

CALCUL DE PRÉCISION DES PRINCIPAUX ESTIMATEURS DE L'ENQUÊTE 2012 SUR LES CONDITIONS DE VIE DES BÉNÉFICIAIRES DE MINIMA SOCIAUX DE LA DREES

*Audrey ISEL*¹ (*), *Emmanuel GROS*² (**)

(*) *Insee, Direction régionale d'Alsace, Service Études et Diffusion*³

(**) *Insee, Direction générale, Division Sondages*

Résumé

L'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux est une enquête d'intérêt général menée par la Drees - service statistique des ministères sanitaires et sociaux. Au-delà des nombreuses informations qualitatives que cette enquête apporte sur ces ménages aux revenus très modestes, son objectif central est de fournir une mesure approchée de leur "reste-à-vivre" et d'estimer leur éventuelle marge de manoeuvre financière. Pour mieux appréhender la qualité de cette enquête, un travail de calcul de précision des principaux estimateurs a été mené.

L'échantillon de l'enquête a été obtenu via un tirage à trois degrés. À chaque degré de tirage, des stratifications préalables ont été effectuées et des ajustements ont pu être nécessaires. Un système récursif basé sur la formule de RAJ a été mis en oeuvre pour estimer la variance d'échantillonnage des principaux estimateurs retenus relatifs aux allocataires du minimum vieillesse. Ces calculs de précision ont dû être ajustés pour prendre en compte toute la complexité du plan de sondage et la réalité du tirage de l'échantillon. Dans un second temps, afin d'intégrer le traitement de la non-réponse totale dans le calcul de précision, cette phase a été assimilée à un degré de sondage supplémentaire via la formule de RAO. Et finalement, la méthode proposée par DEVILLE et SÄRNDAL a été mise en oeuvre pour tenir compte du redressement de l'enquête par calage sur marges.

1. audrey.isel@insee.fr

2. emmanuel.gros@insee.fr

3. Cet article présente les travaux qui ont été menés dans le cadre du Master 2 de Statistique - Économétrie, spécialité statistique publique, proposé par l'Ensaï et l'Université de Rennes 1 [1].

Abstract

In 2012, the living conditions of beneficiaries of the minimum allowances have been investigated by the Drees, which is the statistical service of the French Minister for Social Affairs, Health and Women's Rights. This survey is a centrepiece of the French statistical system about people in financial difficulties. It provides much qualitative information about beneficiaries of the government-guaranteed minimums, and leads in particular to the estimation of their subsistence allowance.

This survey has been sampled using a three-stage sampling design, with stratified random samplings at the first two stages and a systematic sampling at the third stage. This article focuses on the analytical approach used to estimate the variance of estimators in this survey. At first, a recursive formula based on the RAJ formula was used to estimate the sampling variance. Then the weighting adjustment for non-response was taken into account thanks to the RAO formula. Finally, following DEVILLE and SÄRNDAL, the use of calibration techniques is taken into account by applying the variance formula to the residuals, which are the result of a weighted least squares regression of the variable under study on the calibration variables.

Mots-clés

Enquête ménages, estimation de variance, sondage à plusieurs degrés, formule de RAJ, formule de RAO.

Introduction

L'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux menée par la Drees - service statistique des ministères sanitaires et sociaux - s'inscrit dans le cadre du dispositif d'observation statistique des situations des populations en difficultés. Elle répond à une demande ministérielle et à des requêtes du secteur associatif d'avoir des éléments d'information permettant d'éclairer le débat public sur les minima sociaux et leur montant.

Alors que le débat public a surtout porté depuis la fin des années 1990 sur l'incitation au retour à l'emploi des bénéficiaires de minima sociaux d'âge actif, les questions sur leur niveau de vie reviennent sur le devant de la scène ces dernières années à la faveur notamment des débats sur l'évolution du pouvoir d'achat. Le poids grandissant des dépenses pré-engagées dans les revenus les plus modestes pose en effet la question de leur capacité à "boucler leurs fins de mois". Cette enquête permettra à terme d'apporter une mesure approchée du "reste-à-vivre" des bénéficiaires de minima sociaux et d'estimer leur éventuelle marge de manoeuvre financière. En attendant, elle donne de nombreuses informations qualitatives sur les conditions de vie de ces personnes aux revenus très modestes.

Il s'agit de la troisième édition de cette enquête, mais aucune estimation de la précision des estimateurs construits à partir des données recueillies n'avait encore été réalisée jusque

là. Un tel travail paraît d'autant plus nécessaire pour cette édition, que l'un de ses principaux objectifs est de fournir une quantification du reste-à-vivre des bénéficiaires. Il est donc important de connaître le degré de confiance que l'on peut accorder aux résultats ainsi obtenus.

1 Présentation de l'enquête et de son plan de sondage

Pour introduire les travaux de calcul de précision qui ont été menés, il est nécessaire de commencer par présenter le sujet de l'étude : l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux. Son plan de sondage, fondement de l'exercice d'estimation de la variance d'échantillonnage qui constitue l'objet de cet article, est ensuite décrit plus en détails. Les données sur lesquelles a porté plus particulièrement le travail réalisé sont ensuite exposées. En particulier, il convient de s'attarder quelque peu sur les principales caractéristiques des allocataires du minimum vieillesse, qui forment la sous-population d'étude dans le cadre de ces travaux.

1.1 L'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux

L'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux est une enquête ayant obtenu le label d'intérêt général et de qualité statistique, sans obligation de réponse. C'est une enquête en face-à-face d'une durée d'une heure environ, qui a été réalisée au quatrième trimestre 2012. Une centaine d'enquêteurs de TNS Sofres, mandatés par la Drees, ont ainsi interrogé 8 450 personnes qui bénéficiaient, au 31 décembre 2011, d'un minimum social [2]. Les personnes interrogées n'étaient donc plus forcément bénéficiaires de ce minimum social au moment de l'enquête.

Au total, 18 000 fiches-adresses ont été nécessaires pour atteindre les objectifs fixés :

- 7 400 fiches-adresses au titre du revenu de solidarité active (RSA) pour 3 850 répondants (1 450 du RSA socle non majoré, 1 100 du RSA socle majoré et 1 300 du RSA activité seul)
- 3 200 fiches-adresses au titre de l'allocation de solidarité spécifique (ASS) pour 1 800 répondants
- 3 200 fiches-adresses au titre de l'allocation aux adultes handicapés (AAH) pour 1 400 répondants
- 4 300 fiches-adresses au titre du minimum vieillesse (dont 1 000 provenant de l'échantillon complémentaire) pour 1 400 répondants

Par rapport aux précédentes enquêtes réalisées auprès des bénéficiaires de minima sociaux en 2003 et en 2006, deux nouveaux publics ont été inclus dans le champ de cette enquête : les travailleurs pauvres, désormais en partie identifiables car percevant le RSA activité seul, et les allocataires du minimum vieillesse.

Bases de sondage :

Pour le RSA, l'AAH et l'ASS, les unités enquêtées font partie de la vague 2011 de l'Échantillon national interrégimes d'allocataires de minima sociaux (Eniams)⁴.

L'intégration des allocataires du minimum vieillesse dans le champ de cette troisième enquête a nécessité d'utiliser une autre base de sondage que l'Eniams. Cet échantillon ne contient en effet que des personnes d'âge actif. Les bénéficiaires du minimum vieillesse ont ainsi été directement sélectionnés dans les bases administratives des trois principales caisses compétentes : la Caisse nationale de l'assurance vieillesse (Cnav), la Caisse centrale de la mutualité sociale agricole (CCMSA) et le Service de l'allocation de solidarité aux personnes âgées de la Caisse des dépôts (SASPA). Une même personne ne pouvant bénéficier d'une allocation que dans un seul régime, les trois bases de sondage sont bien disjointes. Celles-ci regroupent au total environ 96% de l'ensemble des allocataires du minimum vieillesse.

Cet article se concentre sur les allocataires du minimum vieillesse relevant de la Cnav (Figure 1).

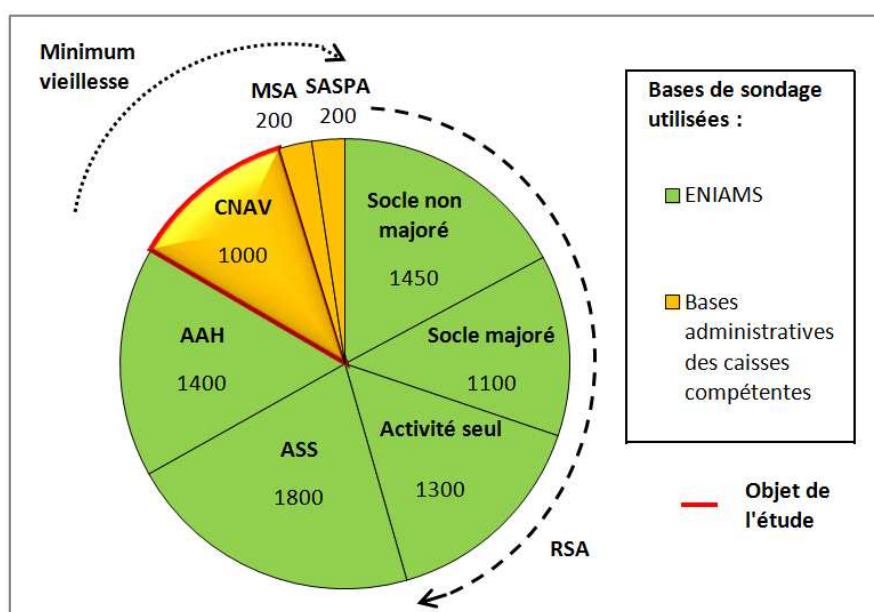


FIGURE 1 – Identification de la population sur laquelle porte l'article parmi l'ensemble des répondants

Les bases de sondage utilisées pour sélectionner les bénéficiaires du minimum vieillesse sont en partie disjointes de la base de sondage dont on dispose pour les bénéficiaires d'âge actif (l'Eniams). En effet, l'ASS et le RSA ne peuvent être, pour des raisons d'âge d'éligibilité, cumulés avec le minimum vieillesse. En revanche, l'AAH et le minimum vieillesse

4. Il s'agit d'un panel annuel constitué par rapprochement de données administratives individuelles. Depuis 2001, il permet de suivre, au 31 décembre de chaque année, la situation par rapport aux minima sociaux et à l'inscription à Pôle emploi d'un échantillon de bénéficiaires de minima sociaux nés entre le 1^{er} et le 14 octobre et âgés de 16 à 64 ans.

sont cumulables dans quelques rares cas. Pour que les bases de sondage soient bien disjointes, les personnes nées entre le 1^{er} et le 14 octobre n'ont pas été sélectionnées dans les bases de sondage des organismes verseurs du minimum vieillesse. Ce défaut de couverture de notre échantillon, qui pourrait entraîner un problème de biais de nos estimateurs, a été corrigé par le calage de notre enquête. Le calage sur marges a en effet été réalisé sur le champ complet des allocataires du minimum vieillesse, et non pas sur ce champ privé des individus nés entre le 1^{er} et le 14 octobre.

Unités statistiques :

L'Eniams contient les allocataires administratifs des différentes prestations, mais aussi leurs éventuels conjoints dans le cas de prestations familialisées.

Dans le cas du RSA, la prestation étant attribuée à un foyer, l'éventuel conjoint bénéficie, au même titre que l'allocataire administratif, de tout le dispositif qui entoure cette prestation, notamment son dispositif d'insertion. Ainsi, dans le cadre de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux, la personne enquêtée au titre du RSA peut être l'allocataire administratif ou son éventuel conjoint.

En revanche, bien que l'AAH soit également attribuée en fonction des ressources du foyer, le comité de pilotage de l'enquête a décidé d'exclure du champ de l'enquête les éventuels conjoints des allocataires de l'AAH. Le droit à cette prestation est en effet conditionné au handicap de l'allocataire, qui est une caractéristique personnelle.

Bien que l'ensemble des ressources du foyer soient prises en compte pour le calcul du droit au minimum vieillesse, seul l'allocataire administratif est repérable dans les bases de données administratives des caisses. Pour lever le flou de l'unité statistique dans le cas d'une personne en couple, une question supplémentaire avait été ajoutée dans le questionnaire pour savoir si le conjoint de la personne enquêtée était aussi allocataire de cette prestation.

Au final, pour l'ensemble de l'article, comme pour tous les résultats issus de l'enquête qui ont été publiés jusque là, on utilise indifféremment les termes de "bénéficiaires" et d'"allocataires" des différentes prestations.

Zones géographiques :

La sélection des bénéficiaires du minimum vieillesse a été réalisée par les caisses elles-mêmes, suivant le plan de sondage fourni par la Drees et dans les zones géographiques que la Drees a sélectionnées. Pour des raisons de coûts d'enquête, ces dernières sont les mêmes quelle que soit la prestation considérée. Pour des contraintes de calendrier des caisses de retraite, la sélection de ces unités géographiques s'est faite à partir de la vague 2010 de l'Eniams (la vague 2011 n'étant pas encore disponible). Il n'a pas semblé gênant de procéder de la sorte, la répartition géographique des bénéficiaires de minima sociaux évoluant en effet peu d'une année sur l'autre.

Par ailleurs, comme pour les enquêtes précédentes, les DOM ne font pas partie des départements retenus dans la collecte. L'enquête fournit donc uniquement des résultats représentatifs sur les bénéficiaires résidant en France métropolitaine.

Enfin, les 2 départements que compte la Corse ont à nouveau d'emblée été exclus du tirage des départements lors de cette enquête en raison de contraintes logistiques⁵. Néanmoins, les résultats obtenus peuvent raisonnablement être extrapolés à l'ensemble de la France métropolitaine, y compris la Corse, les dispositifs législatifs en vigueur étant les mêmes.

Hors-champs :

Les personnes enquêtées peuvent être dans n'importe quelle situation de logement : en logement de droit commun, dans un service d'hébergement, dans une habitation mobile, sans abri... Seules sont exclues du champ les personnes vivant dans des institutions les prenant presque entièrement en charge (maisons de retraite, foyers d'accueil médicalisés, maisons d'accueil spécialisées). Ces bénéficiaires ont été retirés du champ de l'enquête par les enquêteurs lors de la phase de prise de contact, car cette information n'était pas disponible en amont dans les bases de sondage.

Au total, les hors-champs représentent 2% de l'échantillon. Cette proportion est évidemment très variable d'un minimum social à un autre. Ainsi, les prestations d'âge actif (hors AAH) comptent moins de 1% d'hors-champs, tandis que 17% des bénéficiaires du minimum vieillesse relevant de la Cnav sont hors-champ (cette proportion dépasse même les 30% parmi les bénéficiaires du minimum vieillesse relevant des autres organismes).

1.2 Le plan de sondage de l'enquête

L'objectif du plan de sondage de cette enquête a d'abord été de se rapprocher le plus possible de la méthodologie de l'enquête effectuée en 2006 pour faciliter la comparaison des résultats. En raison de l'élargissement du champ de l'enquête, il a tout de même fallu adapter le plan de sondage et procéder à quelques ajustements.

Pour faciliter le travail des enquêteurs et pour des raisons de coût de collecte, les adresses ne devaient pas être trop dispersées. Le tirage a donc été effectué dans un nombre limité de départements et, dans chaque département, concentré autant que faire se peut sur certaines zones géographiques.

Comme en 2003 et en 2006, il s'agissait d'un tirage à trois degrés :

- 1^{er} degré : les départements⁶
- 2^e degré : les communes ou fractions de cantons
- 3^e degré : les individus

1.2.1 1^{er} degré de tirage : les départements

Préalablement au tirage des unités primaires, une stratification des départements en 5 strates a été effectuée, comme pour les enquêtes réalisées en 2003 et en 2006 :

5. La TNS Sofres, prestataire en charge du terrain de cette enquête et des précédentes, ne dispose pas d'un véritable réseau d'enquêteurs en Corse.

6. Hors DOM et Corse

- 2 strates exhaustives
 - Strate 1** : la région parisienne (8 départements)
 - Strate 2** : les 6 plus gros départements en termes d’allocataires de minima sociaux (tous minima confondus)
- 3 strates déterminées selon le taux d’urbanisation du département
 - Strate 3** : les départements très urbains ($\text{taux d'urbanisation} \geq 65\%$)
 - Strate 4** : les départements semi-urbains ($50\% \leq \text{taux d'urbanisation} < 65\%$)
 - Strate 5** : les départements ruraux ($\text{taux d'urbanisation} < 50\%$)

Dans chacune des strates 3, 4 et 5, 15 départements ont ensuite été sélectionnés, proportionnellement à leur taille en termes d’allocataires de minima sociaux. En ajoutant à ces 45 unités primaires, les 14 issues des deux premières strates exhaustives, l’enquête s’est déroulée au final dans 59 départements (Figure 2).

1.2.2 2^e degré de tirage : les communes et fractions de cantons

Comme pour les unités primaires, une stratification préalable des unités secondaires a été réalisée avant le tirage. Ainsi, dans chaque département sélectionné, les communes ont été classées en 3 strates :

- 1 strate exhaustive : les 4 plus grosses communes du département en termes de bénéficiaires de minima sociaux.
- 2 strates de cantons :
 - Strate urbaine** : elle est constituée des cantons dits “urbains”, c’est-à-dire regroupant des communes appartenant à l’espace urbain d’après le zonage en unités urbaines de l’Insee.
 - Strate rurale** : elle est constituée des cantons dits “ruraux”, c’est-à-dire regroupant les communes appartenant à l’espace rural.

Un même canton peut être présent dans les deux strates s’il regroupe des communes urbaines et des communes rurales. Ainsi, on parle de “fractions urbaines” et de “fractions rurales” de cantons.

Quel que soit le département, 6 fractions de cantons ont été tirées, proportionnellement à leur taille en termes de bénéficiaires de minima sociaux : le nombre de fractions de cantons appartenant à la strate de cantons urbains et celui appartenant à la strate de cantons ruraux varient en fonction du poids relatif de ces 2 strates dans chaque département ⁷.

7. 7 tranches ont été constituées selon le poids des bénéficiaires de minima sociaux appartenant à l’espace urbain d’un département dans le total des bénéficiaires de ce département. Pour chaque département, en fonction de la tranche à laquelle appartient son taux d’urbanisation, on détermine le nombre de fractions de cantons urbaines à tirer (de 0 à 6).



FIGURE 2 – Répartition par strate des départements dans lesquels l'enquête s'est déroulée

Adaptation de la procédure aux départements très urbanisés :

Cette procédure de stratification des fractions de cantons entre “rural et urbain” a été remise en cause dans quelques départements⁸. En effet, la règle permettant de déterminer le nombre de fractions à tirer dans chacune des 2 strates conduisait à ne sélectionner que des fractions de cantons urbaines, alors qu’il y avait tout de même quelques bénéficiaires appartenant à des fractions de cantons rurales dans ces départements.

Pour que chaque bénéficiaire ait effectivement une chance, même infime, d’être sélectionné, la stratification des fractions de cantons entre “rural et urbain” n’a pas été effectuée dans ce cas : 6 fractions de cantons ont été tirées dans l’ensemble du département (hors les 4 communes de la strate exhaustive), proportionnellement à leur taille en termes de bénéficiaires de minima sociaux.

Étape supplémentaire pour les minima sociaux d’âge actif :

Pour les seuls minima sociaux d’âge actif, 3 fractions de cantons rurales et 1 fraction de canton urbaine supplémentaires ont été tirées dans chaque département, proportionnellement à leur taille en termes de bénéficiaires de minima sociaux d’âge actif⁹. La base de sondage étant déjà un échantillon, il était nécessaire d’augmenter le nombre d’unités géographiques dans lesquelles il était possible de sélectionner les individus.

Pour les départements dans lesquels aucune stratification entre fractions de cantons rurales et urbaines n’avait été effectuée, ce sont simplement 4 fractions de cantons qui ont été sélectionnées.

Ces nouvelles fractions de cantons ont été tirées parmi l’ensemble des fractions possibles, les 6 fractions déjà sélectionnées précédemment n’ayant pas été retirées de la base. Il a donc fallu ensuite supprimer les éventuels doublons et procéder à un partage des poids pour déterminer le poids des unités.

Au final :

14 unités secondaires ont au maximum été sélectionnées par département :

- les 4 plus grosses communes en termes de bénéficiaires de minima sociaux (y compris minimum vieillesse)
- 6 fractions de cantons, généralement réparties entre “rural et urbain”, tirées proportionnellement à leur taille en termes de bénéficiaires de minima sociaux (y compris minimum vieillesse)
- un maximum de 4 fractions de cantons supplémentaires tirées proportionnellement à leur taille en termes de bénéficiaires de minima sociaux d’âge actif (généralement 3 fractions rurales et 1 urbaine)

8. 10 départements sont concernés : 78, 91, 95, 13, 59, 62, 69, 06, 67, 83.

9. A partir de la vague 2011 de l’Eniams cette fois.

Le cas des Alpes-de-Haute-Provence :

Ce deuxième degré de tirage a également dû être adapté pour une unité primaire en particulier : les Alpes-de-Haute-Provence. En effet, les bénéficiaires du RSA socle majoré et de l'ASS n'étaient pas assez nombreux dans l'Eniams par rapport à l'estimation du nombre de bénéficiaires qu'il aurait fallu sélectionner pour remplir l'objectif global de chaque minimum social¹⁰.

Ainsi, dans ce département, tous les bénéficiaires du RSA socle majoré et de l'ASS présents dans l'Eniams ont d'office fait partie de l'échantillon. Par conséquent, la constitution des 3 strates de communes de ce département, puis la sélection des 6 fractions de cantons, ont été faites en excluant ces 2 minima sociaux. Seuls les bénéficiaires du RSA socle non majoré, du RSA activité seul, de l'AAH et du minimum vieillesse ont été pris en compte dans ce degré de tirage pour la sélection des unités secondaires de ce département.

Le cas de Paris :

Paris a également été traité de façon spécifique, cette unité primaire étant aussi bien une commune qu'un département. Les 20 arrondissements que compte la capitale constituent autant d'unités secondaires. Ce 2^e degré de tirage a ainsi consisté à en sélectionner 10, proportionnellement à leur taille en termes d'allocataires de minima sociaux. Aucune stratification préalable n'a donc été réalisée dans ce cas.

1.2.3 Un 2^e degré de tirage pouvant être composé de 2 phases

Par rapport au plan de sondage des deux enquêtes précédentes, une nouvelle étape a dû être mise en place cette fois-ci au moment du tirage des unités secondaires lors du 2^e degré de tirage, en raison de l'extension du champ de l'enquête à des publics différents de ceux qui avaient été enquêtés jusque-là. En effet, la répartition géographique des bénéficiaires de minima sociaux n'étant évidemment pas la même d'une prestation à l'autre, la multiplicité de celles-ci peut conduire à retenir des unités géographiques regroupant peu de bénéficiaires d'une prestation en particulier.

Pour chaque minimum social d'âge actif et pour chacun des organismes en charge du minimum vieillesse, un test a été effectué dans chaque unité primaire sélectionnée : la prestation en question y est-elle un "critère rare" ? Autrement dit, est-ce que plus de 50% des fractions de cantons de ce département sont vides pour le minimum social considéré (hormis les 4 plus grosses communes du département regroupées dans la 1^{ère} strate) ?

Le nombre de départements concernés est très variable d'une prestation à une autre (Tableau 1).

10. Effectif attendu dans chaque département i (de probabilité d'inclusion π_i) pour un minimum social MS donné :

$$Eff_i = \frac{1}{\pi_i} * \frac{\text{taille de l'échantillon souhaitée pour ce MS}}{\text{nombre de bénéficiaires de ce MS dans la base de sondage}} * \text{nombre de bénéficiaires du MS dans } i$$

	RSA socle non majoré	RSA socle majoré	RSA activité seul	AAH	ASS	Minimum vieillesse - CNAV	Minimum vieillesse - MSA	Minimum vieillesse - SASPA
Nombre de départements dans lesquels le critère est rare	0	23	0	0	1	6	13	7

TABLE 1 – Nombre d’unités primaires concernées par un 2^e degré de tirage en 2 phases selon le minimum social considéré

Au sein de ces départements, le tirage des individus pour une prestation donnée n’a plus lieu dans toutes les fractions de cantons qui avaient été retenues jusque là.

Dans ces unités primaires, pour un minimum social donné, seules 2 fractions de cantons ont été sélectionnées dans les strates urbaine et rurale parmi celles qui avaient été retenues à l’étape précédente. Si chacune de ces 2 strates était représentée initialement, pour qu’elles continuent à l’être dans l’échantillon final, on sélectionnait forcément une fraction de canton au sein de chaque strate (et ce même si toutes les fractions de cantons d’une strate donnée étaient vides pour le minimum social considéré). La sélection de ces 2 unités secondaires se faisait proportionnellement à leur taille en termes de bénéficiaires du minimum social en question.

Au final, lorsque pour une caisse de retraite donnée, le minimum vieillesse est un critère rare dans une unité primaire, seules 6 unités secondaires ont été retenues au maximum : les 4 plus grosses communes du département en termes de bénéficiaires de minima sociaux, auxquelles s’ajoutent un maximum de 2 fractions de cantons tirées spécifiquement parmi les 6 qui avaient été sélectionnées au préalable pour l’ensemble des bénéficiaires. Pour les minima sociaux d’âge actif, la seule différence réside dans le nombre de fractions de cantons au sein desquelles il était possible d’effectuer ce tirage : 10 fractions au maximum au lieu de 6 pour le minimum vieillesse.

1.2.4 3^e degré de tirage : les individus

Dans chaque unité secondaire préalablement sélectionnée, le tirage des individus a été effectué prestation par prestation pour les minima sociaux d’âge actif et organisme gestionnaire par organisme gestionnaire pour le minimum vieillesse. Ces derniers ont d’ailleurs été réalisés par les caisses elles-mêmes selon les instructions fournies par la Drees.

Ce 3^e degré de tirage est un tirage systématique d’allocataires dans chaque commune ou fraction de canton retenue, sur un fichier trié par âge et sexe. Le nombre d’unités à tirer dans les différentes unités secondaires est calculé pour chaque minimum social¹¹, suivant la taille de l’échantillon final souhaitée et de façon à tenir compte de l’inégalité d’effectifs et de répartition sur le territoire des bénéficiaires.

11. Effectif à sélectionner dans chaque unité secondaire q du département i (de probabilités d’inclusion respectives $\pi_{q|i}$ et π_i) pour un minimum social MS donné :

$$Eff_{q|i} = \frac{1}{\pi_i * \pi_{q|i}} * \frac{\text{taille de l'échantillon souhaitée pour ce MS}}{\text{nombre de bénéficiaires de ce MS dans la base de sondage}} * \text{nombre de bénéficiaires du MS dans } q$$

Par ailleurs, les bénéficiaires du RSA socle non majoré appartenant à une famille nombreuse ont été surpondérés à la demande de la Cnaf (d'un coefficient de 1,3), afin de pouvoir mener des études spécifiques sur ce public par la suite.

Enfin, un échantillon principal a été tiré au sein de cet échantillon, de sorte qu'il soit équilibré selon différentes variables (sexe, tranche d'âge, département, rural/urbain). La probabilité de faire partie de l'échantillon principal pour un bénéficiaire présent dans l'échantillon global a été calculée pour chaque minimum social et était la même pour tous ses bénéficiaires¹². Les échantillons complémentaires obtenus permettaient de faire face en cas de difficulté à atteindre les objectifs fixés en termes de répondants. Ces échantillons étaient à débloquent par minimum social et devaient alors être exploités complètement, quitte à dépasser les effectifs prévus, afin d'éviter tout biais dans la collecte des réponses. Seul l'échantillon complémentaire du minimum vieillesse a finalement été débloquent.

12. Pour chaque prestation, cette probabilité est égale au rapport entre la taille souhaitée de l'échantillon principal et la taille de l'échantillon global qui a été sélectionné.

1.2.5 Conclusion sur la procédure de sélection des allocataires du minimum vieillesse

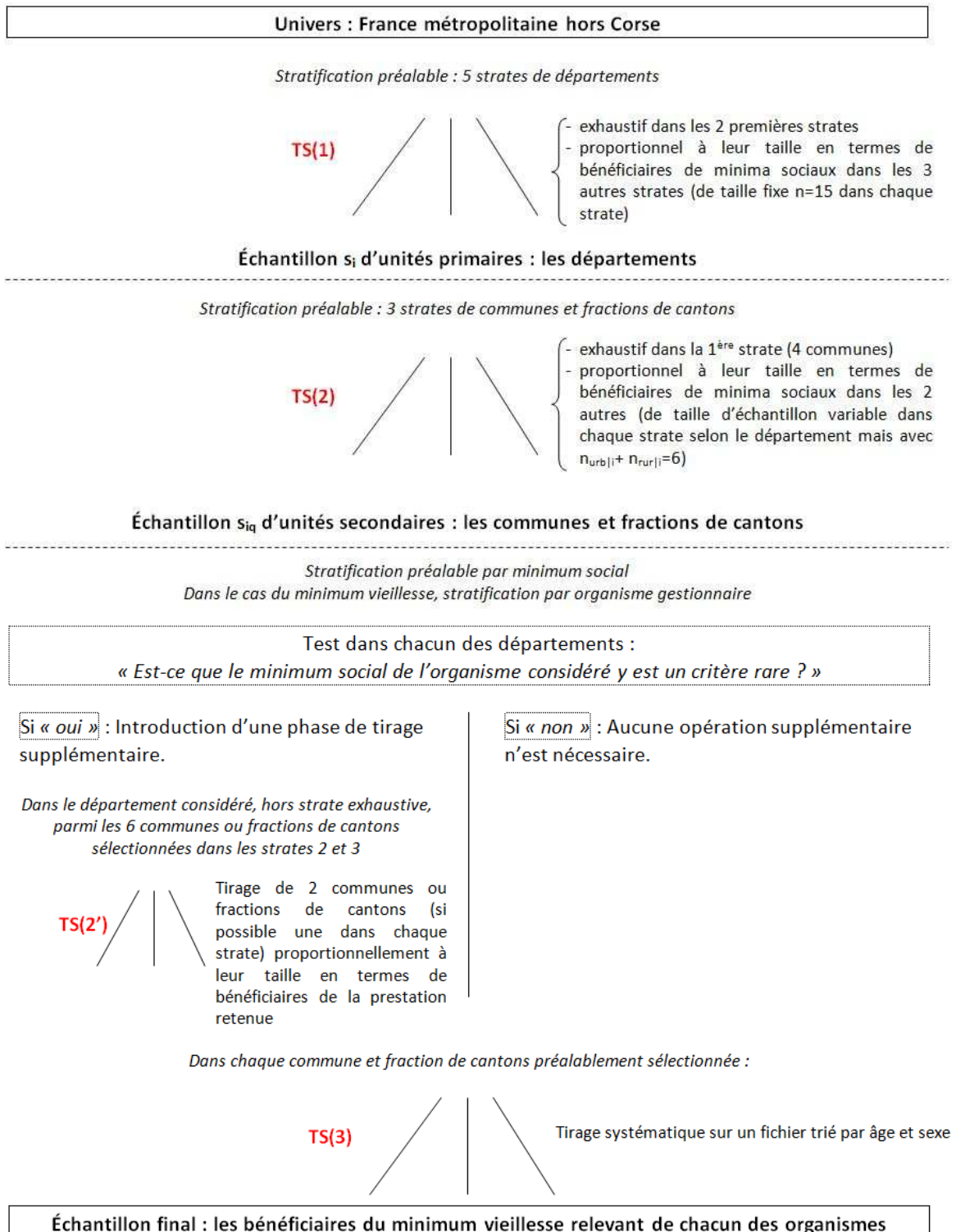


FIGURE 3 – Schéma résumant la procédure de sélection des allocataires du minimum vieillesse

1.3 Les données

Nous nous sommes concentrés, dans le cadre de ce travail, sur une population particulière de bénéficiaires de minima sociaux : les allocataires du minimum vieillesse (en particulier ceux relevant de la Cnav). La première publication parue à partir des résultats de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux leur avait d'ailleurs été consacrée [3].

Par ailleurs, nous nous sommes particulièrement intéressés à un indicateur phare de l'enquête : l'indicateur de pauvreté en conditions de vie. Il a lui aussi fait l'objet d'une publication spécifique [4].

1.3.1 Le minimum vieillesse

Le minimum vieillesse permet aux personnes âgées d'au moins 65 ans (60 ans en cas d'inaptitude au travail) d'atteindre un seuil minimal de ressources. Historiquement, il s'agit du premier minimum social qui a été créé en France (l'allocation aux vieux travailleurs date de 1941).

Au moment de l'enquête, au 31 décembre 2012, le plafond de ressources mensuelles pour pouvoir bénéficier du minimum vieillesse était de 777 euros par mois pour une personne seule et de 1 207 euros pour un couple. L'allocation est différentielle : elle correspond à la différence entre ce plafond de ressources et le revenu mensuel du foyer. Le SASPA verse le minimum vieillesse aux personnes qui ne perçoivent aucune pension de retraite, sinon ce sont les caisses de retraite (Cnav, CCMSA...) qui s'en chargent.

D'après les données administratives fournies par les caisses de retraite sur les allocataires de notre échantillon, un allocataire du minimum vieillesse relevant du régime général (Cnav) perçoit en moyenne 317 euros par mois d'allocation (Table 2).

Organisme	Moyenne	Médiane	Q1	Q3
CNAV	316,87	307,79	177,67	458,87
MSA Vieillesse	240,74	215,33	96,00	373,67
SASPA	664,33	737,10	652,14	742,27

Source : DREES, enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux.

TABLE 2 – Statistiques descriptives de la variable de montant mensuel de minimum vieillesse perçu

La plupart de ces allocataires ont déjà travaillé par le passé. Néanmoins 28% d'entre eux ont connu de longues périodes d'inactivité ou n'ont jamais travaillé d'après les résultats de l'enquête. Parmi les allocataires relevant de la Cnav, cette proportion s'établit à 22%. Les problèmes de santé, d'invalidité ou de handicap semblent être à l'origine de nombreuses

carrières écourtées.

Enfin, peu d'allocataires du minimum vieillesse disposent d'un patrimoine immobilier : ils ne sont que 16% à être propriétaires, accédants à la propriété ou usufruitiers. Avec plus d'un tiers de personnes concernées, ce sont les allocataires du régime agricole qui en détiennent le plus souvent un.

1.3.2 Un indicateur phare de l'enquête : l'indicateur de pauvreté en conditions de vie

Le rôle des minima sociaux est avant tout de garantir un revenu minimum décent à leurs bénéficiaires. Grâce à l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux, il est possible d'objectiver la situation de ces personnes. En attendant l'étude de la pauvreté monétaire¹³, une autre forme de pauvreté a largement été analysée : la pauvreté en conditions de vie.

En France, un ménage est dit pauvre en conditions de vie lorsqu'il cumule au moins 8 difficultés parmi les 27 retenues, regroupées en 4 grands domaines que sont les contraintes budgétaires, les retards de paiement, les restrictions de consommation et les difficultés liées au logement. Les résultats obtenus à partir de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux sont parfaitement comparables avec ceux obtenus en population générale par l'Insee via l'enquête statistique sur les ressources et les conditions de vie (dispositif SRCV).

A la fin 2012, 57% des bénéficiaires de minima sociaux qui résident en France métropolitaine sont pauvres en conditions de vie. Les revenus garantis par ces différentes prestations et leurs conditions d'éligibilité dessinent la hiérarchie entre les minima. Comme le minimum vieillesse garantit des revenus plus élevés que les minima sociaux dits d'insertion (RSA et ASS)¹⁴, ses bénéficiaires sont un peu moins souvent pauvres en conditions de vie. 47% d'entre eux sont tout de même dans cette situation, soit bien plus que les plus modestes en population générale. En se limitant aux allocataires du minimum vieillesse relevant du régime général (Cnav), ce sont 141 784 allocataires qui sont pauvres en conditions de vie.

Conclusion

Avant l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux, deux enquêtes similaires avaient déjà été menées par la Drees auprès des bénéficiaires de minima sociaux (en 2003 et 2006). Afin que les résultats obtenus lors de cette nouvelle enquête soient comparables avec ceux des deux précédentes éditions, le plan de sondage devait se rapprocher le plus possible de ceux mis en oeuvre à l'époque. Toutefois, il a été nécessaire de l'adapter, en raison de l'intégration dans le champ de deux nouveaux publics : les

13. Les résultats des appariements entre l'enquête et les données sociales et fiscales devraient être disponibles à la mi-2015.

14. Pour les minima sociaux dits d'insertion, l'objectif du législateur est d'éviter une désincitation au retour à l'emploi en garantissant un revenu trop élevé. Comme les chances de sortir du minimum vieillesse sont quant à elles faibles, le revenu minimum garanti est plus élevé.

bénéficiaires du RSA activité seul d'une part et les allocataires du minimum vieillesse de l'autre. Ces derniers sont au coeur de cet article.

Particulièrement attendus, les résultats de l'enquête sur les allocataires du minimum vieillesse révèlent un public confronté à de nombreuses difficultés. Bien que le revenu garanti par le minimum vieillesse soit plus élevé que ceux garantis par les minima sociaux dits d'insertion (RSA et ASS), près de la moitié de ses bénéficiaires sont pauvres en conditions de vie.

2 Estimation de la variance d'échantillonnage

Dans cette partie, le travail de calcul de précision a été réalisé sans prendre en compte le traitement de la non-réponse et le calage sur marges. Il s'agit ainsi de déterminer un estimateur de la variance d'échantillonnage des estimateurs relatifs aux allocataires du minimum vieillesse relevant du régime général (Cnav), en considérant que l'ensemble des individus faisant partie de l'échantillon ont répondu à l'enquête. Pour résoudre le problème des valeurs manquantes pour les non-répondants à l'enquête, des imputations ont été réalisées.

On utilisera dans la suite de cet article les notations habituelles de la théorie des sondages :

La variable d'intérêt est notée Y et nous utilisons l'estimateur classique d'Horvitz-Thomson pour estimer sans biais son total t_y :

$$t_{y\pi}^{\hat{}} = \sum_{k \in s} \frac{y_k}{\pi_k} = \sum_{k \in s} w_k y_k \quad (1)$$

avec π_k la probabilité d'inclusion d'ordre 1 et w_k le poids d'échantillonnage.

2.1 Estimation de la variance dans un plan de sondage à plusieurs degrés : les formules de Rao et de Raj

Pour un plan de sondage à plusieurs degrés, on obtient un estimateur de variance sans biais en utilisant un système récursif mis au point par DURBIN (1955), amélioré par RAJ (1966), puis par RAO (1975).

Dans le cas d'un tirage à 2 degrés :

Soit une population U partitionnée en N_I unités primaires $\{U_1, U_2, \dots, U_i, \dots, U_{N_I}\} = U_I$.

- 1^{er} degré : un échantillon s de n_s unités primaires est tiré de U_I selon un plan de sondage $p_I(\cdot)$

- 2^e degré : $\forall i \in s$, un échantillon s_i d'unités secondaires est tiré de U_i selon un plan de sondage $p_i(\cdot|s)$

On s'intéresse à l'estimation du total t_y d'une variable Y .

Condition d'indépendance. Les tirages au 2^e degré sont indépendants d'une unité primaire à l'autre.

On suppose que la condition d'indépendance est respectée et que le plan de sondage conduit à des probabilités de sélection de 1^{er} degré π_i et à des estimateurs sans biais \hat{t}_i du total t_i dans l'unité primaire i . On dispose ainsi de l'estimateur d'Horvitz-Thomson $\hat{t}_{y\pi}$ du total t_y de la forme $\hat{t}_{y\pi} = \sum_{i \in s} \frac{\hat{t}_i}{\pi_i}$.

Considérons les estimateurs sans biais suivants :

- $f(\hat{t}_y)$ l'estimateur de la variance relative au 1^{er} degré de sondage en fonction des vrais totaux t_i de chaque unité primaire i de la forme : $f(t_y) = \sum_{i \in s} q_i y_i^2 + \sum_{(i,j) \in s, i \neq j} q_{ij} y_i y_j$
- \hat{V}_i l'estimateur de la variance de \hat{t}_i liée au 2^e degré de sondage au sein de l'unité primaire i

Proposition 1. La **formule de RAO** fournit un estimateur sans biais de $V(\hat{t}_y)$ de la forme :

$$\hat{V}(\hat{t}_y) = f(\hat{t}_y) + \sum_{i \in s} \left(\frac{1}{\pi_i} - q_i \right) \hat{V}_i \quad (2)$$

Condition d'invariance. A chaque fois que l'unité primaire i est tirée au 1^{er} degré, un même plan de sondage est utilisé au 2^e degré, quelles que soient les autres unités primaires sélectionnées. Ainsi, on a :

$$\forall s, \forall i, p_i(\cdot|s) = p_i(\cdot)$$

Proposition 2. Si la condition d'invariance est vérifiée, on peut appliquer la **formule de RAJ** pour déterminer un estimateur sans biais de $V(\hat{t}_y)$:

$$\hat{V}(\hat{t}_y) = f(\hat{t}_y) + \sum_{i \in s} \frac{\hat{V}_i}{\pi_i} \quad (3)$$

Cette formule peut être facilement utilisée dans le cas des plans de sondage à plus de 2 degrés. Il faut considérer le plan de sondage sous forme d'arborescence (Figure 3) et partir du dernier degré, c'est-à-dire celui au cours duquel les individus sont sélectionnés.

On applique alors de manière récursive la formule de RAJ (3) en remontant l'arborescence comme détaillé dans la figure 4. Au final, la dernière agrégation, au degré 1, donne l'estimation de la variance $\hat{V}(\hat{t}_y)$ du total t_y .

Dans le cas du plan de sondage de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux, les conditions d'indépendance et d'invariance sont bien respectées à chaque degré de tirage. Nous avons donc mis en oeuvre le système récursif qui vient d'être décrit pour calculer la précision des estimateurs de l'enquête.

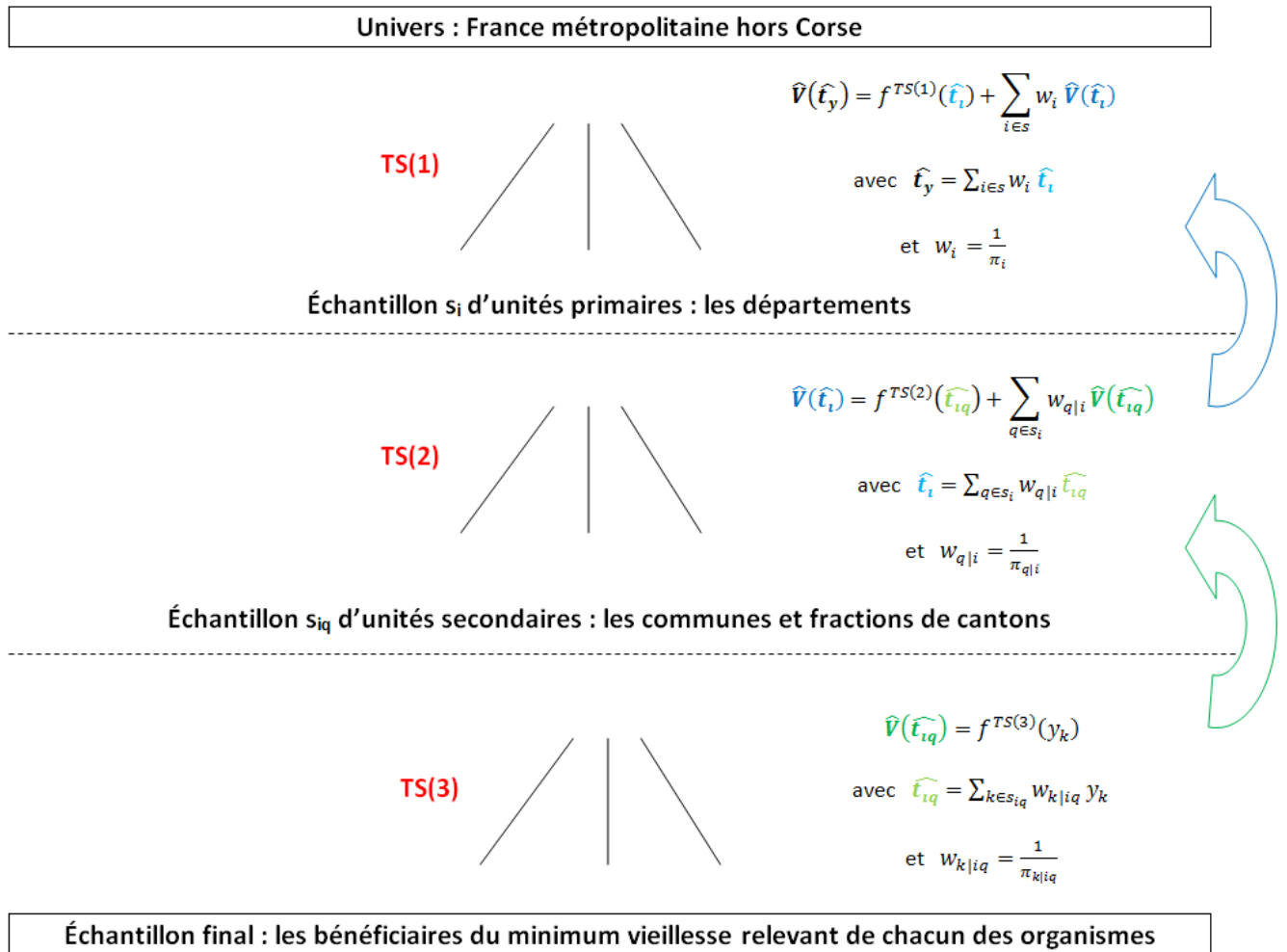


FIGURE 4 – Démarche globale du calcul de la variance d'échantillonnage appliquée à l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux

2.2 Tirage systématique au 3^e degré

Au 3^e degré de tirage, au bout de l'arborescence du plan de sondage de l'enquête, le tirage des individus a été effectué, pour chaque prestation, suivant un sondage systématique dans chacune des unités secondaires iq qui ont été sélectionnées au préalable pour l'ensemble des bénéficiaires de minima sociaux du champ de l'enquête.

Proposition 3. Un sondage systématique est la réalisation d'un sondage par grappe où on ne sélectionne qu'une seule grappe. Il est impossible d'estimer sans biais la variance des estimateurs pour ce type de tirage. Il est néanmoins possible de l'approximer, après s'être assuré que les individus sont bien triés dans le même ordre que celui qui a prévalu au moment de leur tirage.

On utilise la formule suivante, valable pour un tirage systématique à probabilités égales d'un échantillon de taille n parmi une population N [6] :

$$\hat{V}(t_{y\pi}) = N^2 \left(1 - \frac{n}{N}\right) \frac{1}{2n(n-1)} \sum_{k=1}^{n-1} (y_k - y_{k+1})^2 \quad (4)$$

Soit $w_{k|i q}$ le poids de l'individu k qui réside dans la commune ou fraction de canton q du département i . On note $N_{i q}$ le nombre d'allocataires du minimum vieillesse relevant de la Cnav qui compte l'unité secondaire $i q$ et $n_{i q}$ le nombre de ceux qui ont été sélectionnés pour faire partie de l'échantillon. Tous les individus appartenant à la même unité secondaire $i q$ ont la même probabilité de sélection et donc le même poids : $w_{k|i q} = \frac{N_{i q}}{n_{i q}}$.

Après avoir trié le fichier selon l'âge et le sexe des allocataires (comme cela avait été le cas au moment de leur tirage), on obtient les estimateurs suivants :

$$\hat{t}_{i q} = \sum_{k \in s_{i q}} w_{k|i q} y_k \quad (5)$$

$$\begin{aligned} \hat{V}(\hat{t}_{i q}) &= f^{TS(3)}(y_k) \\ &= (w_{k|i q} n_{i q})^2 \left(1 - \frac{1}{w_{k|i q}}\right) \frac{1}{2n_{i q}(n_{i q} - 1)} \sum_{k=1}^{n_{i q}-1} (y_k - y_{k+1})^2 \end{aligned} \quad (6)$$

Un traitement particulier a dû être effectué au niveau des unités secondaires, pour s'assurer qu'au final on ait bien, dans chaque unité géographique, plus d'un allocataire du minimum vieillesse relevant de la Cnav qui fasse partie de l'échantillon (2.4.1). Il faut en effet que $n_{i q}$ soit supérieur à 1 pour que la formule (6) puisse être appliquée.

2.3 Tirages à probabilités inégales aux 1^{er} et 2^e degrés

Après stratification, les unités primaires et secondaires des strates non exhaustives ont été sélectionnées proportionnellement à leur taille en termes de bénéficiaires de minima sociaux.

Proposition 4. Dans le cas des tirages à probabilités inégales, les formules de variance usuelles du type Horvitz-Thomson s'appliquent mal, car elles font intervenir des probabilités d'inclusion d'ordre 2, qui ne sont pas connues et sont difficilement calculables.

Il existe toutefois plusieurs formules approximatives ne faisant pas intervenir les probabilités d'inclusion doubles. Nous utiliserons celle de DEVILLE (1993) [6] :

$$\hat{V}(t_{y\pi}) = \frac{n}{n-1} \sum_{k \in s} (1 - \pi_k) \left(\frac{y_k}{\pi_k} - \frac{\sum_{k \in s} (1 - \pi_k) \frac{y_k}{\pi_k}}{\sum_{k \in s} (1 - \pi_k)} \right)^2 \quad (7)$$

2.3.1 Estimateur de la variance liée au 2^e degré

Soit \hat{t}_i l'estimateur sans biais d'Horvitz-Thomson du total t_i de la variable d'intérêt Y dans l'unité primaire i .

$$\hat{t}_i = \sum_{q \in s_i} w_{q|i} \hat{t}_{iq} \quad (8)$$

$$\text{avec } w_{q|i} = \frac{1}{\pi_{q|i}}$$

L'estimateur sans biais de la variance de \hat{t}_i liée au 2^e degré de sondage s'obtient à partir de la formule de RAJ (3) :

$$\hat{V}(\hat{t}_i) = f^{TS(2)}(\hat{t}_{iq}) + \sum_{q \in s_i} w_{q|i} \hat{V}(\hat{t}_{iq}) \quad (9)$$

Si on exclut le cas de Paris, dans chaque département sélectionné i , une stratification préalable au tirage des unités secondaires a été effectuée entre les 4 plus grosses communes du département en termes de bénéficiaires de minima sociaux, les fractions de cantons urbaines et les fractions de cantons rurales. Etant donné que les tirages sont indépendants d'une strate à une autre, on dispose d'une décomposition de l'estimateur de variance $f^{TS(2)}(\hat{t}_{iq})$ relatif au 2^e degré de sondage que l'on aurait si les vrais totaux t_{iq} de chaque unité secondaire iq étaient connus :

$$f^{TS(2)}(\hat{t}_{iq}) = f_{grandescommunes}(\hat{t}_{iq}) + f_{fractionurbaines}(\hat{t}_{iq}) + f_{fractionsrurales}(\hat{t}_{iq}) \quad (10)$$

avec $f_{grandescommunes}(\hat{t}_{iq}) = 0$ car il s'agit d'une strate exhaustive

$$f_{fractionurbaines}(\hat{t}_{iq}) = \frac{n_{urb|i}}{n_{urb|i} - 1} \sum_{q \in s_{urb|i}} \left(1 - \frac{1}{w_{q|i}} \right) \left(w_{q|i} \hat{t}_{iq} - \frac{\sum_{q \in s_{urb|i}} w_{q|i} (1 - \frac{1}{w_{q|i}}) \hat{t}_{iq}}{\sum_{q \in s_{urb|i}} (1 - \frac{1}{w_{q|i}})} \right)^2 \quad (11)$$

$$f_{fractionsrurales}(\hat{t}_{iq}) = \frac{n_{rur|i}}{n_{rur|i} - 1} \sum_{q \in s_{rur|i}} \left(1 - \frac{1}{w_{q|i}} \right) \left(w_{q|i} \hat{t}_{iq} - \frac{\sum_{q \in s_{rur|i}} w_{q|i} (1 - \frac{1}{w_{q|i}}) \hat{t}_{iq}}{\sum_{q \in s_{rur|i}} (1 - \frac{1}{w_{q|i}})} \right)^2 \quad (12)$$

D'après le document de DEVILLE et VITE SAN-PEDRO (1993), les approximations utilisées pour déterminer les estimations de variance dans les strates urbaines et rurales où des tirages à probabilités inégales ont été effectués, nécessiteraient que la taille des échantillons dans chacune de ces strates dépasse au moins 8. Dans le cas contraire, on s'expose à une sous-estimation de la variance. Or, dans notre plan de sondage, pour le minimum vieillesse, on a $n_{urb|i} + n_{rur|i} = 6$ ¹⁵. Toutefois, pour des tailles d'échantillons aussi petites, il n'existe pas de formule permettant d'estimer de manière satisfaisante la variance d'échantillonnage.

Par ailleurs, la stratification entre les fractions urbaines et les fractions rurales n'a pas été prise en compte dans l'estimation de la variance pour les départements dans lesquels la condition " $n_{urb|i} > 1$ et $n_{rur|i} > 1$ " n'était pas vérifiée (ainsi évidemment que pour les départements très urbains dans lesquels aucune stratification de ce type n'avait été effectuée).

Cas de Paris :

Aucune stratification préalable n'a été réalisée dans ce département. La moitié des arrondissements qui compte la capitale ont été sélectionnés via un tirage à probabilités inégales (proportionnellement à leur taille en termes de bénéficiaires de minima sociaux) :

$$f_{Paris}(\hat{t}_{iq}) = \frac{10}{9} \sum_{q \in S_{Paris}} \left(1 - \frac{1}{w_{q|i}}\right) \left(w_{q|i} \hat{t}_{iq} - \frac{\sum_{q \in S_{Paris}} w_{q|i} \left(1 - \frac{1}{w_{q|i}}\right) \hat{t}_{iq}}{\sum_{q \in S_{Paris}} \left(1 - \frac{1}{w_{q|i}}\right)}\right)^2 \quad (13)$$

2.3.2 Estimateur de la variance liée au 1^{er} degré

Soit \hat{t}_y l'estimateur sans biais d'Horvitz-Thomson du total t_y de la variable d'intérêt Y .

$$\hat{t}_y = \sum_{i \in s} w_i \hat{t}_i \quad (14)$$

$$\text{avec } w_i = \frac{1}{\pi_i}$$

L'estimateur sans biais de la variance de \hat{t}_y liée au 1^{er} degré de sondage s'obtient à nouveau à partir de la formule de RAJ (3) :

$$\hat{V}(\hat{t}_y) = f^{TS(1)}(\hat{t}_i) + \sum_{i \in s} w_i \hat{V}(\hat{t}_i) \quad (15)$$

Une stratification préalable au tirage des unités primaires a également été effectuée. Etant donné que les tirages sont indépendants d'une strate à une autre, on peut à nouveau décomposer l'estimateur de variance $f^{TS(1)}(\hat{t}_i)$ relatif au 1^{er} degré de sondage que l'on aurait si les vrais totaux t_i de chaque unité primaire i étaient connus :

$$f^{TS(1)}(\hat{t}_i) = f_{strate1}(\hat{t}_i) + f_{strate2}(\hat{t}_i) + f_{strate3}(\hat{t}_i) + f_{strate4}(\hat{t}_i) + f_{strate5}(\hat{t}_i) \quad (16)$$

15. Pour les minima sociaux d'âge actif, 4 fractions de cantons supplémentaires avaient été sélectionnées.

avec $f_{strate1}(\hat{t}_i) = f_{strate2}(\hat{t}_i) = 0$ car il s'agit de 2 strates exhaustives

$$f_{strate3}(\hat{t}_i) = f_{strate4}(\hat{t}_i) = f_{strate5}(\hat{t}_i) = \frac{15}{14} \sum_{i \in s} \left(1 - \frac{1}{w_i}\right) \left(w_i \hat{t}_i - \frac{\sum_{i \in s} w_i \left(1 - \frac{1}{w_i}\right) \hat{t}_i}{\sum_{i \in s} \left(1 - \frac{1}{w_i}\right)} \right)^2 \quad (17)$$

Dans les strates 3,4 et 5, le tirage des départements est de taille fixe égale à 15, ce qui garantit le respect de la préconisation de DEVILLE et VITE SAN-PEDRO (1993) vue dans la sous-partie précédente.

2.4 Ajustements

Pour tenir compte de toute la complexité du plan de sondage et de la réalité du tirage de l'échantillon, et afin d'estimer au mieux la variance d'échantillonnage, il a fallu ajuster les formules présentées jusqu'à maintenant.

2.4.1 Regroupement des unités secondaires de taille 1

Pour régler les problèmes liés aux échantillons d'individus de taille 1 dans certaines unités secondaires, ces dernières ont été regroupées avec leur plus proche voisin lorsque cela s'est avéré nécessaire.

On a évidemment essayé de respecter la stratification entre grandes communes, fractions rurales et fractions urbaines.

En regroupant certaines unités secondaires, tous les individus d'une même nouvelle unité secondaire \tilde{q} n'ont donc plus forcément les mêmes poids de sélection de l'unité secondaire. Il convient donc d'adapter la formule d'estimation de la variance en conséquence.

La formule permettant d'estimer la variance liée au 2^e degré de tirage (9) peut être réécrite de la façon suivante :

$$\hat{V}(\hat{t}_i) = f^{TS(2)}(\hat{t}_{iq}) + \sum_{q \in s_i} \hat{V}(\sqrt{w_{q|i}} \hat{t}_{iq}) \quad (18)$$

Soit \tilde{q} une nouvelle unité secondaire qui regroupe α unités q .

On considère les notations suivantes :

- $n_{i\tilde{q}} = \sum_{q=1}^{\alpha} n_{iq}$
- $N_{i\tilde{q}} = \sum_{q=1}^{\alpha} n_{iq} w_{q|i}$
- $(w_{q|i})_k$ le poids initial de l'unité secondaire q qui a été intégrée dans l'unité \tilde{q} et qui contenait l'individu k

L'estimation de la variance liée au 3^e degré de tirage (6) est donc désormais de la forme :

$$\hat{V}(\sqrt{w_{\hat{q}|i}} t_i \hat{q}) = (N_{i\hat{q}})^2 \left(1 - \frac{n_{i\hat{q}}}{N_{i\hat{q}}}\right) \frac{1}{2n_{i\hat{q}}(n_{i\hat{q}} - 1)} \sum_{k=1}^{n_{i\hat{q}}-1} \left(\sqrt{(w_{\hat{q}|i})_k} y_k - \sqrt{(w_{\hat{q}|i})_{k+1}} y_{k+1}\right)^2 \quad (19)$$

2.4.2 Phase de tirage supplémentaire des unités secondaires dans le cas où le minimum vieillesse est un critère rare dans une unité primaire

Dans les unités primaires où le minimum vieillesse relevant de la Cnav est un critère rare, le 2^e degré de tirage dans les strates urbaine et rurale est en réalité composé de 2 phases : les unités secondaires finales ont été tirées au sein d'un échantillon d'unités secondaires préalablement sélectionnées (1.3.3).

Etant donné la complexité de la prise en compte de cette 2^e phase de sondage, il a été décidé de procéder à des réplifications d'échantillons d'unités secondaires dans les strates urbaine et rurale de ces départements. L'objectif est de déterminer la probabilité de sélection de ces fractions de canton. L'idée est en effet d'assimiler ce tirage en 2 phases à un tirage en 1 phase avec des probabilités de tirage égales aux probabilités estimées par réplification.

		100 000 tirages		300 000 tirages	
		Différence maximale entre les probabilités d'inclusion estimées sur la 1ère et la 2nde moitié des tirages simulés	Probabilité d'inclusion de la fraction de canton associée (calculée sur l'ensemble des tirages simulés)	Différence maximale entre les probabilités d'inclusion estimées sur la 1ère et la 2nde moitié des tirages simulés	Probabilité d'inclusion de la fraction de canton associée (calculée sur l'ensemble des tirages simulés)
Département 18	Rural	1,2%	7,0%	0,3%	4,8%
	Urbain	2,5%	37,2%	0,4%	37,6%
Département 28	Rural	-	-	-	-
	Urbain	2,9%	14,0%	0,6%	26,6%
Département 55	Rural	0,6%	11,7%	0,4%	11,9%
	Urbain	3,6%	25,2%	0,4%	16,3%
Département 61	Rural	1,5%	9,2%	0,7%	9,3%
	Urbain	2,7%	18,6%	0,8%	20,1%
Département 70	Rural	1,2%	21,6%	0,4%	21,3%
	Urbain	3,7%	34,9%	0,2%	38,9%
Département 89	Rural	2,0%	6,7%	0,7%	6,4%
	Urbain	3,4%	32,5%	0,3%	12,7%

TABLE 3 – Appréciation de la convergence de la probabilité empirique de sélection d'une fraction de canton dans le cas où le minimum vieillesse relevant de la Cnav est un critère rare

Au total, 300 000 tirages ont été simulés pour obtenir une estimation suffisamment précise. Cette précision est évaluée au regard de la différence maximale de probabilités d'inclusion qui existe entre une estimation obtenue sur la première moitié des tirages simulés et une autre obtenue sur la seconde moitié (Table 3). La probabilité d'inclusion d'une fraction de canton est calculée comme le rapport du nombre d'échantillons dans lesquels elle apparaît

par le nombre de tirages simulés au total.

Si le minimum vieillesse relevant de la Cnav est un critère rare dans le département i , alors, dans la formule d'estimation de la variance liée au 2^e degré (9), le poids $w_{q|i}$ de la fraction de canton q est égal à l'inverse de la probabilité de sélection empirique que l'on vient de calculer.

2.5 Résultats

On a considéré jusqu'à présent qu'il y avait absence de non-réponse totale. Les formules de calcul de variance ont ainsi été appliquées à l'ensemble des individus de l'échantillon s . L'estimation de la variance d'échantillonnage des estimateurs de variables d'intérêt de l'enquête a donc nécessité d'imputer des réponses aux non-répondants à l'enquête.

Les calculs de précision ont également porté sur une variable quantitative de la base de sondage¹⁶. Pour son estimateur, aucun effet parasite lié aux imputations effectuées pour pallier la non-réponse totale ne vient perturber la mesure de sa qualité.

Proposition 5. Lorsque le paramètre d'intérêt est une fonction non linéaire de totaux, il n'existe pas de formule générale pour calculer la variance de son estimateur. Il est cependant possible d'approcher cette variance par la variance de l'estimateur du total de sa variable linéarisée, obtenue grâce à la formule de Taylor de développement en série.

Proposition 6. Dans le cas où la taille totale de la population N est inconnue, la moyenne μ_y de la variable Y est un cas particulier de ratio de cette variable par la variable constante égale à 1. Pour en estimer la variance, on se ramène à l'estimation de la variance de sa variable linéarisée, qui est de la forme :

$$\hat{u}_i = \frac{1}{\hat{t}_N} (y_i - \hat{Y}) \quad (20)$$

$$\text{avec } \hat{Y} = \frac{\hat{t}_y}{\hat{t}_N} \text{ et } \hat{t}_N = \sum_{i \in s} \frac{1}{\pi_i}.$$

Une proportion est simplement la moyenne d'une variable dichotomique.

Le montant mensuel moyen de minimum vieillesse perçu :

On s'intéresse dans un premier temps à la moyenne du *montant mensuel de minimum vieillesse perçu*, variable issue de la base de sondage.

16. Les seules variables quantitatives à disposition au moment de la rédaction de cet article sont celles de la base de sondage. En effet, les variables quantitatives de l'enquête ne seront diffusées qu'une fois que l'appariement avec les données sociales et fiscales aura été réalisé, au 1^{er} semestre 2015.

Pour estimer sa variance et les autres indicateurs d'erreur d'échantillonnage, il convient de se placer dans le cas où la taille de la population N est inconnue. En effet, N correspond à la taille de la population du champ de l'enquête, c'est-à-dire celle dont on a exclu les bénéficiaires vivant dans des institutions les prenant presque entièrement en charge. La proportion de hors-champs n'étant pas connue, on ne peut utiliser qu'une estimation de N à partir des résultats de l'enquête pour nos calculs de précision.

Les résultats des calculs de précision obtenus pour cette variable sont présentés dans le tableau 4.

$\widehat{\mu}_y$	$\widehat{V}(\widehat{\mu}_y)$	$\widehat{V}_{SAS}(\widehat{\mu}_{SAS})$	\widehat{CV}	\widehat{Inf}_{IC}	\widehat{Sup}_{IC}
316.903	14.016	10.319	0.0118	309.566	324.241

TABLE 4 – Précision de l'estimateur de montant mensuel moyen de minimum vieillesse perçu

L'estimateur du coefficient de variation (\widehat{CV}) d'un estimateur est un bon indicateur permettant d'évaluer l'erreur d'échantillonnage. Dans la littérature [5], on considère qu'un "bon" estimateur a un coefficient de variation inférieur à 5%. Cet indicateur est généralement utilisé pour comparer différents estimateurs entre eux. Il est défini de la façon suivante :

$$\widehat{CV} = \frac{\sqrt{\widehat{V}(\widehat{\mu}_y)}}{\widehat{\mu}_y} \quad (21)$$

Le coefficient de variation estimé du *montant mensuel moyen de minimum vieillesse perçu* se révèle très bon : il n'est que de 1,2%.

Par ailleurs, \widehat{Inf}_{IC} et \widehat{Sup}_{IC} sont les estimateurs des bornes inférieures et supérieures de l'intervalle de confiance d'un estimateur à 95%. Celui-ci est de la forme :

$$\widehat{IC} = \left(\widehat{\mu}_y - 1,96\sqrt{\widehat{V}(\widehat{\mu}_y)} ; \widehat{\mu}_y + 1,96\sqrt{\widehat{V}(\widehat{\mu}_y)} \right) \quad (22)$$

Les variables d'intérêt de l'enquête :

On s'intéresse maintenant aux variables d'intérêt de l'enquête. Les résultats des calculs de précision menés jusque là pour ces variables sont à prendre avec précaution, car ils cachent un effet "imputation" lié au traitement qui a été réalisé pour les non-répondants à l'enquête.

On a étudié dans un premier temps la précision du *nombre total d'allocataires du minimum vieillesse relevant de la Cnav pauvres en conditions de vie*. On constate que son coefficient de variation est plus élevé que celui du *montant mensuel de minimum vieillesse*

perçu, mais il reste correct : 3,7%. (Table 5).

\hat{t}_y	$V(\hat{t}_y)$	$V_{SAS}(\hat{t}_{SAS})$	\widehat{CV}	\widehat{Inf}_{IC}	\widehat{Sup}_{IC}
158 294.84	34 038 891.71	8 752 497.41	0.0369	146 859.64	169 730.04

TABLE 5 – Précision de l'estimateur du nombre total d'allocataires du minimum vieillesse relevant de la Cnav pauvres en conditions de vie

On s'est également intéressé à la *proportion d'allocataires disposant d'un patrimoine immobilier*. Cette variable possède le moins bon coefficient de variation des trois variables étudiées jusque là : 4,95%. Celui-ci est tout de même encore inférieur à 5% (Table 6).

\hat{p}_y	$V(\hat{p}_y)$	$V_{SAS}(\hat{t}_{SAS})$	\widehat{CV}	\widehat{Inf}_{IC}	\widehat{Sup}_{IC}
0.1474	0.000 053	0.000 044	0.0495	0.1331	0.1616

TABLE 6 – Précision de l'estimateur de la proportion d'allocataires disposant d'un patrimoine immobilier

L'estimation de l'effet de sondage :

Un sondage complexe, à plusieurs degrés, a de bonnes propriétés pratiques (en particulier une concentration des individus dans des zones géographiques contrôlées), mais il est en revanche forcément moins précis qu'un sondage aléatoire simple de même taille. Il est intéressant de quantifier cette perte de précision.

Le rapport entre l'estimation de la variance obtenue et celle que l'on aurait obtenue en supposant un plan de sondage aléatoire simple est appelé *l'effet de sondage* (noté $Deff$). Cette estimation permet d'apprécier l'efficacité du plan de sondage mis en oeuvre.

Soit \hat{t}_{SAS} l'estimateur du total t_y d'une variable Y pour un plan de sondage aléatoire simple sans remise de taille fixe égale à celle de l'échantillon et $V_{SAS}(\hat{t}_{SAS})$ un estimateur approximativement sans biais de sa variance.

On a donc :

$$Deff = \frac{\hat{V}(\hat{t}_y)}{V_{SAS}(\hat{t}_{SAS})} \quad (23)$$

Sous un plan de sondage complexe, $V_{SAS}(\hat{t}_{SAS})$ diffère de l'estimateur traditionnel obtenu pour un sondage aléatoire simple. La formule retenue est de la forme [6] :

$$V_{SAS}(\hat{t}_{SAS}) = \frac{1}{n} \left(1 - \frac{n-1}{\hat{N}-1} \right) \left(\hat{N} \sum_{k \in s} w_k y_k^2 - \left(\sum_{k \in s} w_k y_k \right)^2 + \hat{V}(\hat{t}_y) \right) \quad (24)$$

En reprenant les notations utilisées jusque là, on a $w_k = w_{k|i_q} * w_{q|i} * w_i$. Et pour estimer la taille de la population N , on applique la formule classique : $\hat{N} = \sum_s w_k$.

Quelle que soit la variable considérée, l'estimation de l'effet de sondage associé est supérieure à 1. Cela traduit bien le fait que la mise en oeuvre du plan de sondage complexe de l'enquête engendre un accroissement de l'erreur de sondage par rapport à un sondage aléatoire simple. Néanmoins, la perte de précision pour ces paramètres n'est pas très grande (Table 7).

L'effet de sondage associé au *montant mensuel moyen de minimum vieillesse perçu* est estimé à 1,4. Il y a donc très peu de perte de précision pour ce paramètre par rapport à un sondage aléatoire simple de même taille. L'estimation de l'effet de sondage de la *proportion d'allocataires disposant d'un patrimoine immobilier* est du même ordre de grandeur.

L'estimation de l'effet de sondage associé au *nombre total d'allocataires pauvres en conditions de vie* est ici bien plus élevée : 3,9. L'estimation de ce total a donc une variance d'échantillonnage près de 4 fois plus grande que celle de l'estimateur obtenu en considérant un sondage aléatoire simple de même taille.

Variable d'intérêt	\widehat{Deff}
Montant mensuel moyen de minimum vieillesse perçu	1.358
Total d'allocataires pauvres en conditions de vie	3.889
Proportion d'allocataires disposant d'un patrimoine immobilier	1.205

TABLE 7 – Estimation de l'effet de sondage

Conclusion

Les calculs menés dans cette deuxième partie conduisent à des premières estimations de la précision des principaux estimateurs de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux. Ces travaux ont été réalisés pour les seuls allocataires du minimum vieillesse relevant de la Cnav.

Les formules mises en oeuvre tiennent compte de nombreux éléments :

- le plan de sondage à 3 degrés de l'enquête : tirage des départements, puis des communes ou fractions de cantons, et enfin des individus
- le tirage des allocataires au 3^e degré par sondage systématique ordonné suivant leur âge et leur sexe
- le tirage des unités géographiques à probabilités inégales aux 1^{er} et 2^e degrés
- le regroupement des unités secondaires dans le cas des échantillons de taille 1
- la 2^e phase de tirage des unités secondaires dans certains départements (cas où le minimum vieillesse est un critère rare)

Dans cette partie, la pondération utilisée correspond au poids de sélection initial des allocataires (poids avant traitement de la non-réponse et calage sur marges).

D'après les résultats obtenus, l'enquête semble être de bonne qualité. Les coefficients de variation estimés sont en effet relativement faibles (toujours inférieurs à 5%), tout comme les estimations de l'effet de sondage associées.

3 Prise en compte de la non-réponse totale et du calage dans le calcul de la précision

Comme pour toute enquête, l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux est confrontée au problème de la non-réponse, qu'elle soit totale ou partielle. Seule la non-réponse totale a été prise en compte pour l'évaluation de la variance de nos estimateurs. En effet, comme les variables d'intérêt étudiées devaient obligatoirement être renseignées pour pouvoir passer à la suite du questionnaire, nous n'avons pas été confronté à la non-réponse partielle.

Le taux de réponse à l'enquête des bénéficiaires de minima sociaux d'âge actif est de 50% en moyenne. Il est plus faible pour les allocataires du minimum vieillesse, notamment parce que les personnes ne pouvant être enquêtées pour des raisons de santé ou car elles sont entièrement prises en charge par une institution sont plus nombreuses. Le taux de réponse des allocataires du minimum vieillesse est ainsi de 33%.

Pour compenser cette non-réponse totale, une méthode de repondération utilisant des groupes de réponse homogène a été mise en oeuvre, prestation par prestation. Le poids d'échantillonnage des répondants a ainsi été augmenté. Dans une première sous-partie, on s'attache à intégrer ce traitement de la non-réponse totale dans le calcul de l'estimateur de la variance obtenu jusqu'à présent.

Par ailleurs, l'enquête a également été redressée, prestation par prestation (et organisme par organisme dans le cas du minimum vieillesse), sur un certain nombre de marges de calage. L'intégration de cette information auxiliaire dans les estimateurs permet de respecter les totaux connus sur l'ensemble de la population et de diminuer la variance des estimateurs obtenue après calage pour les variables d'intérêt corrélées aux variables de calage. La deuxième sous-partie traite de ce redressement et de son impact sur notre calcul de précision.

3.1 Modélisation de la non-réponse totale et calcul de précision

La non-réponse totale peut être modélisée par une phase de sondage supplémentaire :

- 1^{re} phase : tirage de l'échantillon s selon notre plan de sondage avec des probabilités de tirage $\pi_k = \pi_{k|i q} * \pi_{q|i} * \pi_i$
- 2^e phase : tirage de l'échantillon des répondants r , au sein de l'échantillon s préalablement sélectionné, selon un plan de sondage poissonnien où chaque unité répondante

k est sélectionnée de façon indépendante dans s selon sa probabilité de réponse $p_{rép}$

Si la probabilité de réponse $p_{rép}$ était connue pour chaque individu k , on pourrait estimer sans biais le total de n'importe quelle variable Y à l'aide de l'estimateur en expansion¹⁷ suivant :

$$\hat{t}_{y,CNR} = \sum_{k \in r} \frac{1}{\pi_k p_{k,rép}} y_k \quad (25)$$

Malheureusement, les probabilités de réponses $p_{k,rép}$ ne sont pas connues. Il a donc fallu les estimer. Pour ce faire, nous avons procédé à une modélisation du mécanisme de non-réponse prestation par prestation, en analysant le critère "répond / ne répond pas" en fonction de variables connues pour les répondants et les non-répondants à l'aide d'une régression logistique. Nous avons ainsi déterminé les variables auxiliaires les plus explicatives de la non-réponse pour les allocataires du minimum vieillesse, à savoir :

- le montant mensuel de minimum vieillesse perçu (en 3 tranches)
- le type d'allocation du minimum vieillesse perçu (ASPA ou ASV¹⁸)
- être en couple ou isolé

Les groupes de réponse homogène (GRH) ont alors été définis par croisement de toutes les modalités de ces trois variables explicatives. Les 12 GRH ainsi obtenus contenaient suffisamment d'allocataires du minimum vieillesse pour assurer une certaine stabilité des estimations de probabilités de réponse : la probabilité de réponse est considérée constante au sein d'un même GRH. De plus, cette probabilité de réponse est supposée indépendante d'un GRH à l'autre.

La probabilité de réponse d'un individu k donné appartenant au GRH g est alors estimée par le taux de réponse pondéré¹⁹ observé dans ce GRH :

$$\hat{p}_{k,rép} = \frac{\sum_{k \in r} w_k \mathbb{I}_{k \in GRH_g}}{\sum_{k \in (r+nr)} w_k \mathbb{I}_{k \in GRH_g}} \quad (26)$$

Au final, comme le fait de répondre à l'enquête est indépendant entre les individus, conditionnellement au fait que ces individus appartiennent bien à l'échantillon de l'enquête, on peut voir la phase de non-réponse comme un degré de sondage supplémentaire : au sein de chaque individu k de l'échantillon, un tirage bernoullien de 0 ou 1 unité est effectué selon une probabilité de tirage correspondant à sa probabilité de réponse estimée $\hat{p}_{k,rép}$.

Les tirages sont indépendants entre les individus. En revanche, la condition d'invariance n'est pas vérifiée, puisque les $\hat{p}_{k,rép}$ dépendent de l'échantillon s . On applique donc la formule de RAO (2) pour prendre en compte, dans l'estimation de la précision, ce degré de sondage supplémentaire par rapport au plan de sondage initial sans traitement de la

17. Estimateur classique dans le cadre des sondages en plusieurs phases.

18. En 2007, l'allocation de solidarité aux personnes âgées (ASPA) se substitue à l'allocation supplémentaire vieillesse (ASV) pour les nouveaux entrants.

19. Par les poids de sondage initiaux $w_k = \frac{1}{\pi_k}$.

non-réponse.

Soit f la forme quadratique permettant d'estimer la variance de \hat{t}_y en l'absence de non-réponse. La variance avec prise en compte de la non-réponse s'estime donc par :

$$\hat{V}(\hat{t}_{y,CNR}) = f\left(\frac{t_y \Pi_r}{\hat{p}_{rép}}\right) + \sum_{k \in s} \left(\left(\frac{1}{\pi_k}\right)^2 - q_k \right) (1 - \hat{p}_{k,rép}) \left(\frac{y_k \Pi_{i \in r}}{\hat{p}_{k,rép}} \right)^2 \quad (27)$$

avec q_k le coefficient associé aux termes diagonaux de la forme quadratique f

Identification du coefficient $q_{k|i q}$ associé à la forme quadratique permettant d'estimer la variance relative au 3^e degré :

Après avoir trié le fichier des répondants à l'enquête dans le même ordre que celui qui prévalait au moment du tirage systématique des allocataires du minimum vieillesse relevant de la Cnav, la formule (6) donne les coefficient $q_{k|i q}$ suivants :

- pour le premier et le dernier individu de chaque unité secondaire $i q$:

$$q_{k|i q} = (w_{k|i q} n_{i q})^2 \left(1 - \frac{1}{w_{k|i q}}\right) \left(\frac{1}{2n_{i q}(n_{i q} - 1)}\right) \quad (28)$$

- pour tous les autres individus de l'unité $i q$:

$$q_{k|i q} = 2 * (w_{k|i q} n_{i q})^2 \left(1 - \frac{1}{w_{k|i q}}\right) \left(\frac{1}{2n_{i q}(n_{i q} - 1)}\right) \quad (29)$$

Identification du coefficient $q_{q|i}$ associé à la forme quadratique permettant d'estimer la variance relative au 2^e degré :

Dans le cadre de la formule de DEVILLE (2.3.1), les coefficient $q_{q|i}$ s'obtiennent de la façon suivante [8] :

$$q_{q|i} = \frac{n}{n-1} w_{q|i}^2 \left(1 - \frac{1}{w_{q|i}}\right) \left(1 - \frac{1 - \left(\frac{1}{w_{q|i}}\right)}{\sum_{q \in s_i} \left(1 - \frac{1}{w_{q|i}}\right)}\right) \quad (30)$$

avec :

- $n = n_{urb|i}$ si la fraction de canton q du département i appartient à la strate urbaine
- $n = n_{rur|i}$ si elle appartient à la strate rurale
- $n = 6$ si la condition " $n_{urb|i} > 1$ et $n_{rur|i} > 1$ " n'est pas vérifiée dans le département i ou si aucune stratification entre les fractions urbaines et les fractions rurales n'avait été effectuée dans ce département
- $n = 10$ dans le cas de Paris

Identification du coefficient q_i associé à la forme quadratique permettant d'estimer la variance relative au 1^{er} degré :

Comme dans le cas du 2^e degré et d'après la formule (17), les coefficient q_i sont de la forme :

$$q_i = \frac{15}{14} w_i^2 \left(1 - \frac{1}{w_i}\right) \left(1 - \frac{1 - \left(\frac{1}{w_i}\right)}{\sum_{i \in s} \left(1 - \frac{1}{w_i}\right)}\right) \quad (31)$$

Au final :

Dans le cas d'un plan de sondage à 2 degrés, si on note :

- $f_1(\hat{t}_i)$ l'estimateur de la variance relative au 1^{er} degré de sondage en fonction des vrais totaux t_i de chaque unité primaire i , tel que :

$$f_1(\hat{t}_i) = \sum_{i \in s} q_{1,i} \hat{t}_i^2 + \sum_{(i,j) \in s, i \neq j} q_{1,ij} \hat{t}_i \hat{t}_j$$

- $f_2(y)$ l'estimateur de la variance relative au 2^e degré de sondage, tel que :

$$f_2(y) = \sum_{k \in s_i} q_{2,k|i} y_k^2 + \sum_{(k,l) \in s_i, k \neq l} q_{2,kl|i} y_k y_l$$

Alors le coefficient q_k associé à l'estimateur de la variance totale $\hat{V}(\hat{t}_y)$ se calcule à partir des coefficients $q_{1,i}$ et $q_{2,k|i}$ via la formule de RAJ (3) de la façon suivante :

$$q_k = w_{k|i}^2 q_{1,i} + w_i q_{2,k|i} \quad (32)$$

Dans le cas de notre plan de sondage à 3 degrés, il convient d'itérer la formule de RAJ pour déterminer le coefficient q_k de la formule (27). On obtient ainsi :

$$q_k = (w_{q|i} w_{k|i q})^2 q_i + (w_i w_{k|i q}^2) q_{q|i} + (w_i w_{q|i}) q_{k|i q} \quad (33)$$

3.2 Précision obtenue en tenant compte de la correction de la non-réponse

En comparant la précision des estimateurs obtenus dans cette première sous-partie avec ceux obtenus dans la partie précédente, on mesure l'effet de la repondération liée à la correction de la non-réponse totale. Pour se cantonner à cet effet, il convient de travailler sur une variable de la base de sondage, de telle sorte qu'on ne prenne pas en compte l'effet lié à l'imputation qui a été réalisée dans la partie précédente pour les non-répondants à l'enquête. On s'intéresse ainsi à nouveau au *montant mensuel moyen de minimum vieillesse perçu*.

Par rapport à la partie précédente où on considérait qu'il y avait absence de non-réponse totale, la variance estimée de cet estimateur est 3 fois plus élevée (Table 8). Ce résultat est tout à fait cohérent avec le taux de réponse d'un tiers observé parmi les allocataires du minimum vieillesse²⁰.

20. Avec un taux de réponse d'un tiers, la taille de l'échantillon exploitable est divisée par 3. Or, en 1^{ère} approximation, l'évolution de la variance est inversement proportionnelle à la taille de l'échantillon. Il est donc logique d'observer une multiplication de la variance par un facteur proche de 3 avec un tel

Dans le même temps, l'estimateur de son coefficient de variation est logiquement moins bon : 2% contre 1,2%. De la même manière, l'estimation de son intervalle de confiance est plus large.

$\widehat{\mu}_y$	$V(\widehat{\mu}_y)$	$V_{SAS}(\widehat{\mu}_{SAS})$	\widehat{CV}	\widehat{Inf}_{IC}	\widehat{Sup}_{IC}	\widehat{Def}
319.291	41.874	31.182	0.0203	306.608	331.974	1.343

TABLE 8 – Précision de l'estimateur de montant mensuel moyen de minimum vieillesse perçu et effet de sondage associé dans le cas de la prise en compte du traitement de la non-réponse

Il est à nouveau intéressant d'estimer à ce stade l'effet de sondage (\widehat{Def}) qui lui est associé.

Par rapport à la formule (24), l'estimateur sans biais de $V_{SAS}(t_{SAS})$ doit être ajusté à la marge pour tenir compte du fait qu'on se restreint désormais aux seuls répondants à l'enquête :

$$V_{SAS}(\hat{t}_{SAS,CNR}) = \frac{1}{r} \left(1 - \frac{r-1}{\hat{N}-1}\right) \left(\hat{N} \sum_{k \in r} w_{k,CNR} y_k^2 - \left(\sum_{k \in r} w_{k,CNR} y_k \right)^2 + \hat{V}(\hat{t}_{y,CNR}) \right) \quad (34)$$

En reprenant les notations utilisées jusque là, on obtient le poids d'un individu k répondant à l'enquête de la manière suivante : $w_{k,CNR} = \frac{w_k}{\hat{p}_{rép}}$. Et pour estimer la taille de la population N , on applique la formule : $\hat{N} = \sum_r w_{k,CNR}$.

L'effet de sondage associé au *montant mensuel moyen de minimum vieillesse perçu* est désormais estimé à 1,3. Il est ainsi légèrement plus faible que celui obtenu dans la partie précédente. En y regardant de plus près, on constate que les estimations de variance $\hat{V}(\hat{\mu}_y)$ et $V_{SAS}(\hat{\mu}_{SAS})$ augmentent bien toutes les deux lorsqu'on tient compte de la présence de non-réponse. Ce qui fait diminuer l'effet de sondage à ce stade, c'est que cette augmentation de variance est proportionnellement un peu moins forte quand on considère une modélisation de la non-réponse avec des groupes homogènes de réponse, comme c'est le cas avec notre véritable plan de sondage, que quand on traite la non-réponse de manière globalement uniforme en passant d'un sondage aléatoire simple de taille n à un sondage aléatoire simple de taille r .

3.3 Estimation finale de la variance après prise en compte du calage

Le calage sur marges permet de redresser un échantillon, par repondération des individus, en utilisant une information auxiliaire disponible sur un certain nombre de variables. Les

taux de réponse.

pondérations produites par la méthode assurent le calage de l'échantillon sur des totaux connus sur l'ensemble de la population.

Dans notre enquête, le calage sur marges a été réalisé prestation par prestation, voire organisme par organisme dans le cas du minimum vieillesse. Les résultats de l'enquête obtenus auprès des allocataires du minimum vieillesse relevant de la Cnav ont été redressés sur un total de neuf variables :

- l'âge (en 6 tranches)
- le sexe
- être en couple ou isolé
- être né en France ou à l'étranger
- le type d'allocation du minimum vieillesse perçu
- être au minimum vieillesse depuis moins de 10 ans ou depuis 10 ans ou plus
- le montant mensuel de minimum vieillesse perçu (en 3 tranches)
- le type de droit perçu (droit propre, droit dérivé ou les 2)
- la strate de département à laquelle appartient la commune de résidence

D'après DEVILLE et SÄRNDAL [7], pour prendre en compte ce calage dans notre calcul de précision, il suffit d'approximer notre estimateur de variance de la façon suivante :

$$\hat{V}(\hat{t}_{y,calé}) \approx \hat{V}(\hat{t}_{\varepsilon,CNR}) \quad (35)$$

ε est la variable de résidu de la régression de Y sur les variables de calage X pondérée par le poids de non-réponse $w_{k,CNR}$. Cette régression s'effectue sur les répondants à l'enquête et les hors-champs, comme cela avait été le cas lors du calage sur marges. On a ainsi la formule suivante :

$$\varepsilon = Y - X\beta \quad (36)$$

avec Y le vecteur des y_k

X la matrice des variables de calage

et $\beta = (X'DX)^{-1}X'DY$ avec D la matrice diagonale des poids de non-réponse

Au final, on a donc :

$$\hat{V}(\hat{t}_{y,calé}) = f\left(\frac{t_{\varepsilon}\mathbb{I}_r}{\hat{p}_{rép}}\right) + \sum_{k \in s} \left(\left(\frac{1}{\pi_k} \right)^2 - q_k \right) (1 - \hat{p}_{k,rép}) \left(\frac{\varepsilon_k \mathbb{I}_{i \in r}}{\hat{p}_{k,rép}} \right)^2 \quad (37)$$

3.4 Précision des principaux estimateurs en tenant compte du calage

Au final, en plus de la variable de *montant mensuel moyen de minimum vieillesse perçu* issue de la base de sondage, on s'est intéressé à 4 variables d'intérêt de l'enquête :

- le nombre total d'allocataires pauvres en conditions de vie

- la proportion d’allocataires qui ont connu de longues périodes d’activité ou qui n’ont jamais travaillé
- la proportion d’allocataires qui se déclarent en mauvaise santé
- la proportion d’allocataires disposant d’un patrimoine immobilier

Pour mesurer à quel point le redressement effectué par calage a amélioré la précision des estimateurs de nos différentes variables, on peut calculer, pour chacune d’entre elles, *l’effet de calage*. Cet indicateur est défini comme le rapport entre la variance après calage et la variance avant calage, obtenue après prise en compte de la non-réponse.

Pour toutes les variables étudiées, l’effet de calage est inférieur à 1 (Table 9) : le fait d’avoir redressé l’échantillon a augmenté la précision de l’ensemble des variables étudiées.

Variable d’intérêt	Estimation de l’effet de calage
Montant mensuel moyen de minimum vieillesse perçu	0.17
Total d’allocataires pauvres en conditions de vie	0.41
Proportion d’allocataires ayant connu de longues périodes d’activité ou n’ayant jamais travaillé	0.73
Proportion d’allocataires qui se déclarent en mauvaise santé	0.85
Proportion d’allocataires disposant d’un patrimoine immobilier	0.80

TABLE 9 – Estimation de l’effet de calage des principales variables d’intérêt

Plus l’effet de calage est grand, moins le gain de précision dû au calage est important. L’effet de calage est ainsi maximal pour le *montant mensuel moyen de minimum vieillesse perçu*. Ce résultat était attendu, car cette variable, une fois discrétisée en 3 modalités, fait partie des variables sur lesquelles les résultats de l’enquête sont calés. Avec un effet de calage estimé à 0,41, le redressement a permis d’estimer le nombre *total d’allocataires du minimum vieillesse relevant de la Cnav pauvres en conditions de vie* avec une précision égale à celle que l’on aurait eu sans redressement sur un échantillon de taille 2,4 fois plus grande. Il s’agit de la variable d’intérêt étudiée pour laquelle le calage a été le plus bénéfique. Pour les autres, l’effet de calage est compris entre 0,7 et 0,9.

Par ailleurs, outre les indicateurs habituels d’erreur d’échantillonnage que nous avons étudiés (variance, coefficient de variation, intervalle de confiance), l’estimation finale de l’effet de sondage ($D\hat{e}ff$) peut à nouveau être analysée, en considérant cette fois le rapport entre $\hat{V}(\hat{t}_{y,calé})$ et $V_{SAS}(\hat{t}_{SAS,calé})$, avec comme écriture pour cette dernière :

$$V_{SAS}(\hat{t}_{SAS,calé}) = \frac{1}{r} \left(1 - \frac{r-1}{\hat{N}-1} \right) \left(\hat{N} \sum_{k \in r} w_{k,calé} \varepsilon_k^2 - \left(\sum_{k \in r} w_{k,calé} \varepsilon_k \right)^2 + \hat{V}(\hat{t}_{y,calé}) \right) \quad (38)$$

avec $w_{k,calé} = p_{final}$ et $\hat{N} = \sum_r w_{k,calé}$.

Le montant mensuel moyen de minimum vieillesse perçu :

Dans un premier temps, on s'intéresse à la précision de l'estimateur de notre variable de *montant mensuel moyen de minimum vieillesse perçu*, renseignée pour tous les individus de la base de sondage.

L'estimateur de son coefficient de variation est désormais de 0,8% (Table 10). Avec le calage sur marges, on aboutit donc au meilleur estimateur de cette variable. Son intervalle de confiance est également le plus resserré.

$\hat{\mu}_y$	$V(\hat{\mu}_y)$	$V_{SAS}(\hat{\mu}_{SAS})$	\widehat{CV}	\widehat{Inf}_{IC}	\widehat{Sup}_{IC}	\widehat{Deff}
316.868	7.108	6.508	0.0084	311.642	322.094	1.092

TABLE 10 – Précision finale de l'estimateur de montant mensuel moyen de minimum vieillesse perçu et effet de sondage associé après prise en compte du calage

Le graphique (5) récapitule les estimateurs obtenus pour cette variable lors des différentes étapes du calcul de précision.

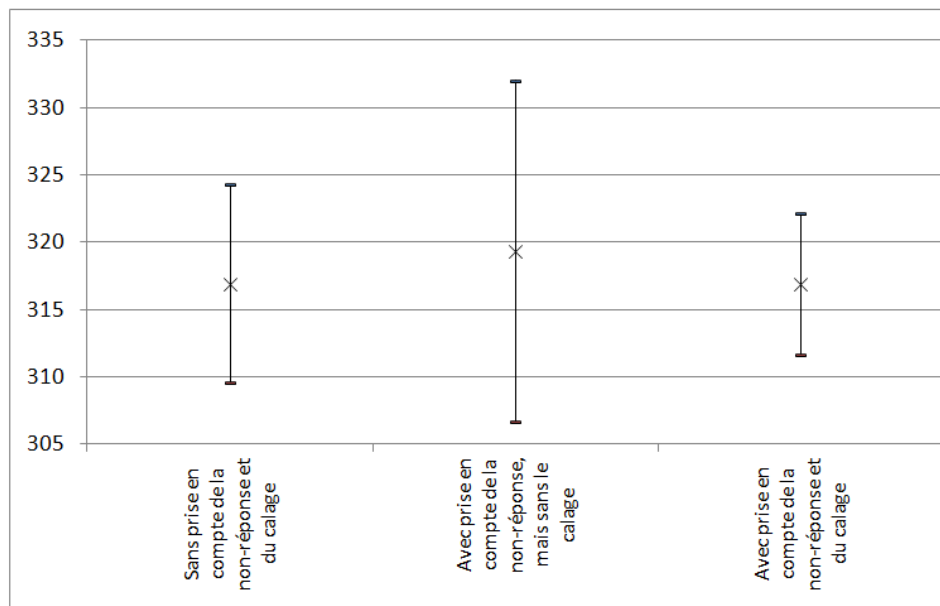


FIGURE 5 – Intervalles de confiance estimés du montant mensuel moyen de minimum vieillesse perçu aux différentes étapes du calcul de précision

L'effet de sondage associé à cette variable est quant à lui désormais estimé à 1,1. En comparaison avec l'effet de sondage obtenu précédemment lorsqu'on s'arrêtait à la prise en compte de la non-réponse (Table 8), cette estimation est bien plus faible. En y regardant de plus près, on constate que les estimations de variance $\widehat{V}(\hat{\mu}_{y,CNR})$ et $V_{SAS}(\hat{\mu}_{SAS,CNR})$ diminuent toutes les deux considérablement lorsqu'on tient compte du calage sur marges, ce qui est bien en conformité avec l'effet recherché lorsqu'on effectue un tel redressement.

Ce qui fait diminuer l'effet de sondage au global, c'est que ce gain de précision est proportionnellement plus marqué quand on considère notre véritable plan de sondage que dans le cas d'un sondage aléatoire simple.

Au final, cette estimation de l'effet de sondage est très proche de 1. On en conclut donc que le plan de sondage mis en oeuvre est de très bonne qualité pour cette variable, puisqu'il n'engendre quasiment pas de perte de précision par rapport à un sondage aléatoire simple.

Les variables d'intérêt de l'enquête :

Les estimateurs des coefficients de variation sont plutôt bons pour les estimateurs du *nombre total d'allocataires pauvres en conditions de vie* et de la *proportion d'allocataires qui se déclarent en mauvaise santé*. Le coefficient de variation estimé de la *proportion d'allocataires disposant d'un patrimoine immobilier* est le plus mauvais, mais il reste tout de même sous le seuil des 10%.

Par ailleurs, quelle que soit la variable considérée, on a toujours un effet de sondage associé très proche de 1 (Table 11). L'accroissement de l'erreur de sondage dû à notre plan de sondage complexe plutôt qu'un sondage aléatoire simple de même taille, est donc relativement limité.

Estimateur	Total d'allocataires pauvres en conditions de vie	Proportion d'allocataires ayant connu de longues périodes d'activité ou n'ayant jamais travaillé	Proportion d'allocataires qui se déclarent en mauvaise santé	Proportion d'allocataires disposant d'un patrimoine immobilier
\hat{V}	141 783.64	0.2228	0.4660	0.1425
\widehat{V}_{SAS}	24 651 594.02	0.000 149	0.000 258	0.000 154
\widehat{CV}	20 037 963.84	0.000 138	0.000 238	0.000 115
\widehat{Inf}_{IC}	0.0350	0.0547	0.0345	0.0872
\widehat{Sup}_{IC}	132 052.17	0.1989	0.4345	0.1182
\widehat{Def}	151 515.11	0.2467	0.4976	0.1669
	1.230	1.079	1.087	1.340

TABLE 11 – Précision finale des estimateurs de différentes variables d'intérêt et effet de sondage associé

Conclusion

A l'issue de cette troisième partie, on obtient de bonnes estimations de la précision des principaux estimateurs de l'enquête pour les allocataires du minimum vieillesse relevant de la Cnav. Les formules tiennent désormais compte du traitement de la non-réponse totale et du calage sur marges.

Par rapport à la partie précédente, on ne travaille plus sur l'ensemble de l'échantillon, mais uniquement sur les répondants à l'enquête (ainsi que sur les hors-champs au moment du calcul des résidus ε lors de la prise en compte du calage).

Comme le montrent les résultats obtenus, la non-réponse se traduit par une diminution de la précision des estimateurs, qui est ici plus que compensée par le gain de précision apporté par le calage. En effet, ce dernier a été particulièrement bénéfique pour nos principales variables d'intérêt, eu égard à la faiblesse des estimations de l'effet de calage.

Au final, les coefficients de variation estimés sont faibles et les estimations de l'effet de sondage très proches de 1. L'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux semble donc être de très bonne qualité.

Conclusion générale

Les résultats de ce travail de calcul de précision réalisé pour l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux sont encourageants : les estimateurs obtenus semblent être de très bonne qualité.

De nombreux enseignements peuvent d'ores et déjà être tirés pour les enquêtes futures menées par la Drees auprès des bénéficiaires de minima sociaux. En effet, certaines contraintes liées au plan de sondage pèsent sur la qualité des estimateurs obtenus et il est intéressant de les identifier précisément :

- La multiplicité des minima sociaux étudiés via cette enquête pose des difficultés au moment du 3^e degré de tirage. En effet, la répartition géographique des bénéficiaires peut être assez différente d'une prestation à une autre. Ainsi, au moment du tirage des individus, on peut être confronté à des échantillons de taille 1.
- L'échelon retenu pour les unités secondaires peut également poser problème, notamment pour le minimum vieillesse. Le découpage en fractions de cantons a été réalisé pour garantir une représentativité suffisante du milieu rural dans notre échantillon, mais ce niveau peut s'avérer trop fin dans certains cas. En effet, dans plusieurs départements, pour une prestation donnée, plus de 50% des fractions de cantons sont vides : la prestation en question est un critère rare.
- Le nombre d'unités secondaires sélectionnées dans chacune des strates de fractions de cantons urbaines et rurales est très faible : leur somme n'excède pas 10 (6 pour le minimum vieillesse). Travailler sur de trop petits échantillons conduit à augmenter la variance liée à ce 2^e degré de tirage. Néanmoins, à budget d'enquête constant, cette restriction d'unités secondaires tirées permet de tirer plus d'unités primaires, ce qui est bon pour la précision globale des estimateurs de l'enquête.

Par ailleurs, plusieurs pistes de travaux complémentaires peuvent être envisagées pour approfondir ces premiers résultats :

- Les travaux réalisés dans le cadre de cet article n'ont été effectués que sur les allocataires du minimum vieillesse relevant de la Cnav. Il serait intéressant d'étendre ces calculs aux autres bénéficiaires de minima sociaux enquêtés. Ce travail est loin d'être immédiat. La base de sondage qui a servi à sélectionner les bénéficiaires d'âge actif n'étant qu'un échantillon, il s'agit de traiter en réalité un tirage en deux phases.
- Les calculs de précision n'ont été effectués que pour un nombre restreint de variables.

Il reste de nombreuses thématiques à explorer, en particulier toutes les nouvelles variables permettant d'estimer le reste-à-vivre des bénéficiaires de minima sociaux. Ces travaux pourront être menés dès lors que les résultats des appariements avec les données sociales et fiscales seront disponibles, à la mi-2015.

Bibliographie

- [1] ISEL A., *Calcul de précision des principaux estimateurs de l'Enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux de la Drees*, Rapport de stage, Master 2 de Statistique-Économétrie, Ensaï et Université de Rennes 1, 2014.
- [2] *Minima sociaux et prestations sociales – édition 2014, Ménages aux revenus modestes et redistribution*, collection Études et Statistiques, Drees, 2014.
- [3] ARNOLD C. et BARTHÉLÉMY N., *Les allocataires du minimum vieillesse : parcours et conditions de vie*, Études et Résultats, n°863, Drees, 2014.
- [4] ISEL A., *Les conditions de vie des bénéficiaires de minima sociaux en 2012 : privations et difficultés financières*, Études et Résultats, n°871, Drees, 2014.
- [5] ARDILLY P., *Les techniques de sondage*, Éditions Technip, 2006.
- [6] CARON N., DEVILLE J.C. et SAUTORY O., *Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel Poulpe*, Document de travail, n°9806, Insee.
- [7] DEVILLE J.C. et SÄRNDAL C.E., *Calibration estimators in survey sampling*, Journal of the American Statistical Association, vol.87, n°418, p. 376–382, 1992.
- [8] MOUSSALLAM K., *Calcul de la précision de l'enquête CVS 2012*, Note interne, n°39/DG75-L110/SF, Insee, 2013.
- [9] TILLÉ Y., *Théorie des sondages : échantillonnage et estimation en populations finies*, Éditions Dunod, 2001.
- [10] VOLANT S., *Estimation de précision pour l'indice d'évolution des loyers du secteur social*, Rapport de stage, Master 2 de Statistique-Économétrie, Ensaï et Université de Rennes 1, 2013.

Annexe 1 : Plan du questionnaire de l'enquête 2012 sur les conditions de vie des bénéficiaires de minima sociaux

Fiche contact pour connexion enquêteur	4
Partie A : renseignements signalétiques.....	8
Sous-partie A : situation de logement.....	8
Sous-partie B : tableau des habitants du logement.....	12
Sous-partie C : modification de la composition familiale.....	21
Sous-partie D : situation du ménage vis-à-vis du travail.....	21
Partie B : situation de l'allocataire.....	23
Sous-partie A : emploi occupé.....	23
Sous-partie B : recherche d'emploi.....	27
Sous-partie C : situation en décembre 2011.....	30
Sous-partie D : vie professionnelle.....	34
Sous-partie E : passé dans les minima sociaux.....	37
Partie C : ressources.....	46
Sous-partie A : revenus mensuels.....	46
Sous-partie B : aides financières reçues.....	52
Partie D : logement.....	54
Sous-partie A : conditions de logement.....	54
Sous-partie B : dépenses de logement.....	57
Sous-partie C : dépenses d'énergie.....	66
Partie E : autres dépenses.....	73
Sous-partie A : emprunts et services financiers.....	73
Sous-partie B : impôts.....	77
Sous-partie C : transports.....	81
Sous-partie D : couverture santé et assurances.....	84
Sous-partie E : téléphone, internet.....	89
Sous-partie F : services.....	91
Sous-partie G : transferts vers d'autres ménages.....	98
Sous-partie H : autres dépenses.....	100
Sous-partie I : habitudes de consommation.....	102
Partie F : situation financière.....	106
Partie G : relations avec les organismes et accompagnement.....	112
Partie H : situation personnelle.....	115
Sous-partie A : études, diplômes et formations.....	115
Sous-partie B : vie sociale.....	118
Sous-partie C : santé, handicap.....	123
Partie I : opinion et perception du travail, du chômage et des minima sociaux.....	126
Coordonnées	129
Observations de l'enquêteur	130