques observables en $t-1$.

Ici, la variable y_{it} (observable) dépend d'une variable latente y_{it}^* qui dépend linéairement :

- des variables individuelles ou du contexte terrain X_{it} observées à l'instant t , tels que le fait d'avoir déménagé⁷, le rang d'interrogation, un changement d'enquêteur ;
- des variables individuelles ou du ménage X_{it-1} observées à l'instant $t-1$, tels que la composition du ménage, la situation d'emploi, l'état de santé...
- de variables constantes dans le temps comme le genre ;
- d'un aléa non observé ε_{it} indépendant de X_{it} , X_{it-1} et de X_i .

$$y_{it}^* = X_{it}\beta + X_{it-1}\alpha + X_i\gamma + \varepsilon_{it}$$

Les aléas sont supposés indépendants entre individus et pour un même individu au cours de périodes différentes.

On ne s'intéresse qu'aux individus qui restent dans le champ de l'enquête en $t+1$ car leur « sortie du panel » est indépendante de l'action des enquêteurs ou de l'institut. On ne modélise par ailleurs la non réponse que pour les individus âgés de 16 ans ou plus dans la mesure où les moins de 16 ans ne sont pas « actifs » dans le cadre de l'enquête puisqu'ils ne renseignent pas de questionnaire individuel : leur disparition dans le panel dépend donc du comportement de réponse de leurs parents.

Les résultats obtenus mettent en évidence l'influence déterminante dans la non réponse du déménagement de l'individu panel (que ce soit avec l'ensemble des membres du ménage ou un départ du ménage de l'individu) et du passage au caractère non obligatoire de l'enquête (cf. *tableau 1*). Avoir connu une situation de pauvreté (monétaire ou en conditions de vie) l'année précédente favorise la non réponse, ainsi qu'être inactif ou en mauvais état de santé. Ce sont des résultats traditionnels que l'on retrouve dans la littérature sur l'analyse de la non réponse (Cf. Uhlig (2008) pour une revue de la littérature concernant les variables explicatives de la non réponse).

Chercher à suivre les individus en cas de déménagement est donc essentiel dans le cadre d'un panel. En effet, le biais potentiel lié à l'exploitation d'un sous-échantillon de personnes n'ayant jamais déménagé est d'autant plus susceptible d'être important que les populations étudiées sont mobiles ou dans des situations précaires (Breuil & Valdelièvre, 2001). Dans l'enquête SRCV, chaque année, près de 8 % des individus panels appartiennent à des ménages ayant déménagé entre deux vagues et seulement entre 65 et 70 % des individus ayant déménagé une année t appartiennent à un ménage répondant l'année $t+1$.

En cas de déménagement de l'ensemble des membres du ménage et malgré les moyens mis en place pour fidéliser les enquêtés et retrouver les adresses, près de 20 % des adresses ne sont pas

⁷ Si les individus n'ont pu être interrogés, l'enquêteur sur le terrain s'est renseigné et transmet l'information sur les raisons de la non interrogation (individus ayant déménagé, impossible à joindre...).

retrouvées. Par-contre, les taux de réponse des personnes ayant déménagé et pour lesquelles on a une adresse sont élevés et proches de ceux des ménages n'ayant pas déménagé.

En cas de départ d'individus panels du ménage d'appartenance de l'année précédente (séparation, divorce, départ d'enfants du domicile parental...), dans 95 % des cas une adresse est retrouvée mais la non réponse s'explique par une proportion d'individus n'ayant pu être contacté quatre fois plus importante que dans le cadre des ménages n'ayant pas déménagé (autour de 20 % contre 5 %).

A défaut de disposer d'informations sur les caractéristiques individuelles et familiales à la date t sur les non répondants, on peut penser que l'information sur le déménagement et les ménages éclatés sont des variables qui permettent de capter les populations mobiles dont le mouvement peut être le résultat d'une modification de la vie familiale ou professionnelle. C'est donc une variable qui est incluse dans les modèles de redressement de la non réponse par vague pour les calculs de pondération mais l'hypothèse implicite derrière est que les individus répondants ayant déménagé ont les mêmes caractéristiques que les individus non répondants ayant déménagé, ce que l'on ne peut vérifier.

Un autre élément impactant les taux de réponse est le changement d'enquêteur d'une vague à l'autre. L'opinion courante dans la conduite d'enquête est de recommander de conserver les mêmes enquêteurs pour un même enquêté d'une vague à l'autre pour permettre de meilleurs taux de réponse. En effet, on peut supposer que dans le cadre d'un panel surtout sur longue période, une relation de confiance s'installe entre l'enquêteur et l'enquêté et que le changement d'enquêteur rompt cette relation qui s'était créée ce qui conduit à la sortie anticipée de l'individu du panel. Le travail conduit par Laurie & co. (1999) sur l'échantillon du panel britannique de Ménage (BHPS) des vagues 2 à 4 montre l'impact positif du maintien de l'enquêteur sur les taux de réponse. Les travaux de Hill and Willis (2001) montre également à partir de l'enquête Share que conserver le même enquêteur permet d'augmenter le taux de réponse de six points entre la vague 1 et 2. Dans notre modèle, on peut constater que le changement d'enquêteur influe positivement sur la non réponse, que l'individu ait déménagé ou non.

Chaque année, dans l'enquête SRCV, le changement d'enquêteur va concerner entre 10 et 15 % des individus panels selon les années et entre 11 et 14 % pour les seuls individus ne déménageant pas : en fait le déménagement n'implique pas nécessairement un changement d'enquêteur car dans la majorité des cas il s'agit d'un déménagement intra ou intercommunale. Dans le cas des déménagements et que l'adresse est retrouvée, 66 % des individus panels ne changent pas d'enquêteur selon SRCV. En effet, la mobilité est souvent à proximité : dans l'enquête logement de 2006, la mobilité infra-départementale concernait 70 % des ménages (37 % en cas de mobilité infra-communale) et la moitié des ménages des ménages déménageait dans un rayon de 5 kilomètres (Minodier, 2006). Dans le cas de l'enquête SRCV, il est cependant vraisemblable qu'une adresse non retrouvée soit plus fréquemment en dehors de la zone d'enquêteur.

En se limitant aux individus n'ayant pas déménagé, le taux de réponse est systématiquement plus bas quand il y a eu un changement d'enquêteur : il est de cinq points plus bas pour les interrogations obligatoires et atteint 10 points d'écart en cas de réinterrogations non obligatoires. Quand on regarde sur les individus ayant déménagé et pour lesquels une adresse a été retrouvée, on retrouve cette même constatation au moins sur les premières années, ensuite les effectifs plus réduits ne permettent pas de l'assurer avec certitude.

Si les variables explicatives les plus significatives sont intégrées dans le redressement de la non réponse en réinterrogation dans le calcul de la pondération⁸, il est évident que nous ne disposons pas de toutes les variables pouvant avoir une influence sur le comportement de réponse car elles ne sont pas observées dans l'analyse. La question qui se pose est de savoir si la correction est suffisante pour limiter le biais de précision dû à l'attrition.

⁸ Plusieurs étapes sont mises en place dans le calcul de la pondération qui se résument en le redressement de la non réponse initiale, le redressement de la pondération en réinterrogation, le partage des poids et le calage sur marge.

Tableau 1 : Paramètres estimés d'un logit dichotomique

	Coefficients et significativités
Constante	- 2,27***
Genre	
Homme	Réf.
Femme	- 0,03
Age	
<30 ans	0,34***
De 30 à moins de 45 ans	Réf.
De 45 à moins de 60 ans	- 0,04
De 45 à moins de 75 ans	- 0,05
75 ans et plus	0,19***
Vague d'interrogation	
Obligatoire	Réf.
1 ^{ère} année non obligatoire (5 ^{ème} interrogation)	0,88***
Autres années non obligatoires (>5 ^{ème} interrogation)	0,36***
Type de ménage	
Personne seule	0,10**
Famille monoparentale	0,14**
Couple sans enfant	0,05*
Couple avec au moins un enfant	Réf.
Autre type de ménage	0,07
Situation d'emploi de l'individu	
en emploi	Réf.
Chômeur	0,06
Inactif	-0,11***
Type d'habitat	
Maisons individuelles	-0,19***
Immeubles	Réf.
Situation de pauvreté	
Pauvreté monétaire et pauvreté en CDV	0,43***
Pauvreté monétaire ou pauvreté en CDV	0,28***
Aucune des 2 formes de pauvreté	Réf.
Etat de sante déclaratif	
Bon ou très bon	-0,09***
Assez bon	Réf.
Mauvais ou très mauvais	0,23***
Mouvement de l'individu	
Déménagement de l'ensemble des membres du ménage et changement d'enquêteur	0,40***
Déménagement de l'ensemble des membres du ménage sans changement d'enquêteur	- 0,34***
Départ du ménage	1,31***
Vit toujours à la même adresse et changement d'enquêteur	0,51***
Vit toujours à la même adresse sans changement d'enquêteur	Réf.

Lecture : *** : significatif au seuil de 1% ; ** : significatif au seuil de 5% ; * significatif au seuil de 10% ; Réf.

Catégorie de référence.

Champ : individus âgés de 16 ans ou plus résidant en ménage ordinaire en métropole

Source : Enquête SRCV 2004-2010

Calcul sur données empilées

*** : significatif à 1% ; ** significatif à 5% ; * : significatif à 10% ; écarts-types entre parenthèses.

3. Les effets de l'attrition dans SRCV

Afin d'étudier les effets de l'attrition sur SRCV, l'approche est de regarder l'impact de l'attrition sur les moyennes de revenu disponible et de ses grandes composantes. Le revenu disponible d'un ménage comprend les revenus d'activité, les revenus du patrimoine, les transferts en provenance d'autres ménages et les prestations sociales (y compris les pensions de retraite et les indemnités de chômage), nets des impôts directs. Nous sélectionnons pour ce travail les catégories de revenus les plus perçues par les ménages : Il en va ainsi des salaires et retraites pour les revenus primaires, des aides au logement pour les transferts, enfin de l'impôt sur le revenu et de la taxe d'habitation pour les impôts. Les revenus portent sur l'année précédant la collecte (nommée année de référence) et dans la mesure où nous traitons de variables de revenus, nous parlerons ici par simplicité d'année de revenu et non de collecte. L'impôt retenu dans le calcul du revenu disponible est l'impôt payé l'année de référence des revenus et relatif aux revenus de l'année précédente : ainsi pour le calcul du revenu disponible de l'année 2003, l'impôt retenu est celui payé en 2003 sur les revenus perçus en 2002.

Cette méthode se décline en deux étapes (Graf E., 2010), la première consiste à présenter les moyennes du revenu disponible du ménage des individus panels et de ces grandes composantes en occultant le jeu de pondérations longitudinales, la seconde l'utilise.

Nous associons aux moyennes du revenu disponible et ses grandes composantes les écarts types correspondants afin de vérifier si ces derniers calculés pour les revenus 2003 et sur les individus toujours présents dans le panel restent dans l'intervalle de confiance de celui obtenu sur la population initiale. Les écarts-types sont estimés à l'aide du Programme Optimal et Universel pour la Livraison de la Précision des Enquêtes, Poulpe. Ce logiciel a été créé à l'Insee et permet d'évaluer la précision de statistiques issues d'enquêtes par sondage complexes, en particulier les enquêtes auprès des ménages ou des entreprises, réalisées par l'Insee (Caron, 2006).

En 2004, il n'y a pas d'attrition par définition, puisqu'il s'agit de la première vague du panel. Le graphique 3 présente les moyennes du revenu disponible des ménages pour l'année 2003 calculées sur les individus panels interrogés en 2004 (abscisse année de revenu 2003) à celles calculées pour les individus encore présents en 2005 (abscisse 2004), puis 2006 (abscisse 2005) et ce jusqu'en 2009. Sont calculés les moyennes non pondérées (en bleu) et les moyennes pondérées (en rose).

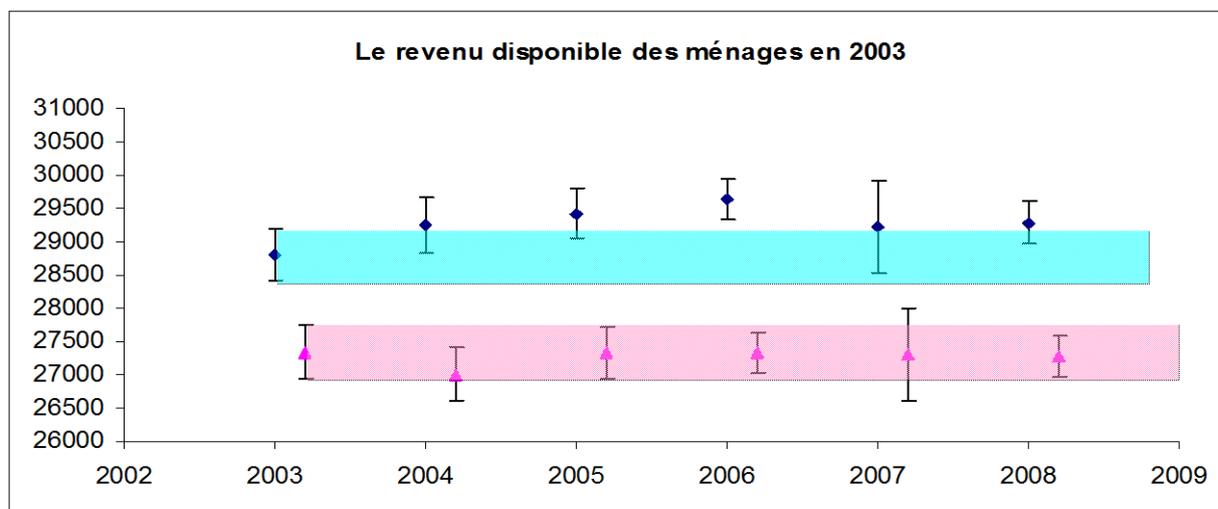
Dans un premier temps, on peut observer l'effet de l'attrition sur le revenu disponible (calculé pour l'année 2003) pour les différentes populations présentes chaque année en vérifiant si les intervalles de confiance calculés chaque année se recoupent avec celui obtenu sur les revenus de l'année 2003 (bande de couleur bleue dans le tableau).

Jusqu'à l'année de revenu 2006, le revenu disponible calculé sans pondération des individus présents aux différentes dates croît, ce qui semblerait indiquer que l'attrition concerne davantage les individus membres de ménages disposant de plus faibles revenus disponibles. Néanmoins, l'attrition ne semble pas avoir une influence significative au plan statistique sur la représentativité du panel puisque les intervalles de confiance se recoupent, hormis pour l'année de revenu 2006 (l'intervalle de confiance calculé pour cette année est totalement extérieure à la bande bleue, soit celui obtenu pour l'année 2003).

Dans un second temps, sont calculés les moyennes et intervalles de confiance à l'aide des pondérations longitudinales (Ardilly, Lavallé, 2007) calculées selon les années de présence⁹ que l'on compare à la moyenne et l'intervalle de confiance des revenus disponibles des ménages des individus panels présents en 2004. Le graphique montre qu'aucun point n'est en dehors de l'intervalle de confiance ce qui permet de penser que la pondération longitudinale corrige correctement l'attrition pour le revenu disponible.

⁹ Des pondérations longitudinales sont calculées pour chaque individu panel présent entre la date t et la date t+n (n=1 à 8 en fin de panel).

Graphique 3 : Moyenne et intervalle de confiance du revenu disponible de l'année 2003 des ménages des individus panels présents en 2004 et suivant leur présence dans le panel.



Note de lecture : La moyenne du revenu disponible sur l'année de revenu 2003 des ménages des individus panels s'élève en pondéré à 27 345 euros annuels. Elle est de 27 012 euros annuels en moyenne en 2003 si on ne retient que les individus panels encore présents l'année suivante (année de revenus 2004).

Source : Enquête SRCV 2004-2009 (revenus 2003 à 2008)

De la même manière, nous regardons l'impact de l'attrition sur les composantes du revenu disponible. La moyenne des salaires hors pondération est bien répartie et l'attrition ne semble pas affecter la moyenne des salaires observés sur les individus restants dans le panel ; les salariés sont pourtant en effectif autant affectés par l'attrition que l'ensemble des individus mais les caractéristiques des non répondants est peut-être peu différente de celles des répondants. Par contre la pondération a pour effet de modifier un peu cette répartition au bout de trois années de présence dans le panel (année de revenu 2005) sans toutefois qu'aucun point ne sorte totalement de l'intervalle de confiance initial.

Graphique 4 : Moyenne et intervalle de confiance des salaires de l'année 2003 des individus panels présents et salariés en 2004 et suivant leur présence dans le panel.

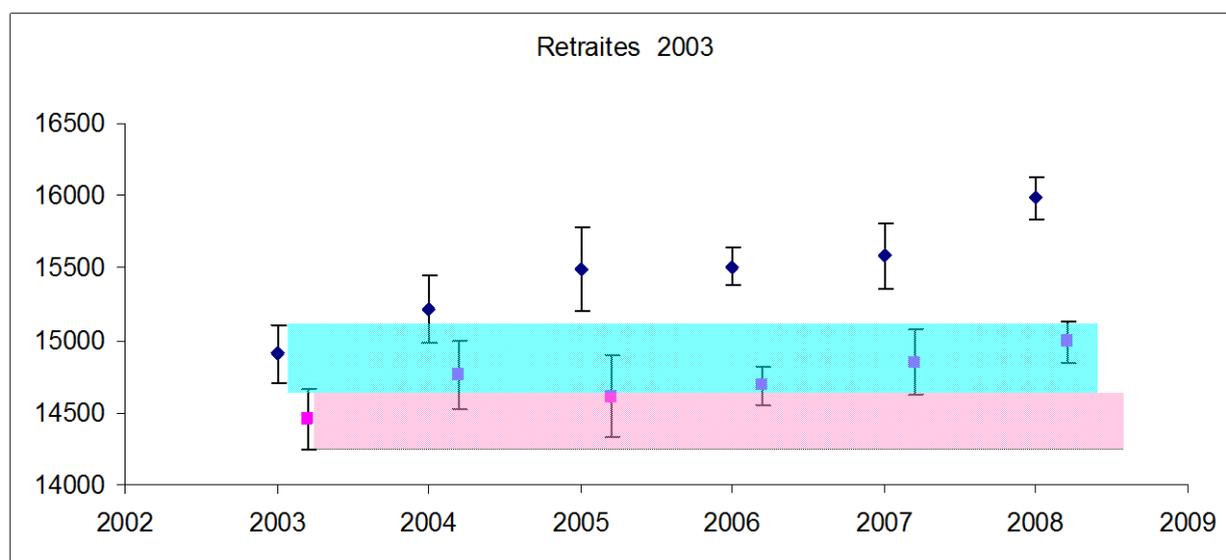


Note de lecture : La moyenne des salaires sur l'année de revenu 2003 des ménages des individus panels s'élève en pondéré à 17 605 euros annuels. Elle est de 17 536 euros annuels en moyenne en 2003 si on ne retient que les individus panels encore présents l'année suivante (année de revenus 2004).

Source : Enquête SRCV 2004-2009 (revenus 2003 à 2008)

L'effet de l'attrition est par-contre marqué sur les retraites. En effet, hors pondération, dès la troisième année de présence dans le panel (année de revenu 2005), les points sortent de l'intervalle de confiance initial et les montants moyens de retraites annuelles calculées pour les individus panels retraités s'élèvent au fur et à mesure de l'augmentation de la durée du panel. La pondération ne change pas le sens de la courbe mais permet de corriger pour partie cet effet puisqu'un seul point est « out », celui calculé sur les individus panels présents encore la sixième année d'interrogation (année de revenus 2008).

Graphique 5 : Moyenne et intervalle de confiance des retraites de l'année 2003 des individus panels présents et retraités en 2004 et suivant leur présence dans le panel.



Note de lecture : Le montant annuel moyen de retraite sur l'année de revenu 2003 des ménages des individus panels retraités s'élève en pondéré à 14 455 euros annuels. Elle est de 14 756 euros annuels en moyenne en 2003 si on ne retient que les individus panels encore présents l'année suivante (année de revenus 2004).

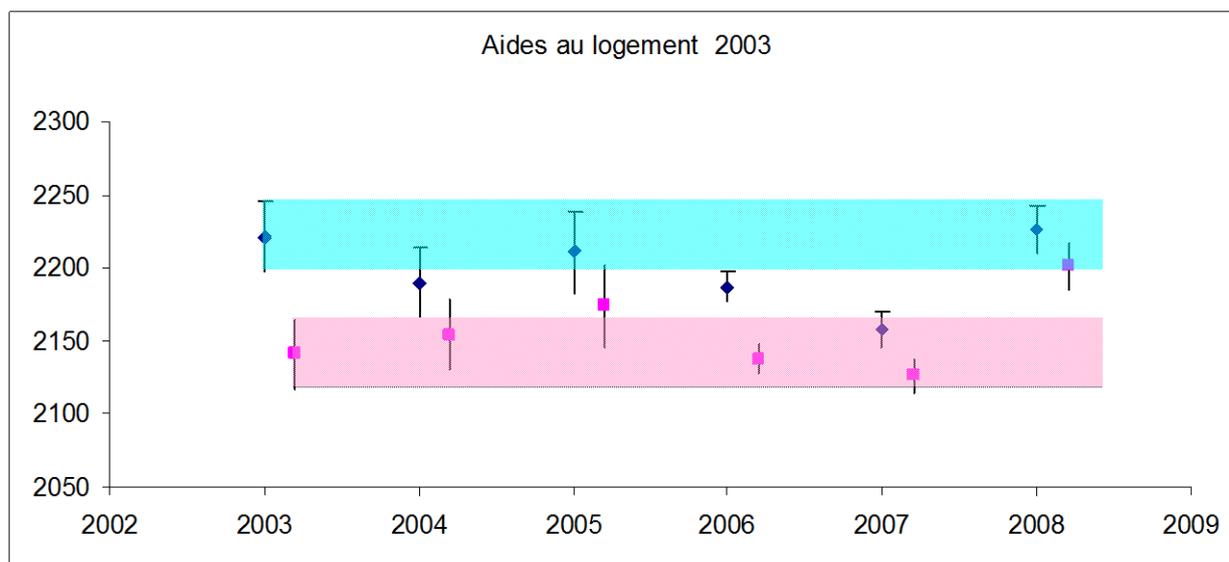
Source : Enquête SRCV 2004-2009 (revenus 2003 à 2008)

L'attrition est donc sélective sur les plus âgés, les individus restant dans le panel ayant des retraites en moyenne plus élevées que les sortants du panel. En effet, Les individus présents en 2004 et toujours présents en 2009 perçoivent des retraites d'un montant supérieur de 11,6 % aux retraites des individus absents en 2009. Ceci peut s'expliquer par les disparitions du panel par décès mais pas seulement : en effet la moyenne des retraites des individus panels décédés n'est inférieure que de de 1,3 point à celle de l'ensemble des absents en 2009. De fait, si on exclut de l'exercice les décédés, cela ne modifie pas la constatation de l'impact sur les retraites, la part des décédés parmi les retraités disparus sur la période du panel n'étant que de 10 %.

Du fait de l'importance des aides au logement dans le système redistributif (cela concerne en moyenne 7,2 millions de bénéficiaires pour 15 milliards d'euros), nous avons regardé également l'impact de l'attrition sur les aides au logement.

Sans pondération, quatre points sont dans l'intervalle de confiance, deux sont en dehors. La pondération corrige l'attrition excepté pour les aides au logement de 2008. Il est à noter que pour l'année 2008 la pondération a un effet néfaste, mais les effectifs deviennent faibles (326 allocataires d'aides au logement la première année d'enquête sont encore présents six ans après).

Graphique 6 : Moyenne et intervalle de confiance des aides au logement de l'année 2003 des individus panels présents et allocataires en 2004 et suivant leur présence dans le panel.

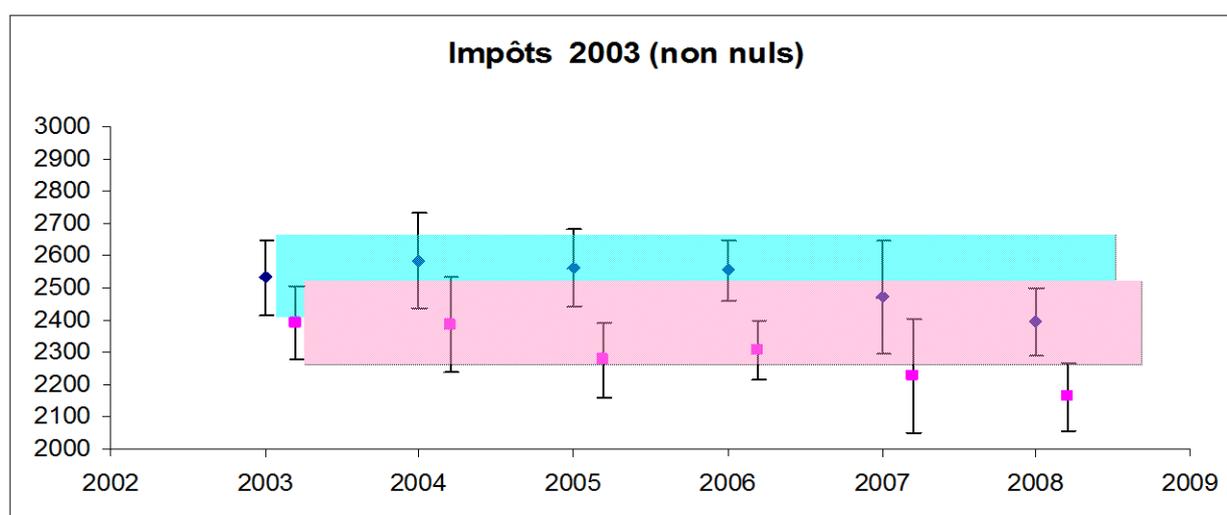


Note de lecture : Le montant annuel moyen des aides au logement sur l'année de revenu 2003 des ménages des individus panels allocataires s'élève en pondéré à 2 141 euros annuels. Elle est de 2 154 euros annuels en moyenne en 2003 si on ne retient que les individus panels encore présents l'année suivante (année de revenus 2004).

Source : Enquête SRCV 2004-2009 (revenus 2003 à 2008)

En ce qui concerne les impôts des ménages imposables (impôt sur le revenu, la taxe d'habitation ainsi que la prime pour l'emploi qui est retranchée), hors pondération on observe une légère diminution de l'impôt, qui reste toujours dans l'intervalle de confiance. La pondération à ce niveau ne change rien, les deux courbes ont des profils proches, les deux derniers points sont moins bons, l'impôt 2008 est même en dehors de l'intervalle de confiance.

Graphique 7 : Moyenne et intervalle de confiance des impôts de l'année 2003 des individus panels présents en 2004 et suivant leur présence dans le panel.



Note de lecture : Le montant annuel moyen des impôts sur l'année de revenu 2003 des ménages des individus panels s'élève en pondéré à 2 226 euros annuels. Elle est de 2 215 euros annuels en moyenne en 2003 si on ne retient que les individus panels encore présents l'année suivante (année de revenus 2004).

Source : Enquête SRCV 2004-2009 (revenus 2003 à 2008)

Le même exercice effectué sur l'année de revenu 2004 (cf .annexe 1) nous amène aux même conclusions, l'effet de l'attrition sur les retraites étant cependant moins marqué. En conclusion, si l'on écarte les retraites, l'attrition n'a pas d'effet majeur sur les revenus catégoriels et un effet nul sur le revenu disponible. Lorsqu'il y a un effet, les pondérations réussissent généralement à corriger l'attrition sur les revenus, avec semble t-il une moins grande efficacité sur une durée du panel plus grande.

Bibliographie

[1] Ardilly P., Lavallée P. « Pondération dans les échantillons rotatifs : le cas de l'enquête SILC en France », 2007.

[2] Breuil-Genier P. et Valdelièvre H., « Le Panel européen : l'intérêt d'un panel d'individus » ÉCONOMIE ET STATISTIQUE N° 349-350, 2001

[3] Caron N., "Le logiciel poulpe : aspects méthodologiques", INSEE-Methodes n°84-85-86, 2006

[4] Ekaterina Kalugina et Boris Najman, « Travail et pauvreté en Russie : évaluations objectives et perceptions subjectives », ÉCONOMIE ET STATISTIQUE N° 367, 2004

[5] Graf E., "Etude empirique de l'attrition du panel suisse de ménages, Vers une caractérisation du profil du non -répondant », Série Statistique de la suisse n°338-0059, 2010.

[6] Hill, Daniel, and Robert Willis. 2001. Reducing Panel Attrition: A Search for Effective Policy Instruments. *The Journal of Human Resources* 36 (3):416-38.

[7] INSEE, « L'économétrie et l'étude des comportements : Présentation et mise en oeuvre de modèles de régression qualitatifs - Les modèles univariés à résidus logistiques ou normaux (LOGIT, PROBIT) », document de travail de la DSDS, 2000.

[8] Laurie H., Smith R., Scott L.« Strategies for reducing nonresponse in a longitudinal panel survey », *Journal of Official Statistics* vol.15, No.2, 1999 pp269-282.

[9] Lipps O., "Attrition of households and individuals in panel surveys", SOEP papers 164, 2009

[10] Minodier c., « Changer de logement dans le même environnement », Données sociales - La société française Edition 2006.

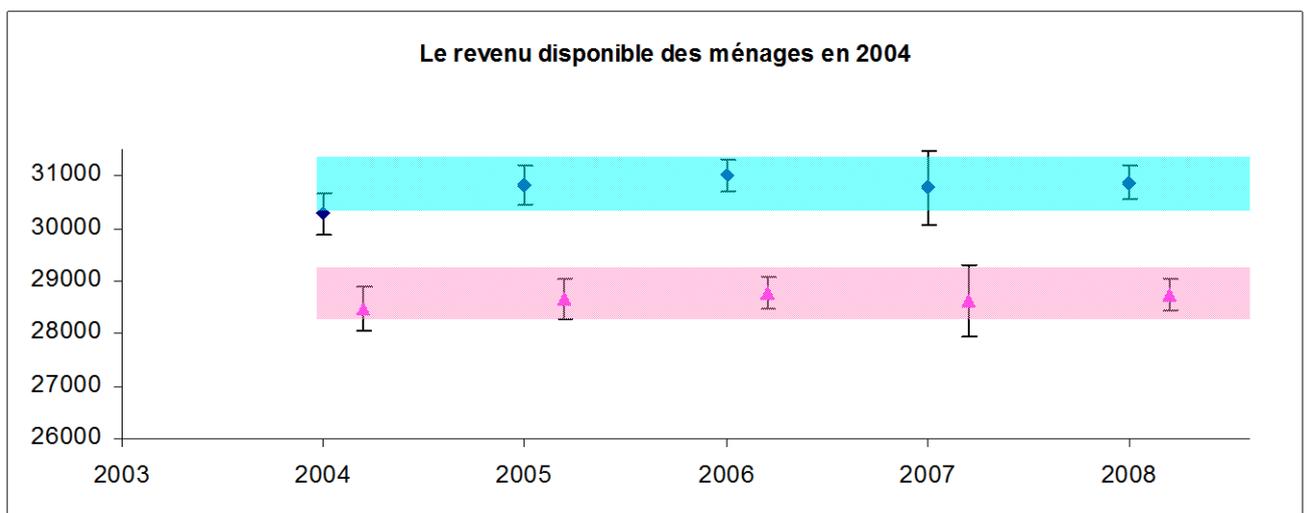
[11] Rendtel U., « Attrition in households Panels : a survey », Chintex Working paper 4, 2002

[12] SC Noah Uhrig, « The nature and causes of Attrition in the British Household Panel Survey" , ESRC Economic & Social research Council N°2008-05, february 2008.

Annexe 1 : Effets de l'attrition en choisissant une autre année de base.

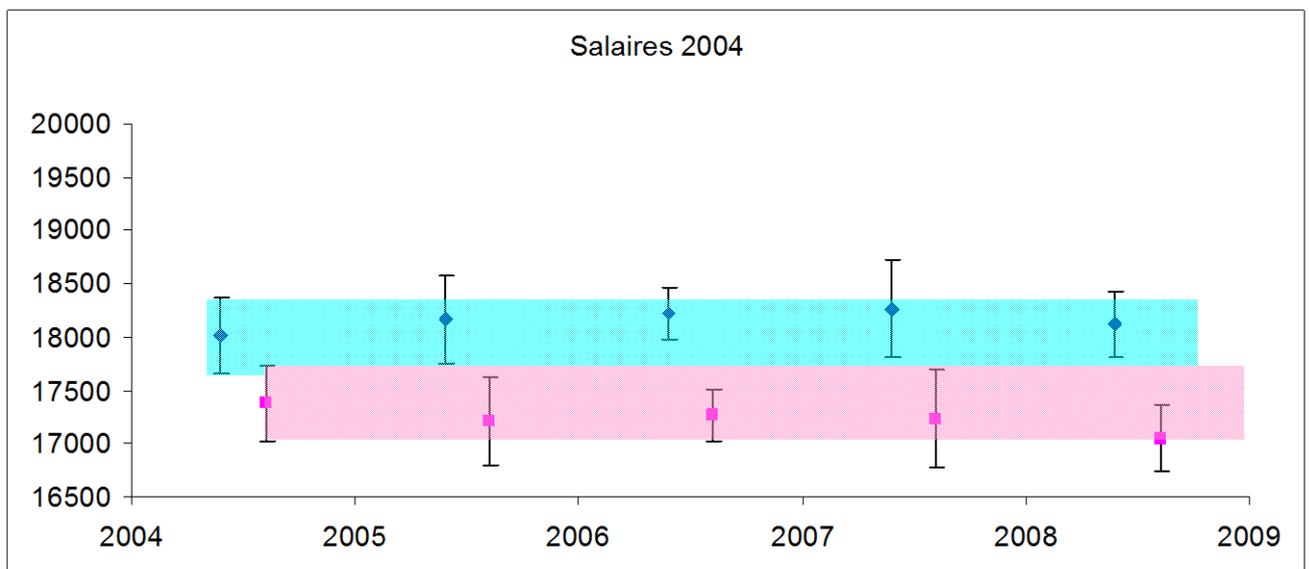
Nous avons voulu contrôler les résultats précédents en choisissant une autre année de base, pour ce faire celle ci sera l'année 2005. Pour ce travail nous retenons le revenu disponible et trois de ces composantes les salaires, les retraites et les impôts. Les graphiques présentent les moyennes du revenu disponible ou des composantes de revenu des individus pour l'année 2004 calculées sur les individus panels interrogés en 2005 (abscisse année de revenu 2004) à celles calculées pour les individus encore présents en 2006 (abscisse 2005), puis 2007 (abscisse 2006) et ce jusqu'en 2009. Sont calculés les moyennes non pondérées (en bleu) et les moyennes pondérées (en rose).

Concernant le revenu disponible en 2004, nous constatons que l'attrition perturbe un peu ce revenu disponible si nous n'utilisons pas de pondération. En revanche la pondération a un effet stabilisateur.



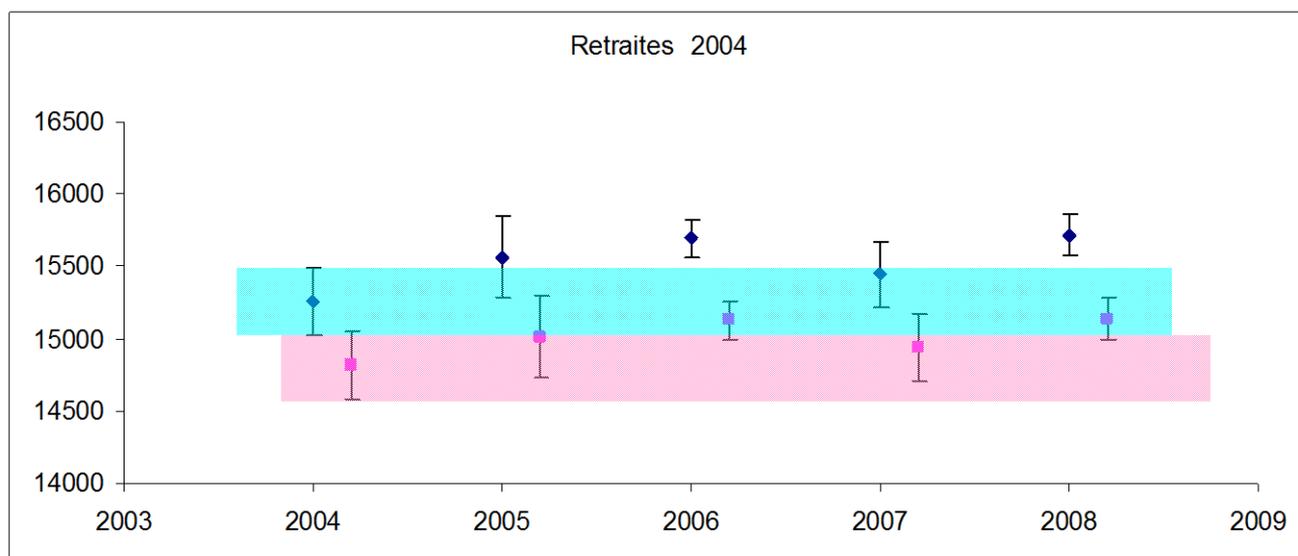
Source : Enquête SRCV 2005-2009 (revenus 2004 à 2008)

Les salaires moyens en 2004 sont tous dans le rectangle de significativité, avec ou sans pondération. Sans pondération les salaires 2004 se situent dans la moitié supérieure du rectangle ce qui n'est pas le cas des salaires 2003. La pondération a pour effet de baisser les salaires 2004 comme elle le faisait avec les salaires 2003, mais dans une moindre mesure.



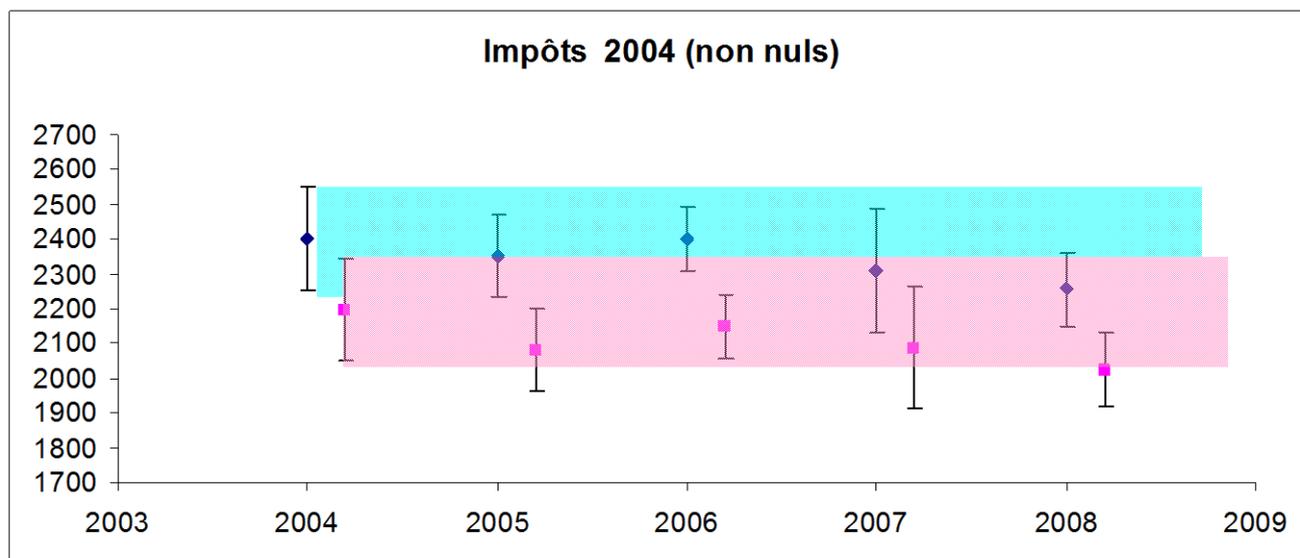
Source : Enquête SRCV 2005-2009 (revenus 2004 à 2008)

Nous constatons, à l'instar des retraites moyennes calculées sur l'année 2003, que les retraites en 2004 sont modifiées par l'attrition, toutefois l'effet est moins marqué en 2004. La pondération atténuée les écarts, après son application aucun point n'est en dehors du rectangle de significativité.



Source : Enquête SRCV 2005-2009 (revenus 2004 à 2008)

Le changement d'année de référence n'introduit pas de nouveauté. Avec ou sans pondération l'impôt diminue. Les individus concernés par l'attrition paient moins d'impôt sur le revenu, ils ont donc des revenus plus faibles.



Source : Enquête SRCV 2005-2009 (revenus 2004 à 2008)