

L'attrition dans l'enquête *SRCV* : déterminants et effets sur la mesure des variables monétaires

Carine Burricand * et Jean-Paul Lorgnet **

Cette étude vise à mesurer l'attrition et ses effets sur le revenu disponible et ses composantes dans l'enquête sur les revenus et les conditions de vie des ménages en France. Cette enquête est basée sur un échantillon rotatif dont les individus sont suivis durant neuf années. Six individus sur dix ne font déjà plus partie de l'échantillon au moment de la huitième interrogation.

La mobilité résidentielle est la principale cause de cette attrition. Avoir connu une situation de pauvreté l'année précédente favorise également la non-réponse, ainsi qu'être inactif ou en mauvais état de santé. La non-réponse s'accroît également lors d'un changement d'enquêteur.

Les effets de cette attrition sont corrigés par repondération. On teste l'efficacité de ces redressements en comparant les revenus de 2003 mesurés sur l'ensemble des répondants de 2004 et ceux qui sont mesurés sur les échantillons de plus en plus restreints des répondants aux vagues 2005 à 2011, après repondérations. Pour le revenu global, les écarts restent systématiquement inférieurs à 2 % environ, avec un léger pic en quatrième interrogation. Les biais après redressements sont également faibles pour la plupart des composantes du revenu : salaires, prestations familiales, allocations logement, allocations chômage. En revanche il subsiste un biais sur les retraites qui croît continuellement d'une réinterrogation à l'autre. Ce biais subsiste même après prise en compte de la mortalité différentielle par catégorie sociale et il ne semble pas non plus explicable par l'intensité des départs en institution.

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni *a fortiori* l'Insee.

* Insee.

** *Chaf.* Au moment de la rédaction initiale de l'article, l'auteur travaillait à l'Insee.

Les auteurs remercient Pascal Ardilly, Pierre Lamarche ainsi que les deux rapporteurs anonymes pour leurs remarques et conseils précieux.

L'enquête européenne *EU-Silc*¹ (*Statistics on Income and Living Conditions*) permet la production d'indicateurs structurels comparables entre les pays européens sur la répartition des revenus, la pauvreté et l'exclusion sociale des personnes vivant en ménage ordinaire. Elle s'inscrit dans le programme d'action communautaire de lutte contre l'exclusion sociale et sert de base au suivi d'un des cinq objectifs majeurs de la stratégie Europe 2020 de la Commission européenne, qui est la sortie de 20 millions de personnes de la pauvreté d'ici 2020².

La qualité de l'enquête représente un enjeu fort pour les États-membres et nécessite une vigilance des équipes en charge de la production à tous les niveaux de la chaîne de traitement. Dans cette recherche de la qualité, l'élément qui nous intéressera plus précisément ici concerne l'attrition et la correction de ses effets par repondération. En effet, pour répondre aux demandes d'indicateurs reposant sur des données transversales ou longitudinales, l'enquête est basée sur un panel rotatif. Or, l'un des risques majeurs concernant la qualité des données de panel résulte de l'attrition (Hill et Willis, 2001). Plus la durée du panel est importante, plus ce risque s'aggrave (Muffels et Ruud, 2000) car l'attrition s'accumule au cours des vagues, réduisant l'échantillon initial et affectant sa représentativité, lorsqu'elle est sélective. La mise en place d'un panel rotatif permet de limiter les effets pervers de cette attrition, au moins en transversal, en faisant entrer de nouveaux individus dans l'échantillon (Ardilly et Lavallée, 2007), mais elle n'y apporte qu'une réponse partielle.

Si la grande majorité des pays européens ont mis en place un panel rotatif sur quatre années,

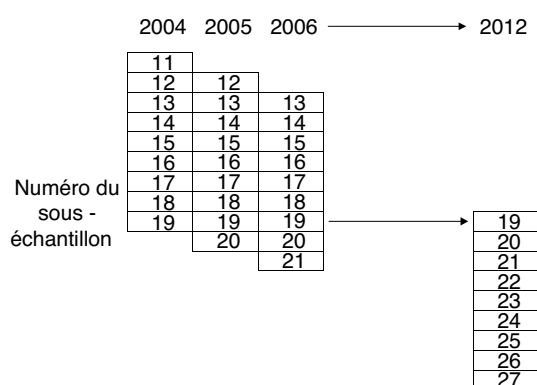
le règlement européen³ exigeant un suivi sur une période d'au moins quatre années, la France, comme la Norvège, a mis en place un panel d'une plus longue durée : neuf et huit années respectivement, afin d'étudier sur une plus grande période les mécanismes d'entrées et de sorties de la pauvreté. La correction de l'attrition est donc un enjeu particulièrement important dans le cas français. La qualité des données de panel dépendra en partie de l'ampleur de cette attrition, des caractéristiques observables et inobservables de celle-ci et de la possibilité *in fine* de pouvoir la corriger de façon satisfaisante par le calcul d'une pondération *ad hoc*. L'objectif de cette étude est ainsi de mesurer l'attrition observée dans l'enquête française, d'en analyser les déterminants et de vérifier la qualité de sa prise en compte par la pondération. En préambule, nous rappellerons les différents cas de non-réponse et leurs traitements dans le suivi de l'enquête.

La non-réponse : concepts et traitement

L'échantillon de l'enquête est constitué d'un panel rotatif sur neuf années, formé par la réunion de neuf sous-échantillons panels. Chaque année, un sous-échantillon sort et un nouveau sous-échantillon est introduit (cf. schéma). La phase d'initialisation s'est achevée en

1. La déclinaison française de l'enquête Silc se nomme SRCV (statistiques sur les revenus et les conditions de vie des ménages).
 2. http://ec.europa.eu/europe2020/index_fr.htm.
 3. RÈGLEMENT (CE) N° 1982/2003 DE LA COMMISSION du 21 octobre 2003 portant application du règlement (CE) n° 1177/2003 du Parlement européen et du Conseil relatif aux statistiques communautaires sur le revenu et les conditions de vie (EU-Silc), en ce qui concerne les modalités d'échantillonnage et les règles de suivi.

Schéma
 Rotation des sous-échantillons de l'enquête



Lecture : en 2004, l'échantillon est constitué de neuf sous-échantillons numérotés de 11 à 19. En 2005, le sous-échantillon 11 est sortant et n'est donc plus interrogé ; à l'inverse l'échantillon 20 est interrogé pour la première fois en 2005.

2012 avec la 9^e interrogation du dernier sous-échantillon encore issu du tirage de 2004.

En termes de représentativité longitudinale et du fait des règles de suivi, seuls les individus dits « panels » (cf. encadré 1) nous intéressent, c'est-à-dire les individus ayant répondu en première interrogation et suivis au moins pendant 2 années consécutives. On se limite à ces individus car notre travail ne porte pas sur la non-réponse en 1^{re} vague : il ne concerne que la non-réponse en réinterrogation responsable de l'attrition. 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 interviews précédentes.

Le succès des enquêtes dépend principalement de la coopération des enquêtés mais la

non-réponse peut résulter d'autres raisons (cf. tableau 1) : elles peuvent être relatives à l'enquête (sorties du champ de l'enquête⁴) ou/et résulter de l'échec de l'institution dans le suivi des ménages (adresse non retrouvée en cas de déménagement, impossible à joindre).

Ces différents processus de non-réponse amènent un traitement différent dans le suivi, conformément aux consignes fixées par Eurostat. 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. En effet, la disparition de l'individu du panel ne

4. 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.

Encadré 1

LE PANEL SRCV/SILC ET LE SUIVI DES INDIVIDUS

L'échantillon de l'enquête est constitué d'un panel rotatif sur neuf années ; le panel est donc constitué par la réunion de neuf 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 nouveau 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 sous-échantillon 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. À partir de 2012, chaque sous-échantillon sera interrogé pour une vague d'interrogation différente, de la première à la neuvième.

Chaque sous-échantillon 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 d'origine comme un panel d'individus. Chaque année, à partir de la deuxième année d'enquête soit 2005, on dispose ainsi d'un échantillon d'individus tous panélisés dont les huit neuvièmes ont déjà été interrogés au moins une fois les années passées. Les individus sont définis comme « *individus panels* » s'ils sont présents en 1^{re} vague dans un ménage répondant et s'ils ont été suivis au moins 2 années consécutives. Ainsi en 2005 (cf schéma), les individus des sous-échantillons numérotés 12 à 19 et ayant déjà été interrogés en 2004 sont individus panels. 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^{re} 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.

se réalise que lorsque le ménage auquel il appartient n'est plus joignable, ne répond plus, etc. En revanche, le fait que l'individu demeure dans l'enquête ne signifie en rien qu'il ait lui-même répondu à l'enquête. En effet, le protocole de collecte autorise le recours au *proxy* dans l'enquête : chaque année, environ 28 % des questionnaires individuels (QI) des personnes âgées de 16 ans ou plus 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. Dans le cas d'un ménage impossible à joindre ou absent pour longue durée (impossibilité de localiser l'individu), le ménage fera toujours partie de l'échantillon l'année suivante. C'est au bout de deux années de non-réponse que le ménage est retiré de l'échantillon.

Plusieurs mesures sont mises en œuvre pour limiter l'attrition sur le terrain. Ainsi, l'envoi de carte-vœux avec coupon-réponse en cas de changement d'adresse vise à rappeler aux enquêtés en janvier qu'ils seront de nouveau contactés au printemps dans le cadre de l'enquête. Par ailleurs, cela permet de recenser les personnes qui n'habitent plus à l'adresse indiquée et de récupérer les nouvelles adresses pour une partie d'entre elles. Fumagalli *et al.* (2010) montrent d'ailleurs à travers une série d'expérimentations visant à encourager les réponses dans les enquêtes longitudinales que le recours à une carte coupon-réponse en cas de changement d'adresse est une stratégie efficace. En cours de collecte, si l'enquêteur ne parvient pas à convaincre l'enquêté de continuer à répondre à l'enquête ou à le contacter, les Directions régionales de l'Insee prennent le relais et envoient des courriers-types aux enquêtés en cas de refus ou pour les personnes impossibles à joindre, pour les inciter à contacter l'enquêteur. Enfin, les enquêtés bénéficient d'un cadeau symbolique

qui vise à les encourager à répondre et à les fidéliser : l'efficacité des incitations financières n'a pas été démontrée dans les enquêtes menées par l'Insee mais Laurie *et al.* (2008), à travers une revue de la littérature sur ce sujet, indiquent que, de toute évidence, « les taux d'attrition seraient plus élevés en l'absence d'incitation », même s'ils reconnaissent que l'on ne dispose que de peu d'éléments sur l'impact à long-terme de ces incitations dans les enquêtes longitudinales.

Au final, parmi les individus panels, définis comme tels en première vague, 88 % en moyenne appartiennent à un ménage répondant la seconde année⁵. 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 retrouvée en cas de déménagement.

Des taux de réponse plutôt élevés mais variables selon le rang d'interrogation

Comme lors de la seconde interrogation, le taux de réponse en troisième et quatrième interrogation s'élève à 88 % : 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.

À partir de la cinquième année, l'enquête n'étant plus obligatoire⁶ en France, le taux de réponse se contracte de 10 points (cf. graphique I). Du fait du principe de la « collecte

5. Ces résultats ont été établis à partir de l'observation des données SRCV 2005 à 2012 parmi les échantillons qu'il est prévu de réinterroger l'année suivante.

6. L'obligation de réponse n'a été accordée par le Comité du Label que pour les quatre premières années d'interrogation du fait de l'obligation des pays Européens à réaliser un panel sur une durée d'au moins quatre années.

Tableau 1

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	Fiche adresse non traitée avant la fin de l'enquête
Refus	Refus du ménage de répondre à une n ^{ième} vague et les suivantes

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. Les taux remontent ensuite progressivement à chaque vague pour finalement atteindre près de 85 % en huitième interrogation ; une hypothèse probable est que les restants sont aussi les individus les plus intéressés par le thème de l'enquête, ou bien ceux qui font preuve de davantage de sens civique, ou encore les individus qui apprécient le contact social avec l'enquêteur.

On n'observe pas de fortes différences entre les sous-échantillons interrogés pour la même vague d'interrogation mais à des années différentes. 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.

40 % d'individus panels encore répondants au bout de huit années

Pour mesurer l'érosion de l'échantillon initial, il est possible de calculer un taux de réponse cumulatif 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 sous-échantillons interrogés pour la dernière fois avant d'avoir atteint la durée complète du panel, soit neuf années d'interrogation, et qui ne représentent pas la réalité en régime permanent du panel), lors de la huitième année d'interrogation, moins de quatre individus panels sur dix font encore partie d'un ménage ayant répondu (cf. graphique II). 37 % des individus ont disparu du panel suite à un refus et 17 % 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 années⁷.

On ne tient par ailleurs pas compte dans ces taux de réponse de retour d'individus 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 % réapparaît l'année suivante dans le ménage⁸.

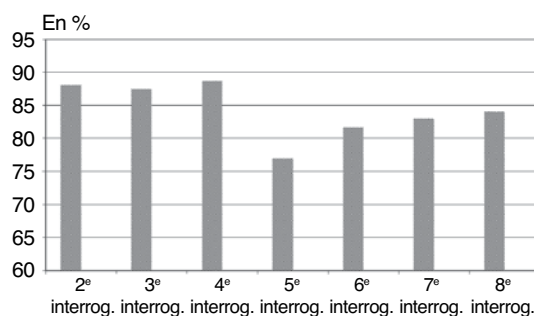
Conformément aux règles de traçabilité, les ménages impossibles à joindre ou absents pour longue durée l'année t sont rééchantillonnés l'année suivante s'ils étaient répondants en $t-1$: les taux de réponse varient selon les années d'interrogation mais entre 35 % et 45 % de ces

7. Ce taux ne porte que sur les sorties de ce type qui ont été portées à notre connaissance. Il est vraisemblable qu'il est en réalité plus élevé puisqu'une partie des personnes que l'on ne retrouve pas peuvent l'être du fait d'un décès ou d'un déménagement en dehors de la métropole.

8. Dans les traitements avals, une recherche est effectuée sur chaque individu entrant dans le ménage pour tester les éventuelles correspondances entre ses caractéristiques et celles des individus du même ménage interrogés les années précédentes. S'il y a correspondance, on réaffecte à l'individu entrant l'identifiant de l'individu déjà enquêté avec lequel il coïncide.

Graphique I

Taux de réponse brut en vague N sachant que l'individu se trouve toujours dans l'échantillon en N



Lecture : il s'agit d'une moyenne arithmétique des taux observés par sous-échantillon (hors échantillon sortant) à chaque vague d'interrogation identique (le calcul du taux moyen de réponse en 2^e vague porte sur les sous-échantillons suivants : 12 à 19 en 2005, 20 en 2006, 21 en 2007, 22 en 2008, 23 en 2009, 24 en 2010 et 25 en 2011 ; se reporter au schéma de rotation des sous-échantillons).

Champ : individus panels.

Source : enquête SRCV 2004-2011, Insee.

ménages sont répondants en $t+1$. 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 $t+1$ sont également systématiquement plus bas que pour les personnes ayant toujours répondu : ainsi, 75 % des personnes impossibles à joindre en 2^e vague et répondant en 3^e vague sont répondants en 4^e vague contre 91 % pour les répondants aux trois premières vagues.

La mobilité résidentielle est la principale cause de l'attrition

L'analyse globale de l'attrition montre donc qu'elle n'est pas négligeable à court-terme et devient importante à long-terme. La qualité des indicateurs issus de l'enquête sera affectée si cette attrition est sélective et mal prise en compte dans les calculs de pondérations. Pour étudier ce caractère sélectif, on cherche à analyser les effets des caractéristiques de l'individu et de son ménage sur la probabilité de non-réponse à l'aide d'un simple modèle binaire (logistique) sur données empilées (cf. encadré 2) qui permet de déterminer les facteurs principaux de la non-réponse.

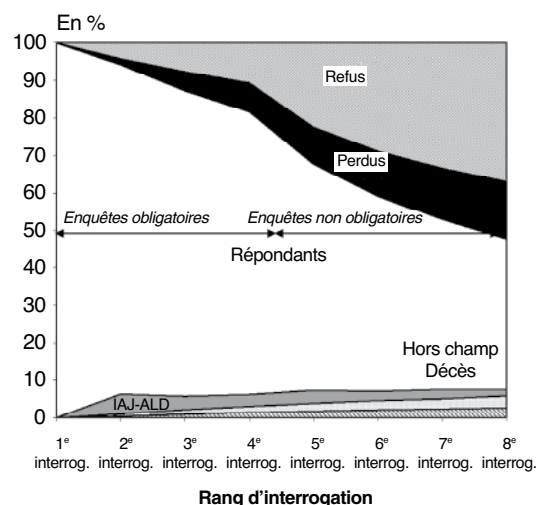
On ne s'intéresse qu'aux individus qui restent dans le champ de l'enquête en $t+1$ car leur sortie du panel est indépendante de l'action des enquêteurs ou de l'institut. On ne modélise par

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 du 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 2). Avoir connu une situation de pauvreté (monétaire ou en conditions de vie) l'année précédente favorise également 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 (pour une revue de la littérature concernant les variables explicatives de cette non-réponse, se reporter à Uhrig (2008)).

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 et Valdelièvre, 2001). Dans l'enquête SRCV, chaque année, près de 8 % des individus

Graphique II
Taux de présence des individus panels sur huit années de panel



Lecture : il s'agit d'une moyenne arithmétique des taux cumulatifs observés par sous-échantillon à chaque vague d'interrogation identique. IAJ-ALD : individu impossible à joindre ou absent de longue durée.
Champ : individus panels.
Source : enquête SRCV 2004-2011, Insee.

panels appartiennent à des ménages ayant déménagé entre deux vagues et seulement entre 65 et 70 % des individus ayant déménagé une année appartiennent à un ménage répondant l'année suivante.

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 retrouvées. En revanche, 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 (séparation, divorce, départ d'enfants du domicile parental...), une adresse est retrouvée dans 95 % de cas : la non-réponse s'explique plutôt par la forte proportion d'individus n'ayant pu être contactés. Elle est quatre fois plus importante que dans le cadre des ménages n'ayant pas déménagé (autour de 20 % contre 5 %).

À défaut de disposer d'informations sur les caractéristiques individuelles professionnelles 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 non-observable l'année précédente. 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 sous l'hypothèse implicite que les individus répondants ayant déménagé ont les mêmes caractéristiques que les individus non-répondants ayant déménagé.

Le changement d'enquêteur augmente la non-réponse

Un autre élément qui affecte les taux de réponse est le changement d'enquêteur d'une vague à l'autre. On s'efforce en général dans le cadre de cette enquête 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. Le changement d'enquêteur rompt cette relation et augmente le risque de sortie anticipée de l'individu du panel. Le travail conduit par Laurie *et al.* (1999) sur les vagues 2 et 4 du panel britannique de Ménages (BHPS) confirme cet impact positif du maintien de l'enquêteur sur les taux de réponse. Les travaux de Hill et Willis (2001) montrent également à partir de l'enquête *Share* que conserver le même enquêteur permet d'augmenter le taux de réponse de six points entre les vagues 1 et 2. Dans notre modèle, on peut également constater que le changement d'enquêteur influe positivement sur la non-réponse, que l'individu ait déménagé ou non.

Chaque année dans l'enquête, le changement d'enquêteur va concerner entre 10 et 15 % des

Encadré 2

LE MODÈLE BINAIRE LOGISTIQUE DE 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 de contexte terrain X_{it} observées à l'instant, telles que le fait d'avoir déménagé¹, le rang d'interrogation, un changement d'enquêteur ;

- des variables individuelles ou du ménage X_{it-1} observées à l'instant $t-1$, telles que la composition du ménage, la situation d'emploi, l'état de santé... ;

- des variables individuelles ou du contexte terrain X_{it} observées à l'instant t , telles que le fait de variables constantes dans le temps telles que le genre de l'individu ;

- d'un aléa non observé ε_{it} indépendant de X_{it} , X_{it-1} et de X_{it} .

On écrit donc :

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

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

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

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. Dans le cas des déménagements pour lesquels l'adresse est retrouvée, 66 % des individus

Tableau 2
Paramètres estimés d'un modèle *logit* dichotomique

	Coefficients et significativités
Constante	- 2,27***
Genre	
Homme	Réf.
Femme	- 0,03
Âge	
Moins de 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.
État de santé 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 : calcul sur données empilées *** : 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, Insee.

panels ne changent pas d'enquêteur. La raison est que la mobilité est souvent de 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 déménageait dans un rayon de 5 kilomètres (Minodier, 2006). Il est cependant vraisemblable qu'une adresse non-retrouvée soit plus fréquemment en dehors de la zone de l'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 l'écart atteint 10 points en cas de réinterrogations non obligatoires. En ce qui concerne les individus qui ont déménagé et pour lesquels une adresse a été retrouvée, ce constat se vérifie au moins sur les premières années. À plus long terme, les effectifs plus réduits ne permettent pas de l'assurer avec certitude.

Des redressements relativement efficaces pour le revenu global et la plupart de ses composantes ...

« Si le plan de sondage rotatif permet de répondre aux besoins longitudinaux et transversaux de l'enquête, il pose des défis en matière de pondération » (Ardilly et Lavallée, 2007). La difficulté essentielle consiste à représenter une population à une date t en transversal ou sur une période $[t ; t+n]$ en longitudinal à partir de sous-échantillons panels tirés à des dates différentes, donc dans des populations différentes. Les pondérations doivent produire des estimateurs approximativement sans biais et doivent corriger différents éléments comme les biais de sondage, la non-réponse initiale ou encore l'attrition étudiée ci-dessus. Plusieurs procédures sont ainsi mises en place pour traiter des différents biais. Elles peuvent se résumer en quatre étapes : le partage des poids⁹, le redressement de la non-réponse totale initiale puis en réinterrogation et enfin le calage sur marge¹⁰. Si des variables explicatives parmi les plus significatives sont intégrées pour le redressement de la non-réponse totale dans le calcul de la pondération, 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'enquête. Par ailleurs, la probabilité de réponse en réinterrogation est expliquée par les caractéristiques sociodémographiques de l'individu en

1^{re} vague (cf. encadré 3) et ne tient pas compte des changements entre les vagues, ni de la non-réponse entre les vagues intermédiaires. La question que l'on se pose ici est de savoir si la correction de la non-réponse en réinterrogation est suffisante pour limiter les biais dus à l'attrition.

Notre analyse des effets de l'attrition et de l'efficacité de sa correction va porter sur le revenu disponible et ses grandes composantes : salaires, retraites, allocations-chômage, prestations familiales et prestations-logement. Nous sélectionnons de cette manière les catégories de revenus les plus fréquemment perçues par les ménages, pour disposer d'effectifs suffisants, et dont la part dans le revenu disponible est non négligeable. Ces éléments composent en effet à eux seuls 90 % du revenu disponible. Les revenus portent sur l'année précédant la collecte (nommée année de référence) : ainsi pour les individus répondants en 2004, les revenus portaient sur l'année 2003.

Nous considérerons que l'échantillon répondant en 2004 joue le rôle de population complète : on s'affranchit donc de l'aléa de sondage. On estime chaque année à la date $[2004 + t]$ le revenu moyen de 2003 à partir des individus encore répondants aux différentes vagues en corrigeant uniquement de la non-réponse en réinterrogation. Puis on calcule l'intervalle de confiance associé à ce revenu moyen. Enfin, on estime le même revenu moyen de 2003 mais pour l'ensemble des individus interrogés en 2004 et encore vivants l'année $2004 + t$. En effet, en l'absence d'attrition, ce sont ces individus encore vivants l'année $2004 + t$ qui auraient dû répondre à notre enquête et c'est à leur revenu moyen de 2003 qu'il convient de comparer le revenu de 2003 mesuré sur les seuls répondants. Si ce revenu moyen reconstitué se trouve dans l'intervalle de confiance calculé pour les répondants, on conclura à l'efficacité de cette repondération (cf. encadré 4).

On constate que le revenu moyen estimé sur les seuls répondants apparaît systématiquement supérieur au revenu reconstitué pour l'ensemble

9. Ce poids dépend du nombre d'années durant lesquelles l'individu est échantillonnable. Dans la pondération transversale par exemple, du fait que les échantillons panels soient tirés à des dates différentes et ne représentent pas la même population, il faut les redresser pour tenir compte de l'évolution de la population dans le temps comme l'arrivée de nouveaux migrants ou les naissances.

10. Sur la population à la date t dans le cas de la pondération longitudinale sur la période $[t ; t+n]$ et sur la population à la date $t+n$ dans le cas de la pondération transversale à la date $t+n$.

des répondants de 2003 survivant à chaque date mais avec un intervalle de confiance qui inclut en général le revenu moyen estimé pour

l'ensemble des survivants (cf. graphique III). L'écart entre les deux moyennes est par ailleurs faible en termes relatifs.

Encadré 3

CALCUL DE LA PROBABILITÉ DE RÉPONSE EN RÉINTERROGATION

La probabilité de réponse en réinterrogation est un des éléments du calcul des pondérations transversale et longitudinale.

Une probabilité de réponse en réinterrogation est attribuée à chaque individu panel i qui est encore dans le champ en t . On attribue aux enfants la probabilité de réponse de leur mère ou à défaut de leur père : en effet, par définition la « non-réponse » n'existe pas pour eux car le questionnaire individuel n'est pas posé aux enfants : le fait de disposer de réponses aux questions sur les enfants dépend donc de la bonne volonté des parents.

Soit $p_i^{t/k}$, la probabilité de l'individu i de répondre à l'enquête à la date t , en k^e réinterrogation, sachant qu'il était répondant lors du premier passage, donc à la date $t-k+1$.

Cette probabilité est estimée pour tous les individus panel $i \in a_{t,k} \cap \Omega_t$ où Ω_t est la population complète du champ de l'enquête l'année t et $a_{t,k}$ le sous-échantillon panel à enquêter l'année t en k^e interrogation.

Pour les adultes en $t-k+1$, $p_i^{t/k}$ est estimée en utilisant les informations individuelles et ménages collectées en $t-k+1$ et le fait d'avoir déménagé ou d'avoir été dans un ménage éclaté en t .

Si $t \geq 2012$, il faut donc estimer 8 modèles de réponse en réinterrogation. Par exemple, en 2012, il faut un modèle pour les entrants en 2004 réinterrogés en 2012 pour la 9^e année, un pour les entrants en 2005 réinterrogés en 2012 pour la 8^e année et ainsi de suite jusqu'au dernier modèle pour les entrants en 2011 réinterrogés pour la 2^e fois en 2012.

À partir de ces résultats, différentes sous-populations de répondants sont formées en croisant les motifs significatifs de la non-réponse (certains groupes à faible effectif doivent être regroupés pour assurer la robustesse des résultats). Au final, les adultes réinterrogés en n^e vague sont répartis en plusieurs groupes et nous supposons le mécanisme de réponse homogène à l'intérieur des sous-populations. La probabilité $p_i^{t/k}$ pour chacun de ces sous-groupes vaut donc :

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

La théorie voudrait que, pour chaque sous-échantillon, soit calculée entre la date n et $n+t$ la probabilité de non-réponse entre n et $n+1$ puis entre n et $n+2$ successivement jusqu'à $n+t-1$ et $n+t$ mais la taille des sous-échantillons a conduit à utiliser cette approche simplifiée.

Encadré 4

CALCUL DU BIAIS D'ATTRITION

On note respectivement :

- Y_i^{2003} , le revenu 2003 de l'individu i calculé à partir des données de SRCV 2004 (les revenus d'une enquête collectée l'année N portent sur l'année $N-1$).
- $p_i^{t/2004}$, la probabilité de l'individu i de répondre à l'enquête à la date $2004 + t$ sachant qu'il était répondant à l'enquête de 2004, t étant la t^e réinterrogation.
- w_i , la pondération calculée en 2004.

On peut alors écrire :

$$\tilde{Y}_t^{2003} = \frac{1}{\left(\sum_{i \in r_{2004+t}} \frac{w_i}{p_i^{t/2004}} \right)} \times \sum_{i \in r_{2004+t}} w_i \cdot \frac{1}{p_i^{t/2004}} \cdot Y_i^{2003}$$

qui estime, à partir de l'échantillon répondant en $2004 + t$ (noté r_{2004+t}), le revenu total 2003 calculé sur la population des individus vivants en 2004 et en $2004 + t$. L'échantillon r_{2004+t} est bien constitué des individus répondant à *Silc* en 2004 et en $2004 + t$. Le poids w_i est le poids transversal associé à l'échantillon répondant de 2004 (en particulier il inclut la probabilité de réponse en 2004) : la pondération finale dans l'échantillon répondant à $2004 + t$ est donc bien le produit de w_i et de $1/p_i^{t/2004}$.

On calcule la variance de cet estimateur :

$$V(\tilde{Y}_t^{2003}) = \frac{1}{\left(\sum_{i \in r_{2004+t}} \frac{w_i}{p_i^{t/2004}} \right)^2} \times \sum_{i \in r_{2004+t}} (1 - p_i^{t/2004}) \times p_i^{t/2004} \times \left(w_i \cdot \frac{1}{p_i^{t/2004}} \cdot Y_i^{2003} \right)^2$$

→

Les deux seuls cas d'écarts significatifs correspondent aux années 2007 et 2011. En 2007

par exemple parmi les six sous-échantillons réinterrogés pour la 4^e année consécutive,

Encadré 4 (suite)

On en déduit l'intervalle de confiance associé à cette variance, sous l'hypothèse (raisonnable) d'une distribution Gaussienne :

$$IC = \left[\bar{Y}_t^{2003} \pm 1,96 \sqrt{V(\bar{Y}_t^{2003})} \right]$$

l'aléa pris en compte étant seulement l'aléa de non-réponse entre t et $2004 + t$ (calcul de variance conditionnelle à l'échantillonnage).

Soit alors \bar{Y}_{2003} le vrai revenu moyen 2003 défini sur la population vivant en 2004 et en $2004 + t$. Cette moyenne (inconnue) est estimée à partir de l'échantillon des individus répondants en 2004 mais encore vivants en $2004 + t$ (on exclut donc les individus décédés entre 2004 et $2004 + t$)

$$\bar{Y}_{2003} = \frac{1}{\sum_{i \in r_{2004/2004+t}} w_i} \cdot \left(\sum_{i \in r_{2004/2004+t}} w_i \cdot Y_i \right)$$

où $r_{2004/2004+t}$ désigne l'échantillon des répondants en 2004 encore vivants en $2004 + t$ (r_{2004+t} est inclus dans $r_{2004/2004+t}$).

Néanmoins, du fait du panel rotatif et de la sortie d'un sous-échantillon par année, on ne sait pas

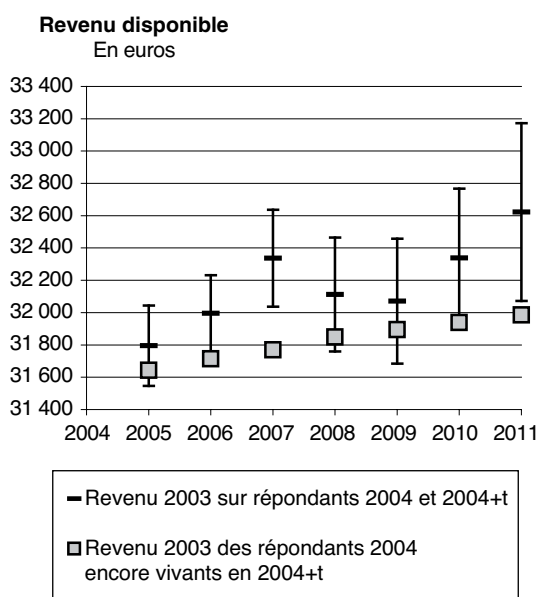
systématiquement quels sont les individus interrogés en 2004 qui sont encore vivants à la date $2004 + t$. On ne le sait que pour les échantillons encore suivis en $2004 + t$ et seulement en partie puisqu'on ne sait pas ce qui est arrivé aux individus sortis du champ. La prise en compte de la mortalité se fait donc plutôt en pondérant les individus interrogés en 2004 par leurs probabilités de survie par sexe et âge entre 2004 et $2004 + t$, telles que calculées par l'Insee.

En résumé :

- l'échantillon r_{2004} pondéré par les w_i permet une inférence sur la population 2004 ;
- l'échantillon $r_{2004/2004+t}$ pondéré par les w_i permet une inférence sur la population 2004 encore vivante à $2004 + t$; on note \bar{Y}_{2003} l'estimateur associé.
- l'échantillon r_{2004+t} des répondants en $2004 + t$, pondéré par les $w_i \cdot \frac{1}{p_i^{t/2004}}$ permet une inférence sur la population 2004 encore vivante à $2004 + t$

Sous l'hypothèse d'absence de biais dû à la non-réponse, l'espérance de Y_t^{2003} conditionnelle à l'échantillonnage *Silc* de 2004 est égale à \bar{Y}_{2003} . \bar{Y}_{2003} doit donc se trouver dans l'IC calculé précédemment. Si ce n'est pas le cas on peut conclure à un biais d'attrition avec une erreur de première espèce de l'ordre de 5 %.

Graphique III
Moyenne et intervalle de confiance du revenu disponible de 2003 des individus



Lecture : le revenu disponible moyen de 2003 des individus panels encore vivants en 2005 s'élève à 31 633 euros annuels : ce chiffre se trouve compris dans l'intervalle de confiance [31 546-32 044] du revenu disponible moyen calculé sur les individus panels répondants en 2005.

Champ : individus panels âgés de 16 ans ou plus.

Source : enquête SRCV 2004-2011, Insee.

tous sauf un, subissent une augmentation du revenu moyen disponible entre 2006 et 2007, en moyenne de l'ordre de 1 % alors que le revenu moyen estimé parmi la population encore vivante ne progresse que de 0,2 % en moyenne chaque année. L'écart se resserre ensuite fortement lorsqu'on passe aux répondants de 2008. Le fait que cette réduction coïncide avec la fin de l'obligation de réponse suggère que celle-ci affecte davantage le comportement des ménages à revenu plus élevé, d'une manière que les redressements ne corrigent qu'imparfaitement. Après cela, l'écart reprend un mouvement ascendant et il retrouve pour les répondants de 2011 le niveau atteint pour ceux de 2007.

Au total, le redressement ne gomme pas totalement les effets de l'attrition, mais les écarts qui subsistent restent cependant faibles et ne sont significatifs que pour ces deux pics de 2007 et 2011. Même pour ces deux années-là, l'écart entre les moyennes ne dépasse cependant pas 2 %. L'effet résiduel de l'attrition après redressements apparaît au final très modéré. Mais, les résultats observés sur l'ensemble du revenu disponible ne présagent pas de ceux qu'on peut avoir sur ses différentes composantes.

Sur la principale composante du revenu disponible, à savoir les revenus salariaux, on peut

encore conclure à l'absence de biais d'attrition une fois introduites les corrections pour la non-réponse. Le profil des écarts est particulièrement plat, ils sont d'au plus 1 % et jamais significatifs (cf. graphique IV).

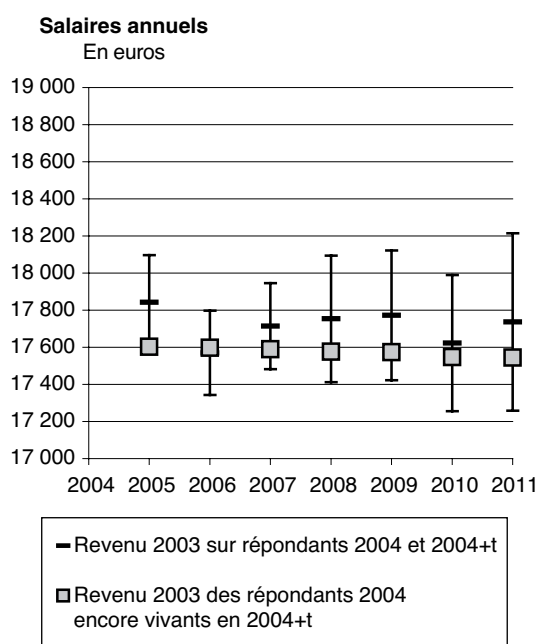
En ce qui concerne les aides au logement, les prestations familiales et les allocations-chômage, on peut également constater l'absence de biais d'attrition sur le montant annuel moyen de prestation, hormis en 2007 où l'on est en dehors de l'intervalle de confiance pour les aides au logement et les allocations-chômage et sur la borne supérieure de l'intervalle de confiance pour les prestations familiales (cf. graphiques V, VI et VII). Cette constatation rejoint l'observation faite sur le revenu disponible sur 2007 et semble indiquer que le poids des personnes bénéficiaires de prestations est sous-estimé au bénéfice de personnes non bénéficiaires et disposant en moyenne d'un revenu plus élevé.

... à l'exception des retraites

En revanche, un biais d'attrition significatif semble subsister concernant la retraite¹¹

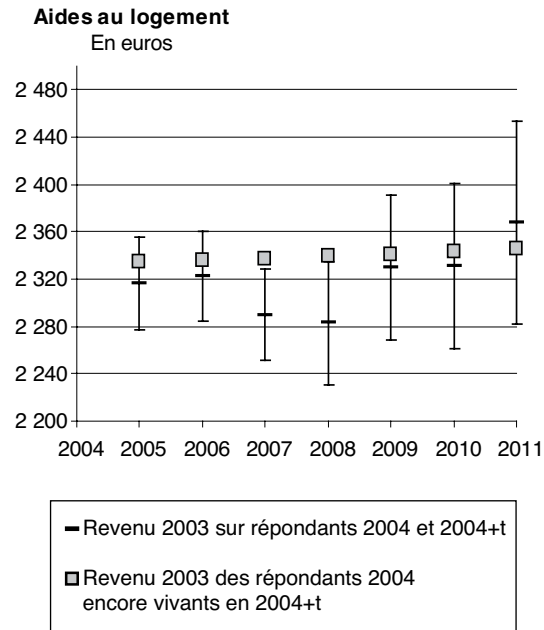
11. On entend ici par retraite, les retraites de droits directs et de droits dérivés des régimes obligatoires de sécurité sociale.

Graphique IV
Moyenne et intervalle de confiance des salaires annuels de 2003 des salariés



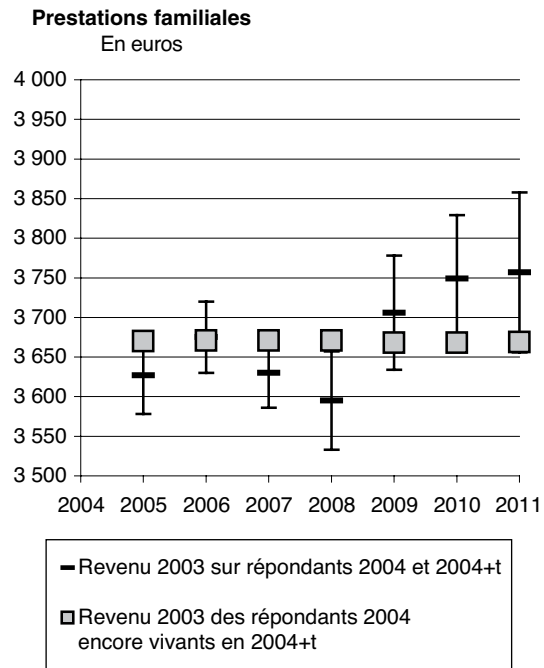
Lecture : le salaire estimé moyen de 2003 des individus panels salariés répondants en 2004 et encore vivants en 2006 s'élève à 17 600 euros. Il se trouve dans l'intervalle confiance [17 340 - 17 800] du salaire estimé moyen de 2003 (corrigé uniquement de la non-réponse en réinterrogation) des individus panels salariés répondants en 2004 et 2006.
Champ : individus panels salariés âgés de 16 ans ou plus.
Source : enquête SRCV 2004-2011, Insee.

Graphique V
Moyenne et intervalle de confiance du montant annuel des aides au logement de 2003 des individus bénéficiaires



Lecture : le montant estimé moyen de 2003 des aides au logement des individus panels bénéficiaires répondants en 2004 et encore vivants en 2006 s'élève à 2 320 euros. Il se trouve dans l'intervalle confiance [2 280 - 2 360] du montant estimé moyen de 2003 des aides au logement (corrigé uniquement de la non-réponse en réinterrogation) des individus panels bénéficiaires répondants en 2004 et 2006. Champ : individus panels âgés de 16 ans ou plus et bénéficiaires d'aides au logement. Source : enquête SRCV 2004-2011, Insee.

Graphique VI
Moyenne et intervalle de confiance du montant annuel des prestations familiales de 2003 des individus bénéficiaires



Lecture : le montant estimé moyen de 2003 des prestations familiales des individus panels bénéficiaires répondants en 2004 et encore vivants en 2006 s'élève à 3 670 euros. Il se trouve dans l'intervalle confiance [3 630 - 3 720] du montant estimé moyen de 2003 des prestations familiales (corrigé uniquement de la non-réponse en réinterrogation) des individus panels bénéficiaires répondants en 2004 et 2006. Champ : individus panels âgés de 16 ans ou plus et bénéficiaires de prestations familiales. Source : enquête SRCV 2004-2011, Insee.

(cf. graphique VIII). En effet, à partir de la quatrième année de présence dans le panel (année de revenu 2007), le revenu moyen estimé sur la population des retraités encore vivants sort de l'intervalle de confiance de la retraite moyenne des répondants ; l'effet est particulièrement marqué pour la 7^e et la 8^e interrogation. Le mouvement est monotone croissant. Les montants moyens de retraites annuelles calculées pour les individus panels s'élèvent d'une réinterrogation à la suivante : on ne conserve donc que les retraités les plus aisés et contrairement au constat effectué précédemment sur les autres prestations, il n'y a pas de réajustement de niveau qui s'opère en 2008.

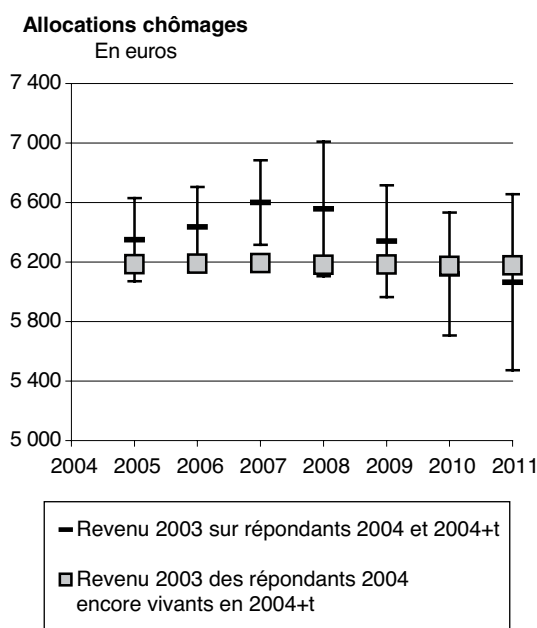
L'attrition est en fait particulièrement sélective sur les personnes les plus âgées, les individus restant dans le panel ayant des retraites en moyenne plus élevées que les sortants du panel. Ainsi, les individus présents en 2004 et toujours présents en 2009 perçoivent des retraites d'un montant supérieur de 11,6 % aux retraites des individus absents en 2009.

Plusieurs pistes peuvent être envisagées pour expliquer cet écart croissant d'une vague à l'autre.

Une première explication possible est l'effet de la mortalité différentielle. Pour calculer le niveau de vie moyen des individus encore vivants à la date $2004 + t$ on s'est appuyé dans un premier temps sur les quotients de mortalité générale par âge, sexe et année (Beaumel *et al.*, 2012). Ceci ne tient pas compte de la sélection induite par la mortalité différentielle selon le revenu (Jusot, 2004). Pour mieux tenir compte de cet effet, on a reconduit le calcul en recourant aux quotients de mortalité différentielle par catégorie-socio-professionnelle, sexe et âge détaillés (Blanpain *et al.*, 2011) sur la période 2000-2008, la catégorie socio-professionnelle étant un descripteur du revenu. L'estimation de la pension moyenne à partir de ces taux de mortalité différentielle ne se révèle cependant que peu éloignée du calcul précédent par âge et sexe. Au demeurant, la moyenne des retraites des individus panels décédés (décès connus) n'est inférieure que de 1,3 point à celle de l'ensemble des absents en 2009. Donc l'écart observé entre les différentes estimations du graphique VIII ne peut s'expliquer par la seule mortalité différentielle.

Le seconde explication potentielle est le départ en institution des retraités. En toute rigueur, il

Graphique VII
Moyenne et intervalle de confiance du montant annuel des allocations chômages de 2003 des individus bénéficiaires



Lecture : le montant estimé moyen de 2003 des allocations chômages des individus panels bénéficiaires répondants en 2004 et encore vivants en 2006 s'élève à 6 430 euros. Il se trouve dans l'intervalle confiance [6 160 - 6 700] du montant estimé moyen de 2003 des allocations chômages (corrigé uniquement de la non-réponse en réinterrogation) des individus panels bénéficiaires répondants en 2004 et 2006.

Champ : individus panels âgés de 16 ans ou plus et bénéficiaires d'allocations chômages.

Source : enquête SRCV 2004-2011, Insee.

conviendrait de comparer la pension moyenne des individus panels répondants à celle des individus encore vivants et continuant de résider en ménage ordinaire. Par construction, les individus panels répondants ne peuvent représenter que des individus vivant en ménages ordinaires, les collectivités n'étant pas couvertes par le champ de l'enquête. Cependant, selon les résultats du recensement de 2010, seuls 6 % de la population âgée de 65 ans ou plus vivent en collectivité ; il est donc assez peu probable qu'une estimation de la pension moyenne des individus encore vivants en ménage ordinaire soit très différente de l'estimation faite sur l'ensemble des individus panels encore en vie. Du reste, comme on ne dispose d'aucune information sur les différences de niveaux de revenus entre les populations vivant en collectivité et celles vivant en ménage ordinaire, il est impossible de dire a priori dans quel sens jouerait cette correction. L'exploitation statistique des fichiers fiscaux devrait permettre à terme d'éclairer cette question.

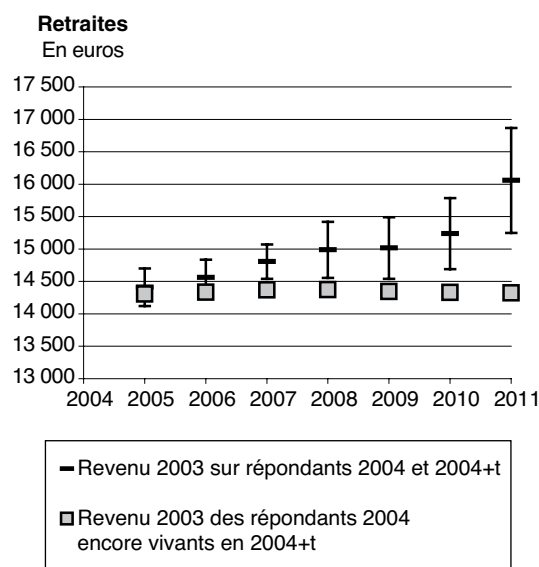
En attendant, les résultats montrent clairement un biais d'attrition non complètement résolu sur la mesure des retraites dans l'enquête. Les résultats de la régression logistique précédemment présentés indiquent bien que l'âge a un effet significatif sur la non-réponse et ce, indépendamment de l'effet de déménagement qui

concerne un certain nombre de jeunes retraités. Pour remédier à cette erreur de mesure, il pourrait être utile de tester des modèles de correction de la non-réponse séparément sur les personnes âgées de plus de 65 ans et sur les moins de 65 ans. Ainsi pourrait-on distinguer d'autres variables explicatives (santé, conditions de vie) que celles utilisées en population générale et tenir compte en particulier des changements récents de situation des individus. Mais la situation peut résulter d'autres variables explicatives inobservables, relevant de l'enquêté, voire de l'enquêteur.

* *
*

L'expérience du panel *SRCV* en France confirme que l'attrition est un phénomène important sur longue période (60 % des individus ont disparu du panel au bout de huit années) et qu'il n'est pas si simple de le prendre en compte dans la pondération. La correction de la non-réponse réussit à éliminer une bonne partie du biais qui résulte de cette attrition mais en laissant subsister un certain nombre d'écarts. Ainsi, si la plupart des composantes du revenu disponible semblent correctement corrigées du biais d'attrition, il en est autrement pour les retraites.

Graphique VIII
Moyenne et intervalle de confiance du montant annuel des pensions de 2003 des retraités



Lecture : le montant estimé moyen de 2003 des retraites (de droit direct et dérivé) des individus panels répondants en 2004 et encore vivants en 2006 s'élève à 14 560 euros. Il se trouve dans l'intervalle de confiance [14 290 - 14 830] du montant estimé moyen de 2003 des retraites (corrigé uniquement de la non-réponse en réinterrogation) des individus panels bénéficiaires répondants en 2004 et 2006. Champ : individus panels retraités âgés de 16 ans ou plus. Source : enquête SRCV 2004-2011, Insee.

Par ailleurs, l'analyse sur les individus présents en 2004 montre également que la dernière année en réinterrogation obligatoire, soit la quatrième, n'est pas correctement corrigée du biais d'attrition. Il conviendrait de voir si ces résultats se retrouvent sur d'autres années d'exploitations et, si tel est le cas, de mettre en place des mesures pour améliorer la correction de la non-réponse dans ces deux cas. Celles-ci pourraient utiliser davantage d'informations sur le passé récent, notamment sur le passage à des situations de pauvreté. Enfin, on pourrait également tester l'intégration d'autres variables explicatives de la non-réponse spécifiques aux retraités. La méthodologie mise en œuvre dans cette étude pourrait ensuite servir à valider ces méthodes de redressement alternatif.

Ceci étant, l'incidence de ces biais d'attrition sur la qualité générale de l'enquête ne doit pas être surestimée. Les données de l'enquête font aussi l'objet d'une repondération transversale issue d'un calage sur marges sur variables catégorielles, et les masses de revenu comme des indicateurs d'inégalité sont contrôlés avec l'enquête française de référence sur les revenus

fiscaux. On peut supposer que ce calage redresse une partie des biais non traités dans les étapes précédentes du calage.

Il est cependant plus difficile de contrôler la pondération longitudinale, en l'absence de données de validation disponibles en longitudinal à partir d'autres enquêtes ou des données administratives. Néanmoins, le principe de l'estimation longitudinale consistant à pratiquer une inférence sur la base d'une population considérée à une date initiale, cette pondération fait également l'objet d'un calage sur marges, sur variables catégorielles de la population initiale visant la représentativité de cette population. Il reste cependant nécessaire d'apporter quelques précautions d'utilisation quant à l'exploitation à ce stade des retraites sur une durée longue de panel au vu des résultats.

Il serait enfin intéressant de prolonger ce type de travaux en étudiant l'impact sur les indicateurs d'inégalité de pauvreté mais aussi sur les indicateurs composant l'indicateur européen de pauvreté et d'exclusion sociale dont l'enquête *EU-Silc/SRCV* constitue le support. □

BIBLIOGRAPHIE

Ardilly P. et Lavallée P. (2007), « Pondération dans les échantillons rotatifs : le cas de l'enquête Silc en France », *Statistique Canada techniques d'enquêtes*, vol. 33, n° 2, pp. 149-156.

Blanpain N. et Chardon O. (2011), « Les inégalités sociales face à la mort - Tables de mortalité par catégorie sociale et indices standardisés de mortalité pour quatre périodes », *Document de travail* n°F1108, Insee.

Beaumel C., Bellamy V. et Pla A. (2012), *Données détaillées des statistiques d'état civil sur les décès en 2011*, Insee Résultats, n° 136, Série société.

Breuil-Genier P. et Valdelièvre H. (2001), « Le Panel européen : l'intérêt d'un panel d'individus », *Économie et Statistique*, n° 349-350, pp. 17-40.

Brzinsky-Fay C. (2007), « Lost in transition? Labour market entry sequences of school leavers in Europe », *European Sociological Review*, n° 23(4), pp. 409-422.

Fumagalli L., Laurie H., Lynn P. (2010), « Experiments with methods to reduce attrition in longitudinal surveys », *ISER Working Paper*, Institute for Social and Economic Research, n° 2010-04.

Graf E. (2010), « Étude 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.

Hill D. et Willis R. (2001), « Reducing Panel Attrition: A Search for Effective Policy Instruments », *The Journal of Human Resources*, n° 36 (3), pp. 416-38.

Jusot F. (2004), « Mortalité et inégalités de revenu en France », *Document de travail Delta*, 2004/09, Insee.

Laurie H. et Lynn P. (2008), « The Use of Respondent Incentives on Longitudinal Surveys », *ISER Working Paper*, Institute for Social and Economic Research, n° 2008-42.

Laurie H., Smith R. et Scott L. (1999), « Strategies for reducing nonresponse in a longitudinal panel survey », *Journal of Official Statistics*, vol.15 n° 2, pp. 269-282.

Lipps O. (2009), « Attrition of households and individuals in panel surveys », *SOEP papers*, n° 164.

Lynn P., Kaminska O. et Goldstein H. (2011), « Panel attrition: how important is to keep the same interviewer? », *ISER Working Paper Series*, Institute for Social and Economic Research, n° 2011-2.

Minodier C. (2006), « Changer de logement dans le même environnement », *Données sociales - La société française*, pp. 515-523, Insee.

Rendtel U. (2002), « Attrition in households Panels : a survey », *Chintex Working paper*, n° 4 .

Uhrig N. (2008), « The nature and causes of Attrition in the British Household Panel Survey », *ESRC Economic & Social research Council*, n°2008-05.
