

Article

Estimation de la variance d'indicateurs transversaux pour l'enquête SILC en Suisse

par Anne Massiani

Juin 2013



Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

Service de renseignements 1-800-635-7943
Télécopieur 1-800-565-7757

Comment accéder à ce produit

Le produit n° 12-001-X au catalogue est disponible gratuitement sous format électronique. Pour obtenir un exemplaire, il suffit de visiter notre site Web à www.statcan.gc.ca et de parcourir par « Ressource clé » > « Publications ».

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens ».

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2013

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada (<http://www.statcan.gc.ca/reference/licence-fra.html>).

This publication is also available in English.

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, ses entreprises, ses administrations et les autres établissements. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- P provisoire
- r révisé
- X confidentiel en vertu des dispositions de la *Loi sur la statistique*
- E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Estimation de la variance d'indicateurs transversaux pour l'enquête SILC en Suisse

Anne Massiani¹

Résumé

L'enquête SILC (*Statistics on Income and Living Conditions*) est une enquête européenne annuelle visant à mesurer la répartition des revenus, la pauvreté et les conditions de vie de la population. Elle est réalisée en Suisse depuis 2007 selon un schéma rotatif de quatre panels qui permet de produire à la fois des estimations transversales et des estimations longitudinales. Dans cet article, nous abordons le problème de l'estimation de la variance des indicateurs transversaux sur la pauvreté et l'exclusion sociale retenus par Eurostat. Nos calculs tiennent compte de la non-linéarité des estimateurs, de la non-réponse totale à différentes phases d'enquête, du sondage indirect et du calage. Nous adaptons la méthode d'estimation de variance en cas de non-réponse après un partage des poids proposée par Lavallée (2002) et obtenons un estimateur de variance asymptotiquement sans biais et très simple à programmer.

Mots clés : Enquête SILC; panel rotatif; indices d'inégalité; estimation de variance; méthode du partage des poids.

1 Introduction

L'enquête SILC (*Statistics on Income and Living Conditions*) est une enquête européenne annuelle qui permet d'obtenir des indicateurs comparables d'un pays à un autre sur la pauvreté, l'exclusion sociale et les conditions de vie de la population. Une présentation détaillée de cette enquête peut être trouvée dans Clemenceau et Museux (2006). Conformément aux recommandations d'Eurostat (Eurostat 2003), l'enquête est réalisée en Suisse selon un schéma rotatif de quatre panels, le premier panel ayant été interrogé en 2007. La durée de chaque panel est de quatre ans et, tous les ans, un panel est remplacé. En rythme de croisière, l'échantillon sélectionné pour un panel comporte environ 3 600 ménages. Nous nous intéressons dans cet article à l'estimation transversale, pour la population présente lors d'une année t , des indicateurs sur la pauvreté et sur les conditions de vie retenus par Eurostat tels que le taux de risque de

1. Anne Massiani, Institut de Statistique, Université de Neuchâtel, Pierre-à-Mazel 7, 2000 Neuchâtel, Suisse & Office Fédéral de la Statistique, Espace de l'Europe 10, 2010 Neuchâtel, Suisse. Courriel : anne.massiani@unine.ch.

pauvreté et le rapport interquintile de revenu. Nous renvoyons à Osier (2009) pour une présentation de ces indicateurs et à Ardilly et Lavallée (2007) pour une description de l'approche transversale dans le cadre de l'enquête SILC. Dans cette vision transversale, l'évolution de la composition des ménages des panels sélectionnés est prise en compte grâce à l'utilisation de la théorie du sondage indirect développée par Lavallée (2002).

Nous présentons dans cette étude des formules d'estimation de la variance des indicateurs qui tiennent compte de la grande complexité de l'enquête SILC-Suisse. Les résultats que nous obtenons, bien que développés spécifiquement pour l'enquête réalisée en Suisse, sont susceptibles d'intéresser les autres pays participants, puisque leurs méthodologies sont voisines. Les principaux éléments pris en compte dans l'estimation de la variance des indicateurs sont la non-linéarité des estimateurs, la non-réponse totale à différentes phases d'enquête, le sondage indirect et le calage. Une solution pour estimer la variance d'indicateurs non-linéaires est de recourir à des techniques de linéarisation (voir Deville 1999). Des formules spécifiquement adaptées aux indicateurs retenus par Eurostat ont été développées par Osier (2009). Une formule alternative pour le rapport interquintile de revenu est disponible dans Langel et Tillé (2011). Une fois que les techniques de linéarisation sont appliquées, subsiste encore le problème de l'estimation de la variance d'un total dans un schéma d'enquête complexe. Une des difficultés tient à la présence de non-réponse après un partage des poids. Lavallée (2002) propose un estimateur de la variance d'un total qui en tient compte. Cet estimateur n'est cependant généralement pas sans biais, même lorsque les probabilités de réponse sont connues. Nous proposons ici une adaptation qui corrige ce biais.

Nous présentons brièvement dans la section 2 le déroulement de l'enquête et son plan de sondage. Dans la section 3, nous exposons la pondération utilisée. La section 4 décrit l'application des techniques de linéarisation à ces estimateurs pour obtenir une approximation de leur variance. La section 5 est consacrée au problème de l'estimation de la variance d'un total en présence de non-réponse après un partage des poids. La formule finale utilisée pour l'estimation

de la variance est donnée dans la section 6. Enfin, nous proposons une application numérique dans la section 7, avant de tirer les conclusions de notre étude.

2 Plan de sondage et déroulement de l'enquête

L'enquête est réalisée en Suisse selon un schéma rotatif de quatre panels. Chaque année t , un nouvel échantillon de ménages $s_m^{A_1,t}$ est sélectionné pour remplacer un panel sortant. L'indice m est utilisé ici pour les échantillons de ménages, l'indice p sera réservé pour les échantillons de personnes. En accord avec les notations de Lavallée (2002), les exposants A_i se réfèrent à des échantillons sélectionnés de manière directe, tandis que l'exposant B sera utilisé dans la suite du document pour les échantillons sélectionnés de manière indirecte. Le nouveau panel sélectionné sera suivi durant quatre ans selon le scénario suivant :

- Lors de la première année t , on sollicite tout d'abord les ménages de $s_m^{A_1,t}$ pour remplir un questionnaire préliminaire appelé « grille ». Les variables relevées dans la grille sont communes à tout le ménage et concernent principalement sa composition. On désigne par $s_m^{A_2,t}$ le sous-échantillon de $s_m^{A_1,t}$ qui répond à la grille. L'échantillon des individus appartenant aux ménages de $s_m^{A_2,t}$ est noté $s_p^{A_2,t}$ et ces individus sont appelés des longitudinaux. Les ménages de $s_m^{A_2,t}$ atteints lors de cette première année d'enquête t sont ensuite invités à remplir un questionnaire commun à tout le ménage, appelé questionnaire ménage.
- Lors des trois années suivantes $t + i$, pour $i = 1$ à 3, on tente de recontacter les longitudinaux de $s_p^{A_2,t}$ âgés de 16 ans et plus dans le but d'enquêter leurs ménages tels qu'ils sont composés lors de l'année $t + i$, les ménages pouvant en effet se modifier au cours du temps. Tous les individus de ces ménages sont intégrés à l'enquête de l'année $t + i$ et sont appelés individus transversaux. Les ménages de l'année $t + i$ atteints par le biais des longitudinaux, donc de manière indirecte, sont appelés à remplir à nouveau la grille, puis le questionnaire ménage. On ne parvient cependant pas à recontacter tous les longitudinaux, notamment à cause des déménagements, ce qui constitue une cause significative de non-réponse et d'érosion de l'échantillon.

Nous nous fixons maintenant une année d'enquête t donnée et adoptons une approche transversale, c'est-à-dire que nous nous intéressons aux estimations que l'on peut produire pour

la population présente lors de l'année t . L'échantillon interrogé durant l'année d'enquête t est composé de quatre panels contactés respectivement pour la première, deuxième, troisième ou quatrième fois. On désigne par $s_m^{B,\tau}$ l'échantillon de ménages répondant au questionnaire ménage en τ^e interrogation (ou vague), pour $\tau = 1, \dots, 4$. Comme le montre la figure 2.1, l'échantillon $s_m^{B,\tau}$ a été contacté par l'intermédiaire des longitudinaux sélectionnés lors de l'année $t_\tau = t - \tau + 1$. Les parties grisées à droite représentent les échantillons participant à l'enquête lors de l'année t (sondage indirect), tandis que les échantillons de gauche contiennent les ménages initiaux des longitudinaux ayant permis de les contacter (sondage direct).

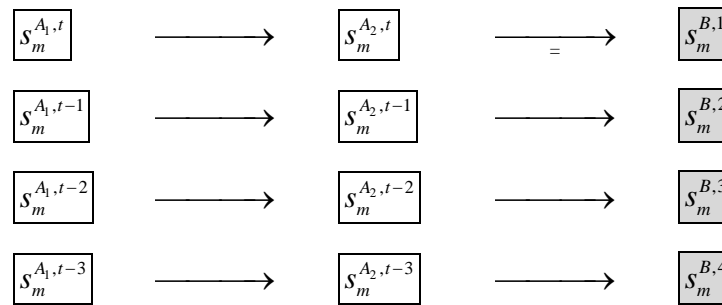


Figure 2.1 Panels composant l'échantillon enquêté lors de l'année t

Nous décrivons ci-dessous de façon plus détaillée le plan de sondage de $s_m^{A_2,t_\tau}$, pour $\tau = 1, \dots, 4$. La composition des ménages pouvant évoluer au fil du temps, nous indiquerons dans la suite du document par k_1 les ménages de la première année d'interrogation t_τ et par k ceux de l'année d'enquête t afin de les distinguer. Chaque échantillon $s_m^{A_2,t_\tau}$ provient de deux phases de sélection.

- **Phase 1** : on sélectionne un échantillon de ménages $s_m^{A_1,t_\tau}$ selon un plan stratifié par grande région, qui sont en Suisse au nombre de sept. Au sein de chaque strate, les tirages sont réalisés selon un plan simple. On note $\pi_{k_1}^{A_1}$ la probabilité d'inclusion d'ordre 1 du ménage k_1 et $\pi_{k_1,k'_1}^{A_1}$ la probabilité d'inclusion d'ordre 2 des ménages k_1 et k'_1 .

- **Phase 2** : la deuxième phase de sélection provient de la non-réponse à la grille lors de la première année d'interrogation t_τ . Cette non-réponse est modélisée par un plan de Poisson sur les ménages et on note, pour tout ménage k_1 :

$$P(k_1 \in s_m^{A_2, t_\tau} \mid k_1 \in s_m^{A_1, t_\tau}) = q_{k_1}^a. \quad (2.1)$$

Pour tous ménages k_1 et k'_1 , on définit alors :

$$\pi_{k_1}^{A_2} = P(k_1 \in s_m^{A_2, t_\tau}) = \pi_{k_1}^{A_1} q_{k_1}^a \quad (2.2)$$

$$\pi_{k_1 k'_1}^{A_2} = P(k_1, k'_1 \in s_m^{A_2, t_\tau}) = \begin{cases} \pi_{k_1}^{A_1} q_{k_1}^a & \text{si } k_1 = k'_1 \\ \pi_{k_1 k'_1}^{A_1} q_{k_1}^a q_{k'_1}^a & \text{si } k_1 \neq k'_1. \end{cases} \quad (2.3)$$

Enfin, pour tous longitudinaux j et j' appartenant respectivement aux ménages k_1 et k'_1 , on pose :

$$\pi_j^{A_2} = P(j \in s_p^{A_2, t_\tau}) = \pi_{k_1}^{A_2} \quad \pi_{j j'}^{A_2} = P(j, j' \in s_p^{A_2, t_\tau}) = \pi_{k_1 k'_1}^{A_2}. \quad (2.4)$$

Pour des raisons pratiques, les échantillons $s_m^{A_1, t_\tau}$, pour $\tau = 1, \dots, 4$ sont tirés de façon à être disjoints, ce qui implique que les échantillons $s_m^{A_2, t_\tau}$ le sont également. En théorie, les quatre échantillons $s_m^{B, \tau}$ ne sont en revanche pas nécessairement disjoints puisqu'il n'est pas exclu que deux longitudinaux sélectionnés lors de deux vagues différentes se trouvent dans le même ménage lors de l'année d'enquête. Compte tenu de la faible taille des échantillons, cela est extrêmement peu probable et n'a jamais été observé en pratique jusqu'à présent. En effet, en rythme de croisière, un panel comporte environ 3 600 ménages tirés avec des probabilités de sélection $\pi_{k_1}^{A_1}$ dont la valeur la plus élevée est de 0,0011. Soulignons toutefois que la méthodologie utilisée permettrait de traiter des échantillons $s_m^{B, \tau}$ non-disjoints. On note :

$$s_m^B = \bigcup_{\tau=1}^4 s_m^{B, \tau}$$

et un poids w_k , appelé poids transversal ménage, est calculé pour chaque ménage k de s_m^B . On désigne par s_p^B les individus des ménages de s_m^B et chaque individu j d'un ménage k de s_m^B reçoit le poids $w_j = w_k$.

L'information utilisée pour calculer les indices de pauvreté et d'exclusion sociale se base sur le revenu de chaque ménage k de s_m^B . Certaines composantes du revenu du ménage sont relevées dans le questionnaire ménage rempli par tous les ménages de s_m^B . D'autres composantes du revenu, comme les salaires, nécessitent un relevé supplémentaire et sont obtenues à partir de questionnaires individuels soumis aux membres des ménages de s_m^B . En cas de non-réponse complète ou partielle aux questionnaires individuels ou bien en cas de non-réponse partielle au questionnaire ménage, certaines composantes du revenu des ménages de s_m^B sont manquantes. Dans un certain nombre de cas, notamment en cas d'absence d'information sur les salaires, les valeurs manquantes peuvent être obtenues par d'autres sources provenant de données administratives. Dans les cas restants, elles sont imputées. Dans le présent article, nous ne tenons pas compte de l'effet des imputations. Dans ces conditions, le revenu de chaque ménage k de s_m^B est connu et les non-réponses observées lors des différentes phases de l'enquête sont des non-réponses totale de ménage : la non-réponse aux grilles successives et la non-réponse totale au questionnaire ménage. Comme les indices de pauvreté et d'exclusion sociale sont calculés au niveau des individus et non des ménages, on divise le revenu de chaque ménage par un facteur qui dépend du nombre de personnes dans le ménage et de leur âge. Le résultat de ce calcul, appelé revenu équivalent, est attribué à chaque individu j de s_p^B . Les indices de pauvreté et d'exclusion sociale sont calculés à partir du revenu équivalent et des poids w_j .

3 Pondération transversale ménage

La pondération a été réalisée par Graf (2008). Elle fait appel à des techniques de non-réponse et de calage (*cf.* Deville et Särndal 1992, pour plus de détails sur le calage). Elle utilise également la méthode du partage des poids (Lavallée 2002) afin d'attribuer un poids aux individus atteints de manière indirecte. Cette pondération est effectuée en cinq grandes étapes. L'objectif des quatre premières étapes est de calculer un poids pour les ménages de chaque panel $s_m^{B,\tau}$, pour

$\tau = 1, \dots, 4$. Ces poids se calculent séparément pour chacun des panels. La dernière étape permet ensuite de combiner ces panels. Nous rappelons ci-dessous le principe de chaque étape dans le cas où les probabilités de réponse intervenant dans les différentes phases de l'enquête sont connues. Des explications sur la façon dont ces paramètres sont estimés sont fournies dans l'application numérique de la section 7.

1. Calcul du poids initial

Soit $\tau \in \{1, \dots, 4\}$ fixé. Pour tout longitudinal $j \in s_p^{A_2, t_\tau}$, on calcule le poids $w_j^{A_2}$ qui tient compte de la non-réponse à la grille en vague 1 :

$$w_j^{A_2} = \frac{1}{\pi_j^{A_2}}. \quad (3.1)$$

2. Correction de la non-réponse à la grille lors de l'année d'enquête

On observe une érosion de l'échantillon entre la première interrogation et l'année d'enquête. D'une part, on ne parvient pas à recontacter tous les longitudinaux de $s_p^{A_2, t_\tau}$. D'autre part, les ménages que l'on parvient à recontacter par l'intermédiaire des longitudinaux n'acceptent pas tous de remplir à nouveau la grille. On désigne par q_k^b la probabilité de réponse à la grille du ménage k lors de l'année d'enquête, conditionnellement à $s_p^{A_2, t_\tau}$. Pour le panel répondant en première interrogation, les q_k^b sont tous égaux à 1 puisqu'il n'y a pas encore eu d'érosion. Pour tout longitudinal j appartenant au ménage k , on pose $q_j^b = q_k^b$ puis on calcule le poids :

$$w_j^g = w_j^{A_2} \frac{1}{q_j^b}. \quad (3.2)$$

Pour des raisons qui seront explicitées dans la section 7, l'estimation des probabilités de réponse q_j^b est un problème délicat. Compte tenu des informations disponibles, ces estimations ne sont pas toutes égales pour les longitudinaux d'un même ménage k , ce qui n'est pas en adéquation avec le mécanisme de réponse. Cette difficulté, ainsi que la façon dont nous la traitons dans le cadre de l'estimation de la variance des indicateurs, seront abordées dans la section 7.

3. Partage des poids

En suivant la méthodologie de Lavallée (2002, chapitre 6), nous introduisons la définition de cohabitant initialement présent : il s'agit d'un individu qui n'a pas été sélectionné lors de la vague 1, mais qui faisait partie de la population cible lors de la

sélection en vague 1. Les nouveaux-nés ou les immigrants sont appelés des cohabitants initialement absents. On désigne par L_k le nombre de longitudinaux et par P_k le nombre de cohabitants initialement présents dans un ménage k de l'année d'enquête. Le poids d'un ménage k répondant à la grille lors de l'année d'enquête est calculé grâce à un partage des poids de la façon suivante :

$$w_k^p = \frac{1}{L_k + P_k} \sum_{j=1}^{L_k} w_j^g. \quad (3.3)$$

4. Correction de la non-réponse au questionnaire ménage

On modélise la non-réponse observée entre la grille de l'année d'enquête et le questionnaire ménage par un plan de Poisson sur les ménages. On note q_k^c la probabilité pour que le ménage k réponde au questionnaire ménage, sachant qu'il a répondu à la grille de l'année d'enquête. Pour tout ménage $k \in s_m^{B,\tau}$, on calcule alors le poids :

$$w_k^{nr} = w_k^p \frac{1}{q_k^c}. \quad (3.4)$$

5. Combinaison des panels puis calage

- Combinaison des panels

On souhaite obtenir une pondération adaptée à la réunion s_m^B des quatre échantillons $s_m^{B,\tau}$, pour $\tau = 1, \dots, 4$. A cet effet, le poids w_k^{nr} calculé à l'étape précédente doit être divisé, en ordre de grandeur, par un facteur 4. On pose, pour tout ménage $k \in s_m^B$:

$$w_k^u = \phi_k^\tau \cdot w_k^{nr}, \quad (3.5)$$

où les ϕ_k^τ sont des facteurs d'allocation proche de 1 / 4 et devant être optimisés. Merkouris (2001) a utilisé des critères de minimisation de variance pour calculer les valeurs optimales des ϕ_k^τ . Ces valeurs dépendent des variances, donc des effets de plan, associés à chaque panel. Ces effets de plan étant inconnus, on suppose qu'au sein de chaque grande région, ils sont identiques pour chacun des panels. Cela conduit à calculer, au sein de chaque grande région, des facteurs ϕ_k^τ proportionnels à la taille de l'échantillon de ménages répondant en τ^e interrogation.

- Calage

On cale ensuite les poids w_k^u sur des marges connues sur la population des ménages de l'année d'enquête. On note w_k le poids calé ainsi obtenu pour le ménage k .

- *Attribution des poids aux individus*

Pour tout individu j appartenant à un ménage k de s_m^B , on pose :

$$w_j = w_k.$$

Remarque 1 : Les différentes étapes de la pondération ont été présentées dans l'ordre où elles sont réalisées en pratique. Nous exposons ci-dessous une autre façon d'aboutir exactement aux mêmes poids finaux lorsque les probabilités de réponse q_j^b sont connues. Dans ce cas, comme $q_j^b = q_k^b$ pour tous les longitudinaux j d'un même ménage k de l'année d'enquête, il est aisé de constater que le poids w_k^{nr} donné par la formule (3.4) peut se réécrire de la façon suivante :

$$w_k^{nr} = \frac{1}{L_k + P_k} \left(\sum_{j=1}^{L_k} w_j^{A_2} \frac{1}{q_j^b} \right) \frac{1}{q_k^c} = w_k^{p*} \frac{1}{q_k^b q_k^c}, \quad (3.6)$$

où

$$w_k^{p*} = \frac{1}{L_k + P_k} \sum_{j=1}^{L_k} w_j^{A_2}. \quad (3.7)$$

La formule (3.6) montre que l'on peut aboutir au même poids w_k^{nr} que Graf (2008), donc aux mêmes poids finaux, en procédant comme suit :

- attribuer au ménage k de l'année d'enquête le poids partagé w_k^{p*} ,
- corriger dans une même étape la non-réponse à la grille et au questionnaire ménage de l'année d'enquête en la modélisant par un plan de Poisson sur les ménages de paramètre $q_k^b q_k^c$, et en appliquant le facteur de correction $1 / (q_k^b q_k^c)$.

Cette constatation sera utile pour les calculs de variance dans la section 5.

4 Linéarisation et approximation de la variance

On souhaite estimer la variance $\text{var}(\hat{\Theta})$ d'un estimateur $\hat{\Theta}$ calculé sur l'échantillon d'individus transversaux s_p^B affectés des poids w_j . Lavallée (2002, pages 122-123) a développé un cadre asymptotique pour une population sondée de manière indirecte. Ce cadre permet

d'utiliser des techniques de linéarisation (*cf.* Deville 1999) pour obtenir une approximation de la variance d'un estimateur complexe calculé sur une population sondée de manière indirecte. Si $\hat{\Theta}$ est l'estimateur d'un des indices d'inégalité retenus par Eurostat, on emploie des techniques de linéarisation de façon à ramener l'estimation de la variance de $\hat{\Theta}$ à l'estimation de la variance d'un total. Les macros de Osier (2009) peuvent être utilisées à cet effet. Ses formules de linéarisation sont rappelées dans l'annexe A dans le cas des quatre indicateurs considérés dans l'application numérique de la section 7. On note ℓ_j les linéarisées de $\hat{\Theta}$. On a alors :

$$\text{var}(\hat{\Theta}) \simeq \text{var} \left(\sum_{j \in s_p^B} w_j \ell_j \right). \quad (4.1)$$

L'utilisation des résidus de la régression de la variable d'intérêt par rapport aux variables de calage permet de tenir compte de l'effet du calage dans les calculs de variance (*cf.* Deville et Särndal 1992). Comme les variables de calage x_k sont définies au niveau du ménage, on calcule tout d'abord, pour tout ménage k de l'année d'enquête :

$$\ell_k = \sum_{j \in m_k} \ell_j,$$

où m_k désigne l'ensemble des membres du ménage k , puis on définit $e_k = \ell_k - x_k^T \beta$, le paramètre β étant calculé ici à partir de l'ensemble des ménages présents dans la population. On a alors :

$$\text{var}(\hat{\Theta}) \simeq \text{var} \left(\sum_{j \in s_p^B} w_j \ell_j \right) = \text{var} \left(\sum_{k \in s_m^B} w_k \ell_k \right) \simeq \text{var} \left(\sum_{k \in s_m^B} w_k^u e_k \right). \quad (4.2)$$

Remarque 2 : Les linéarisées ℓ_j font intervenir des quantités calculées sur toute la population, comme le montre par exemple la formule (A.6) de l'annexe A. Conformément à la pratique habituelle, les linéarisées ℓ_j seront remplacées *in fine* par des estimations $\hat{\ell}_j$. De même, les quantités e_k étant inconnues, elles seront remplacées par les estimations

$$\hat{e}_k = \hat{\ell}_k - x_k^T \hat{\beta}, \quad (4.3)$$

où

$$\hat{\ell}_k = \sum_{j \in m_k} \hat{\ell}_j$$

et

$$\hat{\beta} = \left(\sum_{k \in s_m^B} w_k x_k x_k^T \right)^{-1} \left(\sum_{k \in s_m^B} w_k x_k \hat{\ell}_k \right).$$

Enfin, les quatre échantillons $s_m^{B,\tau}$, pour $\tau = 1, \dots, 4$, qui composent s_m^B étant atteints par l'intermédiaire d'échantillons $s_m^{A_1, \tau}$ disjoints, ils ne sont en toute rigueur pas indépendants. Cependant, nous faisons l'approximation que ces quatre échantillons sont indépendants car les probabilités de sélection $\pi_k^{A_1}$ sont très faibles. Nous supposons également que les facteurs d'allocation ϕ_k^τ intervenant dans la formule (3.5) ne sont pas aléatoires. En posant, pour tout ménage k :

$$e'_k = \phi_k^\tau \cdot e_k \quad (4.4)$$

et en rappelant (3.5), on peut réécrire la somme intervenant dans le dernier membre de (4.2) sous la forme suivante :

$$\sum_{k \in s_m^B} w_k^u e_k = \sum_{\tau=1}^4 \sum_{k \in s_m^{B,\tau}} w_k^{nr} (\phi_k^\tau e_k) = \sum_{\tau=1}^4 \left(\sum_{k \in s_m^{B,\tau}} w_k^{nr} e'_k \right). \quad (4.5)$$

Cela permet d'obtenir l'approximation suivante de la variance :

$$\text{var}(\hat{\Theta}) \simeq \sum_{\tau=1}^4 \text{var}(\hat{T}_\tau), \quad (4.6)$$

où

$$\hat{T}_\tau = \sum_{k \in s_m^{B,\tau}} w_k^{nr} e'_k. \quad (4.7)$$

Les quatre composantes de variance intervenant dans la formule (4.6) peuvent être calculées et estimées de la même manière. Nous donnons dans la section suivante un estimateur de la variance de \hat{T}_τ , pour τ quelconque.

5 Estimation de la variance et partage des poids

Les calculs présentés dans cette section sont des adaptations des techniques développées par Lavallée (2002, chapitre 8.5) pour le traitement de la non-réponse de grappe (NRG) dans le cadre du sondage indirect. Les résultats sont donnés dans le cas fictif où les probabilités de réponse intervenant aux différentes phases de l'enquête sont connues. Les quantités e_k ainsi que les quantités e'_k définies par (4.4) sont également supposées connues. Toutes ces quantités seront remplacées par des estimations dans la section 7. On désigne par $\mathbf{1}_{\{k \in s_m^{B,\tau}\}}$ la variable indicatrice qui vaut 1 si le ménage k présent lors de l'année d'enquête t fait partie de l'échantillon $s_m^{B,\tau}$ répondant en τ^e vague. Conditionnellement au fait que l'on ait tenté de le contacter par l'intermédiaire des longitudinaux qu'il contient, le ménage k appartient à $s_m^{B,\tau}$ si il a répondu à la grille puis au questionnaire lors de l'année d'enquête t . Par conséquent, on a :

$$E(\mathbf{1}_{\{k \in s_m^{B,\tau}\}} \mid s_p^{A_2,t_\tau}) = P(k \in s_m^{B,\tau} \mid s_p^{A_2,t_\tau}) = q_k^b q_k^c. \quad (5.1)$$

En utilisant le théorème 8.1 de Lavallée (2002, page 151) et la remarque 1, on vérifie aisément que l'estimateur (4.7) peut se réécrire sous la forme :

$$\hat{T}_\tau = \sum_{j \in s_p^{A_2,t_\tau}} w_j^{A_2} \hat{Z}_j = \sum_{j \in s_p^{A_2,t_\tau}} \frac{1}{\pi_j^{A_2}} \hat{Z}_j, \quad (5.2)$$

où l'on a noté, pour tout longitudinal $j \in s_p^{A_2,t_\tau}$ faisant partie du ménage k lors de l'année d'enquête :

$$\hat{Z}_j = \frac{e'_k}{L_k + P_k} \frac{1}{q_k^b} \frac{1}{q_k^c} \mathbf{1}_{\{k \in s_m^{B,\tau}\}}. \quad (5.3)$$

L'estimateur \hat{T}_τ est ainsi ramené à une somme sur des individus sélectionnés de manière directe. On décompose maintenant la variance de \hat{T}_τ de façon standard en conditionnant par rapport à $s_p^{A_2,t_\tau}$:

$$\text{var}(\hat{T}_\tau) = \text{var}_{s_p^{A_2,t_\tau}} \left[E(\hat{T}_\tau \mid s_p^{A_2,t_\tau}) \right] + E_{s_p^{A_2,t_\tau}} \left[\text{var}(\hat{T}_\tau \mid s_p^{A_2,t_\tau}) \right]. \quad (5.4)$$

Pour tout individu j présent dans la population lors de l'année du tirage de $s_p^{A_2, t_\tau}$ et appartenant au ménage k lors de l'année d'enquête, on pose :

$$Z_j = \frac{e'_k}{L_k + P_k}. \quad (5.5)$$

En utilisant (5.1), on vérifie que pour tous longitudinaux j et j' appartenant à $s_p^{A_2, t_\tau}$ et faisant partie respectivement des ménages k et k' lors de l'année d'enquête, on a :

$$E(\hat{Z}_j | s_p^{A_2, t_\tau}) = Z_j \quad (5.6)$$

et

$$\text{cov}_{jj'} = \text{cov} \left[(\hat{Z}_j, \hat{Z}_{j'}) | s_p^{A_2, t_\tau} \right] = \begin{cases} 0 & \text{si } k \neq k' \\ \left(\frac{e'_k}{L_k + P_k} \right)^2 \frac{1 - q_k^b q_k^c}{q_k^b q_k^c} & \text{si } k = k'. \end{cases} \quad (5.7)$$

La formule (5.4) devient alors :

$$\text{var}(\hat{T}_\tau) = \underbrace{\sum_{j=1}^{J_\tau} \sum_{j'=1}^{J_\tau} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} Z_j Z_{j'}}_{V_{A_2}^\tau} + E_{s_p^{A_2, t_\tau}} \left\{ \underbrace{\sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{1}{\pi_j^{A_2}} \frac{1}{\pi_{j'}^{A_2}} \text{cov}_{jj'}}_{V_{\text{NRG}}^\tau} \right\}, \quad (5.8)$$

où J_τ désigne le nombre de personnes présentes dans la population lors de l'année du tirage de $s_p^{A_2, t_\tau}$. Le premier terme $V_{A_2}^\tau$ est la part de variance due au mécanisme de sélection des longitudinaux de $s_p^{A_2, t_\tau}$, tandis que le deuxième terme V_{NRG}^τ est celle due à la non-réponse des ménages à la grille puis au questionnaire ménage lors de l'année t , qui constituent une non-réponse de grappe.

Pour obtenir un estimateur de la variance de \hat{T}_τ , nous adaptons la formule d'estimation de variance (8.37) de Lavallée (2002, page 154). Les différences suivantes sont à noter. Tout d'abord, nous négligeons le fait, qu'en pratique, les probabilités de réponse devront être estimées alors que Lavallée (2002) en tient compte. D'autre part, la méthode d'estimation proposée par

Lavallée (2002) fournit des estimations biaisées, même lorsqu'elle est appliquée dans le cas où les probabilités de réponse sont connues. Par conséquent, nous adaptons sa méthode de façon à obtenir un estimateur sans biais de la variance. Afin de justifier notre propos, nous explicitons tout d'abord ci-dessous le biais obtenu en appliquant la méthode de Lavallée (2002) dans le cas où les probabilités de réponse sont connues. Sa méthode consiste à estimer V_{NRG}^τ par un estimateur sans biais $\hat{V}_{\text{NRG}}^\tau$ puis à estimer $V_{A_2}^\tau$ par $\hat{V}_1^\tau = \tilde{V}_{A_2}^\tau(\hat{Z}_1, \hat{Z}_2, \dots)$ où :

$$\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots) = \sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} \frac{1}{\pi_{jj'}^{A_2}} Z_j Z_{j'} \quad (5.9)$$

est l'estimateur de Horvitz-Thompson de la variance de

$$\tilde{T}_\tau = \sum_{j \in s_p^{A_2, t_\tau}} \frac{1}{\pi_j^{A_2}} Z_j. \quad (5.10)$$

Ceci conduit à l'estimateur de variance :

$$\widehat{\text{var}}_L(\hat{T}_\tau) = \hat{V}_1^\tau + \hat{V}_{\text{NRG}}^\tau. \quad (5.11)$$

L'utilisation de \hat{V}_1^τ , qui est construit en remplaçant les Z_j intervenant dans (5.9) par les \hat{Z}_j , est motivée par le fait que les Z_j ne sont pas connus pour tous les longitudinaux j de $s_p^{A_2, t_\tau}$, mais seulement pour les longitudinaux qui font partie lors de l'année d'enquête d'un ménage k ayant répondu au questionnaire, c'est-à-dire un ménage k appartenant à $s_m^{B, \tau}$. Le recours aux \hat{Z}_j permet d'attribuer plus de poids aux longitudinaux de $s_p^{A_2, t_\tau}$ pour lesquels les Z_j sont connus. Le problème est que l'estimateur \hat{V}_1^τ ainsi construit n'estime pas sans biais $V_{A_2}^\tau$. Ceci se constate aisément en observant ce qui se produit pour les termes diagonaux : pour tout longitudinal j faisant partie du ménage k lors de l'année d'enquête, la quantité Z_j^2 intervenant dans (5.9) est remplacée par $(\hat{Z}_j)^2 = Z_j^2 / (q_k^b q_k^c)^2 \mathbf{1}_{\{k \in s_m^{B, \tau}\}}$ alors qu'une augmentation de poids d'un facteur $1 / (q_k^b q_k^c)$ seulement est sans doute un meilleur choix. Le même type de problème se pose pour le produit $Z_j Z_{j'}$ lorsque les longitudinaux j et j' appartiennent au même ménage lors de

l'année d'enquête. Plus précisément, on a, pour tous longitudinaux j et j' appartenant à $s_p^{A_2, t_\tau}$:

$$E(\hat{Z}_j \hat{Z}_{j'} | s_p^{A_2, t_\tau}) = Z_j Z_{j'} + \text{cov}_{jj'} \quad (5.12)$$

ce qui implique

$$E[\hat{V}_1^\tau] - V_{A_2}^\tau = E_{s_p^{A_2, t_\tau}} \left\{ E \left[\tilde{V}_{A_2}^\tau (\hat{Z}_1, \hat{Z}_2, \dots) | s_p^{A_2, t_\tau} \right] \right\} - V_{A_2}^\tau \quad (5.13)$$

$$= E_{s_p^{A_2, t_\tau}} \left[\sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} \frac{1}{\pi_{jj'}^{A_2}} \text{cov}_{jj'} \right]. \quad (5.14)$$

Comme en revanche $\hat{V}_{\text{NRG}}^\tau$ estime sans biais V_{NRG}^τ , il vient :

$$B^\tau = E \left[\widehat{\text{var}}_L(\hat{T}_\tau) \right] - \text{var}(\hat{T}_\tau) \quad (5.15)$$

$$= E_{s_p^{A_2, t_\tau}} \left[\sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} \frac{1}{\pi_{jj'}^{A_2}} \text{cov}_{jj'} \right]. \quad (5.16)$$

Le biais B^τ dépend notamment du terme $\text{cov}_{jj'}$ défini par (5.7), et donc des probabilités de réponse q_k^b et q_k^c à la grille et au questionnaire lors de l'année d'enquête. Plaçons-nous dans le cas simple du panel répondant en première interrogation, $\tau = 1$, dans lequel la composition des ménages n'a pas encore pu évoluer. La quantité $\text{cov}_{jj'}$ définie par (5.7) est positive si les longitudinaux j et j' appartiennent au même ménage k , et nulle sinon. De plus, pour tous longitudinaux j et j' appartenant au même ménage k , on a d'après la relation (2.4) :

$$\frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_{jj'}^{A_2}} = 1 - \pi_k^{A_2}. \quad (5.17)$$

Cette dernière quantité étant également positive, il découle de l'expression du biais donné par la formule (5.16) que B^1 est positif. D'autre part, les probabilités d'inclusion $\pi_k^{A_2}$ sont très faibles, de sorte que $1 - \pi_k^{A_2} \simeq 1$. En se servant de (5.16) et (5.17), ceci permet de faire l'approximation suivante :

$$B^1 \simeq E_{s_p^{A_2, t_1}} \left[\sum_{j \in s_p^{A_2, t_1}} \sum_{j' \in s_p^{A_2, t_1}} \frac{1}{\pi_j^{A_2}} \frac{1}{\pi_{j'}^{A_2}} \text{cov}_{jj'} \right] = V_{\text{NRG}}^1, \quad (5.18)$$

où on rappelle que V_{NRG}^τ est défini par (5.8) et correspond en première vague à la part de variance due à la non-réponse observée entre la grille et le questionnaire. Par conséquent, l'estimateur $\widehat{\text{var}}_L(\hat{T}_1)$ surestime la variance de \hat{T}_1 et l'erreur commise est de l'ordre de grandeur de V_{NRG}^1 . Le biais peut-être relativement important si les probabilités de réponse au questionnaire sont faibles. En ce qui concerne les autres vagues, la quantité $(\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}) / \pi_{jj'}^{A_2}$ intervenant dans (5.16) dépend des ménages k_1 et k_1' auxquels les longitudinaux j et j' appartaient lors de l'année de leur sélection et obtenir un ordre de grandeur du biais B^τ n'est plus chose aisée.

Nous introduisons un terme correcteur du biais B^τ et donnons notre formule d'estimation de la variance dans la proposition 1 ci-dessous. On rappelle que m_k désigne l'ensemble des membres qui composent le ménage k lors de l'année d'enquête t (cf. page 11) et on note \tilde{m}_k l'ensemble, de cardinal L_k , constitué des longitudinaux j appartenant à m_k .

Proposition 1 : On peut estimer sans biais la variance de \hat{T}_τ par

$$\widehat{\text{var}}(\hat{T}_\tau) = \hat{V}_1^\tau + \hat{V}_2^\tau, \quad (5.19)$$

où

$$\hat{V}_1^\tau = \tilde{V}_{A_2}^\tau(\hat{Z}_1, \hat{Z}_2, \dots) = \sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} \frac{1}{\pi_{jj'}^{A_2}} \hat{Z}_j \hat{Z}_{j'}$$

et

$$\hat{V}_2^\tau = \sum_{k \in s_m^{B, \tau}} \sum_{j, j' \in \tilde{m}_k} \frac{1}{\pi_{jj'}^{A_2}} \left(\frac{e'_k}{L_k + P_k} \right)^2 \frac{1 - q_k^b q_k^c}{(q_k^b q_k^c)^2}.$$

La démonstration de la proposition 1 est donnée dans l'annexe B.

Remarque 3 : L'estimateur $\widehat{\text{var}}(\hat{T}_\tau) = \hat{V}_1^\tau + \hat{V}_2^\tau$ est la somme de deux estimateurs biaisés dont les biais s'équilibrent par construction, de sorte que $\widehat{\text{var}}(\hat{T}_\tau)$ estime sans biais la variance de \hat{T}_τ .

Remarque 4 : La proposition 1 est énoncée sous l'hypothèse que les e_k et les probabilités de réponse sont connues, ce qui permet de conclure à l'absence de biais pour l'estimateur $\widehat{\text{var}}(\hat{T}_\tau)$.

donné par la formule (5.19). En pratique, ces quantités doivent être estimées. Ceci a pour conséquence que l'estimateur de variance ainsi obtenu n'est plus sans biais mais seulement asymptotiquement sans biais, sous réserve que les modèles de non-réponse puissent être considérés comme justes et que leurs paramètres soient estimés par une méthode appropriée.

6 Formule finale

Le terme \hat{V}_2^τ qui intervient dans la proposition 1 comporte une somme double, mais celle-ci ne pose pas de problème sur le plan opérationnel. En effet, elle ne compte que très peu de termes puisqu'elle ne porte que sur les individus d'un même ménage. L'expression de $\hat{V}_1^\tau = \tilde{V}_{A_2}^\tau(\hat{Z}_1, \hat{Z}_2, \dots)$ doit en revanche être transformée pour être calculable plus facilement. Nous commençons donc par donner une autre expression du terme $\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots)$ défini par la formule (5.9). On observe pour cela que :

$$\tilde{T}_\tau = \sum_{j \in s_p^{A_2, t_\tau}} \frac{1}{\pi_j^{A_2}} Z_j = \sum_{k_1 \in s_m^{A_2, t_\tau}} \frac{1}{\pi_{k_1}^{A_2}} T_{k_1}, \quad (6.1)$$

où

$$T_{k_1} = \sum_{j \in m_{k_1}} Z_j. \quad (6.2)$$

Rappelons que la sélection de $s_m^{A_2, t_\tau}$ résulte de la succession d'un plan stratifié par grandes régions et d'une phase de non-réponse modélisée par un plan de Poisson sur les ménages de la vague 1 (cf. section 2). Donner une expression simple de l'estimateur de variance de Horvitz-Thompson d'un total pour ce plan très classique est un problème déjà largement étudié, notamment dans le cadre du logiciel de calcul de précision POULPE (cf. Caron, Deville et Sautory 1998, page 13). Afin de donner une telle expression, nous introduisons les notations suivantes. Pour chacune des sept strates h , on désigne par N_h le nombre de ménages qu'elle contient et par n_h le nombre de ménages sélectionnés. Pour tout ménage $k_1 \in s_m^{A_1, t_\tau}$, on pose :

$$T_{k_1}^* = \begin{cases} \frac{T_{k_1}}{q_{k_1}^a} & \text{si } k_1 \in s_m^{A_2, t_\tau} \\ 0 & \text{sinon.} \end{cases} \quad (6.3)$$

On pose également, pour tout h :

$$(s_h^*)^2 = \frac{1}{n_h - 1} \sum_{k_1 \in \{s_m^{A_1, t_\tau} \cap h\}} (T_{k_1}^* - \bar{T}_h^*)^2 = \frac{1}{n_h - 1} \left[\sum_{k_1 \in \{s_m^{A_2, t_\tau} \cap h\}} \left(\frac{T_{k_1}}{q_{k_1}^a} \right)^2 \right] - \frac{n_h}{n_h - 1} (\bar{T}_h^*)^2 \quad (6.4)$$

où

$$\bar{T}_h^* = \frac{1}{n_h} \sum_{k_1 \in \{s_m^{A_1, t_\tau} \cap h\}} T_{k_1}^* = \frac{1}{n_h} \sum_{k_1 \in \{s_m^{A_2, t_\tau} \cap h\}} \frac{T_{k_1}}{q_{k_1}^a}.$$

D'après Caron et coll. (1998, page 13), le terme $\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots)$ peut s'écrire ici [voir également la formule (11.12) de Särndal et Lundström 2005) :

$$\begin{aligned} \tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots) &= \sum_{k_1 \in s_m^{A_1, t_\tau}} \sum_{k'_1 \in s_m^{A_1, t_\tau}} \frac{\pi_{k_1 k'_1}^{A_1} - \pi_{k_1}^{A_1} \pi_{k'_1}^{A_1}}{\pi_{k_1}^{A_1} \pi_{k'_1}^{A_1}} \frac{1}{\pi_{k_1 k'_1}^{A_1}} T_{k_1}^* T_{k'_1}^* - \sum_{k_1 \in s_m^{A_2, t_\tau}} \frac{1 - \pi_{k_1}^{A_1}}{(\pi_{k_1}^{A_1})^2} \frac{1 - q_{k_1}^a}{(q_{k_1}^a)^2} T_{k_1}^2 \\ &+ \sum_{k_1 \in s_m^{A_2, t_\tau}} \frac{1 - q_{k_1}^a}{(q_{k_1}^a)^2} \left(\frac{T_{k_1}}{\pi_{k_1}^{A_1}} \right)^2. \end{aligned} \quad (6.5)$$

En regroupant les deux derniers termes et en utilisant le fait que le plan de sondage de $s_m^{A_1, t_\tau}$ est un plan stratifié, on aboutit à l'expression simple suivante pour $\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots)$:

$$\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots) = \sum_{h=1}^7 \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h} \right) (s_h^*)^2 + \sum_{k_1 \in s_m^{A_2, t_\tau}} \frac{1 - q_{k_1}^a}{(q_{k_1}^a)^2} \frac{1}{\pi_{k_1}^{A_1}} T_{k_1}^2. \quad (6.6)$$

On note \hat{T}_{k_1} et \hat{s}_h^* les estimateurs obtenus en remplaçant les variables Z_j par les \hat{Z}_j dans les formules (6.2) et (6.4). La relation (6.6), combinée à la formule (4.6) et à la proposition 1, permet d'obtenir la formule finale ci-dessous pour l'estimation de la variance de l'estimateur complexe $\hat{\Theta}$:

$$\widehat{\text{var}}(\hat{\Theta}) \simeq \sum_{\tau=1}^4 \left[\hat{V}_1^{\tau} + \hat{V}_2^{\tau} \right] = \underbrace{\sum_{\tau=1}^4 \hat{V}_{1,1}^{\tau}}_{\hat{V}_{1,1}} + \underbrace{\sum_{\tau=1}^4 \hat{V}_{1,2}^{\tau}}_{\hat{V}_{1,2}} + \underbrace{\sum_{\tau=1}^4 \hat{V}_2^{\tau}}_{\hat{V}_2}, \quad (6.7)$$

où

$$\hat{V}_{1,1}^{\tau} = \sum_{h=1}^7 \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h} \right) (\hat{s}_h^*)^2 \quad (6.8)$$

$$\hat{V}_{1,2}^{\tau} = \sum_{k_1 \in s_m^{A_2, t_{\tau}}} \frac{1 - q_{k_1}^a}{(q_{k_1}^a)^2} \frac{1}{\pi_{k_1}^{A_1}} (\hat{f}_{k_1})^2 \quad (6.9)$$

$$\hat{V}_2^{\tau} = \sum_{k \in s_m^{B, \tau}} \sum_{j, j' \in \tilde{m}_k} \frac{1}{\pi_{jj'}^{A_2}} \left(\frac{\hat{e}'_k}{L_k + P_k} \right)^2 \frac{1 - q_k^b q_k^c}{(q_k^b q_k^c)^2}. \quad (6.10)$$

Remarque 5 : La formule d'estimation de variance (6.7) fournit toujours des estimations positives. De plus, les trois termes qui la composent peuvent être programmés très aisément.

7 Application numérique et discussion

Nous présentons maintenant une application numérique permettant d'illustrer les résultats développés dans les sections précédentes. Jusqu'à présent, nous avons supposé que les probabilités de réponse intervenant dans les différentes phases de l'enquête sont connues, mais en pratique, elles doivent être estimées. En ce qui concerne la non-réponse à la grille lors de la première interrogation, l'estimation des probabilités de réponse $q_{k_1}^a$ définies par (2.1) utilise des informations sur les ménages non-répondants collectées à partir d'une enquête de non-réponse. L'échantillon $s_m^{A_1, t_{\tau}}$ des ménages contactés en première interrogation lors de l'année t_{τ} est divisé en groupes de réponse homogènes au sein desquels des estimations $\hat{q}_{k_1}^a$ des probabilités de réponse $q_{k_1}^a$ sont calculées. Les facteurs q_k^c intervenant dans (3.4) et qui représentent les probabilités de réponse entre la grille et le questionnaire lors de l'année d'enquête sont également estimés au sein de groupes de réponse homogènes. Ce sont cette fois les ménages ayant répondu à la grille lors de l'année d'enquête qui sont répartis en groupes de réponse homogènes. On note

\hat{q}_k^c l'estimation de la probabilité de réponse q_k^c ainsi obtenue. L'estimation des facteurs q_j^b intervenant dans (3.2) et qui représentent les probabilités de réponse entre la grille de la première interrogation et celle de l'année d'enquête est un peu plus problématique. Les deux options suivantes sont envisageables. La première option est de constituer des groupes de réponse homogènes au sein de l'ensemble des ménages que l'on a tenté de contacter lors de l'année d'enquête. On est alors confronté au problème que très peu d'informations sont disponibles pour constituer ces groupes. On ne connaît en effet pas la nouvelle composition des ménages que l'on n'a pas réussi à recontacter ou qui refusent de remplir à nouveau la grille. La deuxième option est d'estimer les q_j^b au sein de groupes de réponse homogènes définis sur les longitudinaux de $s_p^{A_2, t_1}$. Cela n'est pas en totale adéquation avec le mécanisme de réponse mais permet d'utiliser toutes les informations qui ont été collectées au sujet des longitudinaux lors des vagues précédentes. Depuis la première exploitation officielle de l'enquête, le choix s'est porté sur la deuxième option (voir Graf 2008). Compte tenu de cette méthode d'estimation, les quantités \hat{q}_j^b ne sont pas toutes égales au sein d'un même ménage k de l'année d'enquête, ce qui pose problème pour l'application des formules d'estimation de variance développées dans les sections précédentes. On contourne cette difficulté grâce à la constatation suivante. On désigne par \tilde{w}_k^p le poids partagé obtenu à partir de (3.3) et de l'estimation des probabilités de réponse $q_{k_1}^a$ et q_j^b :

$$\tilde{w}_k^p = \frac{1}{L_k + P_k} \sum_{j=1}^{L_k} \left(\tilde{w}_j^{A_2} \frac{1}{\hat{q}_j^b} \right), \quad (7.1)$$

où $\tilde{w}_j^{A_2} = 1 / (\pi_{k_1}^{A_1} \hat{q}_{k_1}^a)$ pour tout longitudinal j appartenant au ménage k_1 en première interrogation. On peut réécrire le poids partagé \tilde{w}_k^p du ménage k de la façon suivante :

$$\tilde{w}_k^p = \frac{1}{L_k + P_k} \sum_{j=1}^{L_k} \left(\tilde{w}_j^{A_2} \frac{1}{\hat{q}_j^b} \right) = \frac{1}{L_k + P_k} \sum_{j=1}^{L_k} \left(\tilde{w}_j^{A_2} \frac{1}{\tilde{q}_j^b} \right) \quad (7.2)$$

où, pour tous les longitudinaux j du ménage k , les quantités \tilde{q}_j^b sont égales à une valeur commune \tilde{q}_k^b définie par

$$\tilde{q}_k^b = \frac{\sum_{j=1}^{L_k} \tilde{w}_j^{A_2}}{\sum_{j=1}^{L_k} \left(\tilde{w}_j^{A_2} \frac{1}{\hat{q}_j^b} \right)}. \quad (7.3)$$

Par conséquent, que l'on estime les quantités q_j^b par \hat{q}_j^b ou par \tilde{q}_j^b , on obtient exactement le même poids partagé \tilde{w}_k^p , et donc le même poids final et les mêmes estimations des indices de pauvreté. En se basant sur cette constatation, nous obtenons une approximation de la précision de $\hat{\Theta}$ en remplaçant dans les formules d'estimation de variance les quantités inconnues $q_{k_1}^a, q_k^b, q_k^c$ par $\hat{q}_{k_1}^a, \tilde{q}_k^b, \hat{q}_k^c$. Les e_k sont également remplacés par les \hat{e}_k donnés par (4.3). Afin de ne pas compliquer inutilement les notations, on note dans ce qui suit de la même manière les estimateurs de variance calculés à partir des paramètres inconnus et ceux calculés à partir de l'estimation de ceux-ci. Si les estimateurs des probabilités de réponse sont convergents, l'estimateur $\widehat{\text{var}}(\hat{\Theta})$ que nous proposons dans la formule (6.7) est asymptotiquement sans biais. Le caractère asymptotique de cette propriété est dû à l'effet de l'estimation des probabilités de réponse ainsi qu'aux approximations faites lors de la linéarisation, de la prise en compte du calage, et de l'obtention de la formule (4.6).

Les calculs présentés dans cette section ont été réalisés à partir de l'enquête SILC-Suisse 2009. L'enquête SILC n'a débuté en Suisse qu'en 2007. En 2009, l'enquête n'a donc pas encore atteint son rythme de croisière et n'est composée que de trois panels au lieu de quatre. Toutefois cela ne change rien à la méthodologie présentée dans les sections précédentes. En 2009, 7 372 ménages, soit 17 561 personnes, ont accepté de participer à l'enquête. Nous donnons dans le tableau 7.1 les taux de réponse obtenus aux différentes phases de l'enquête pour chacun des panels.

Tableau 7.1
Taux de réponse aux différentes phases de l'enquête

Taux de réponse	Panel 07	Panel 08	Panel 09
A la grille en vague 1	0,688	0,694	0,687
Entre la grille de la vague 1 et celle de la vague courante	0,803	0,834	1,000
Entre la grille et le questionnaire ménage de la vague courante	0,966	0,951	0,942

Sur la base de l'enquête de 2009, les taux de pauvreté de différents sous-groupes de population ont été calculés, analysés et commentés par l'Office fédéral de la statistique (OFS 2010). Nous présentons dans le tableau 7.2 différentes mesures de précision pour quelques uns des indices d'inégalité utilisés le plus fréquemment : taux de risque de pauvreté, rapport interquintile de revenu, indice de Gini, écart relatif médian du risque de pauvreté. Pour chaque indicateur $\hat{\Theta}$ considéré, le tableau 7.2 contient les mesures suivantes :

$$ET = \sqrt{\widehat{\text{var}}(\hat{\Theta})}, \quad CV = \frac{ET}{\hat{\Theta}}, \quad ET_{1,1} = \sqrt{\hat{V}_{1,1}}, \quad \Delta_1 = \frac{ET_{1,1}}{ET} - 1,$$

où les variances $\widehat{\text{var}}(\hat{\Theta})$ et $\hat{V}_{1,1}$ sont données par la formule (6.7). On désigne par $\widehat{\text{var}}_L(\hat{\Theta})$ l'estimateur de variance obtenu par la méthode de Lavallée (2002) :

$$\widehat{\text{var}}_L(\hat{\Theta}) = \sum_{\tau=1}^3 \widehat{\text{var}}_L(\hat{T}_\tau), \quad (7.4)$$

où $\widehat{\text{var}}_L(\hat{T}_\tau)$ est défini par l'équation (5.11). Les mesures :

$$ET_L = \sqrt{\widehat{\text{var}}_L(\hat{\Theta})}, \quad CV_L = \frac{ET_L}{\hat{\Theta}}, \quad \Delta_2 = \frac{ET_L}{ET} - 1,$$

figurent également dans le tableau 7.2.

Tableau 7.2
Précisions estimées à partir des données 2009 pour différents indicateurs

Indicateur	$\hat{\Theta}$	ET	CV	$ET_{1,1}$	Δ_1	ET_L	CV_L	Δ_2
Taux de risque de pauvreté								
Population totale	14,6 %	0,581	4,0 %	0,580	-0,61 %	0,632	4,3 %	8,9 %
0-17 ans	18,3 %	1,310	7,2 %	1,310	-0,60 %	1,426	7,9 %	8,8 %
18-24 ans	11,9 %	1,302	10,9 %	1,301	-0,61 %	1,388	11,6 %	6,6 %
25-49 ans	10,7 %	0,585	5,5 %	0,585	-0,62 %	0,645	6,1 %	10,2 %
50-64 ans	9,9 %	0,752	7,6 %	0,752	-0,63 %	0,828	8,4 %	10,1 %
65 ans et plus	26,4 %	1,291	4,8 %	1,291	-0,62 %	1,442	5,4 %	11,6 %
Rapport interquintile de revenu	4,4	0,109	2,4 %	0,109	-0,55 %	0,118	2,7 %	8,9 %
Indice de Gini	*** ^a	***	1,7 %	***	-0,56 %	***	1,8 %	8,0 %
ERM RP	*** ^a	***	5,8 %	***	-0,60 %	***	6,4 %	9,7 %

^a Nous ne communiquons pas la valeur de ces indicateurs car ils ne sont pas encore publiés officiellement par l'OFS.

Les comparaisons entre les différentes mesures de précision que permet le tableau 7.2 conduisent aux constatations suivantes. Tout d'abord, pour les indicateurs considérés, la quasi totalité de l'estimation de l'écart-type ET de l'indicateur $\hat{\Theta}$ est contenue dans le terme $ET_{1,1}$. La différence relative Δ_1 entre ET et $ET_{1,1}$ varie en effet de -0,55 % pour le rapport interquintile de revenu à -0,63 % pour le taux de risque de pauvreté des 50-64 ans. En ne calculant que le terme dominant $ET_{1,1}$, on obtient donc généralement une excellente approximation de l'estimation de l'écart-type de $\hat{\Theta}$. Ce terme est extrêmement simple à programmer. D'autre part, la méthode d'estimation de la variance de $\hat{\Theta}$ proposée par Lavallée (2002), qui, comme le montre la formule (5.14), n'est pas sans biais même lorsque les probabilités de réponse sont connues, conduit ici à des estimations de précisions légèrement plus élevées qu'avec la méthode d'estimation que nous proposons. La différence relative Δ_2 entre ET et ET_L varie en effet de 6,6 % pour le taux de risque de pauvreté des 18-24 ans à 11,6 % pour celui des 65 ans et plus. Il découle des formules (5.7) et (5.16) que l'estimateur ET_L est asymptotiquement sans biais et égal à ET lorsque les probabilités de réponse à la grille et au questionnaire ménage q_k^b et q_k^c lors de l'année d'enquête sont égales à 1. Bien que les taux de réponse à la grille et au questionnaire ménage lors de l'année d'enquête soient assez élevés (*cf.* tableau 7.1), les différences entre ET et ET_L sont non négligeables.

Conclusion

Nous avons proposé un estimateur de la variance pour les indicateurs sur la pauvreté et l'exclusion sociale qui tient compte de la non-linéarité des estimateurs, de la non-réponse totale à différentes phases de l'enquête, du sondage indirect et du calage. Idéalement, les effets des imputations devraient également être pris en compte. Notons toutefois que notre démarche est compatible avec les exigences de Eurostat (2010) à qui de nombreux pays européens ne fournissent qu'une approximation de la variance due à l'échantillonnage et à la non-réponse

totale, ainsi que des indications minimales sur les imputations telles que le pourcentage de valeurs imputées. Nous avons modifié la méthode d'estimation de la variance en présence de non-réponse après un partage des poids proposée par Lavallée (2002) de façon à corriger le biais qu'elle induit. Notre estimateur est toujours positif dans le cas de l'enquête SILC-Suisse et est composé de trois termes très simples à programmer. De plus, en ne calculant que le premier de ces trois termes, on obtient une excellente approximation de la variance.

Remerciements

Je tiens à remercier Yves Tillé et Lionel Qualité de l'université de Neuchâtel pour les discussions fructueuses, leurs conseils et leur relecture attentive. Un grand merci également à Eric Graf, Johan Pea, Thomas Christin et Stéphane Fleury de l'Office fédéral de la statistique qui ont fourni les informations nécessaires pour le bon déroulement du projet. Je souhaite aussi remercier Philippe Eichenberger et Monique Graf de l'Office fédéral de la statistique pour leur soutien. Enfin, je suis très reconnaissante envers les arbitres et le rédacteur associé d'avoir pris le temps de relire attentivement ce travail. Leurs commentaires ont été une grande source d'améliorations.

Annexe A

Formules de linéarisation

Nous rappelons ci-dessous les formules de linéarisation de Osier (2009) dans le cas des quatre indicateurs de pauvreté considérés dans la section 7. On note \mathcal{U} la population des individus présents lors de l'année d'enquête. Conformément aux notations à la section 5, J_1 désigne la taille de \mathcal{U} , puisque \mathcal{U} correspond à la population cible du panel répondant en vague 1. Cependant, afin d'alléger les notations de cette annexe, la taille de \mathcal{U} est simplement notée J . On désigne par R_j le revenu équivalent de l'individu $j \in \mathcal{U}$ utilisé pour calculer les indicateurs de pauvreté et on note pour tout $x \in \mathbb{R}$:

$$F(x) = \frac{1}{J} \sum_{j \in \mathcal{U}} \mathbf{1}_{\{R_j \leq x\}}, \quad (\text{A.1})$$

où $\mathbf{1}_{\{R_j \leq x\}}$ est une variable indicatrice qui vaut 1 si le revenu R_j de l'individu j est inférieur ou égal à x , et 0 sinon. Les formules de linéarisation des indicateurs de pauvreté sont obtenues dans un premier temps en supposant que F est dérivable et que F' est non nulle. Comme cela n'est pas le cas, on contourne cette difficulté en approchant F par la fonction F_K définie, pour tout $x \in \mathbb{R}$, par :

$$F_K(x) = \int F(z)K(x, z)dz, \quad (\text{A.2})$$

où

$$K(x, z) = \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{(x-z)^2}{2h^2}\right], \quad (\text{A.3})$$

le paramètre h étant un paramètre de lissage.

1. Taux de risque de pauvreté

On désigne par SRP seuil de risque de pauvreté calculé à partir du revenu médian MED de la population \mathcal{U} :

$$\text{SRP} = 0,6 \times \text{MED}. \quad (\text{A.4})$$

Le taux de risque de pauvreté, noté TRP, est défini par :

$$\text{TRP} = \frac{1}{J} \sum_{j \in \mathcal{U}} \mathbf{1}_{\{R_j \leq \text{SRP}\}}. \quad (\text{A.5})$$

La linéarisée ℓ_j intervenant dans l'approximation de la variance de l'estimateur du taux de risque de pauvreté donnée dans la formule (4.1) s'écrit, pour chaque individu j :

$$\ell_j = \frac{1}{J} \left[\mathbf{1}_{\{R_j \leq \text{SRP}\}} - \text{TRP} \right] - \frac{0,6}{J} \cdot \frac{F'_K(\text{SRP})}{F'_K(\text{MED})} \cdot [\mathbf{1}_{\{R_j \leq \text{MED}\}} - 0,5]. \quad (\text{A.6})$$

Dans la section 7, le taux de pauvreté a également été estimé au sein de différentes sous-populations. La généralisation de la formule (A.6) dans le cas de sous-populations s'obtient facilement et n'est pas rappelée ici.

2. Rapport interquintile de revenu

On désigne par γ_+ le quantile d'ordre 0,8 et par γ_- le quantile d'ordre 0,2, puis on pose :

$$R = \sum_{j \in \mathcal{U}} R_j \text{ et } S(x) = \sum_{j \in \mathcal{U}} R_j \mathbf{1}_{\{R_j \leq x\}}. \quad (\text{A.7})$$

Le rapport interquintile de revenu, noté RIR, est défini par :

$$\text{RIR} = \frac{R - S(\gamma_+)}{S(\gamma_-)}. \quad (\text{A.8})$$

De façon analogue à ce qui a été appliqué pour la fonction F , on approche la fonction S par la fonction dérivable S_K définie par :

$$S_K(x) = \int S(z)K(x, z)dz. \quad (\text{A.9})$$

La linéarisée ℓ_j intervenant dans l'approximation de la variance de l'estimateur du rapport interquintile de revenu donnée dans la formule (4.1) s'écrit, pour chaque individu j :

$$\ell_j = \frac{S(\gamma_-) [R_j - Q_j(\gamma_+)] - [R - S(\gamma_+)] Q_j(\gamma_-)}{[S(\gamma_-)]^2}, \quad (\text{A.10})$$

où la quantité $Q_j(\gamma)$ est définie, pour tout quantile γ d'ordre α , par

$$Q_j(\gamma) = R_j \mathbf{1}_{\{R_j \leq \gamma\}} - \frac{S'_K(\gamma)}{JF_K(\gamma)} [\mathbf{1}_{\{R_j \leq \gamma\}} - \alpha]. \quad (\text{A.11})$$

3. Indice de Gini

L'indice de Gini, noté G , est défini par :

$$G = \frac{2M - R}{JR} - 1, \quad (\text{A.12})$$

où

$$M = \sum_{j \in \mathcal{U}} R_j \left(\sum_{j' \in \mathcal{U}} \mathbf{1}_{\{R_{j'} \leq R_j\}} \right) \quad (\text{A.13})$$

et R est défini par (A.7). La linéarisée ℓ_j intervenant dans l'approximation de la variance de l'estimateur de l'indice de Gini donnée dans la formule (4.1) s'écrit, pour chaque individu j :

$$\ell_j = \frac{JR(2U_j - R_j) - (2M - R)(R + JR_j)}{(JR)^2}, \quad (\text{A.14})$$

où la quantité U_j est définie par

$$U_j = \sum_{j' \in \mathcal{U}} R_{j'} \mathbf{1}_{\{R_j \leq R_{j'}\}} + R_j \sum_{j' \in \mathcal{U}} \mathbf{1}_{\{R_{j'} \leq R_j\}}. \quad (\text{A.15})$$

4. Écart relatif médian du risque de pauvreté

L'écart relatif médian du risque de pauvreté, noté ERMRP, est défini par :

$$\text{ERM RP} = \frac{\text{SRP} - \text{MED}^p}{\text{SRP}}, \quad (\text{A.16})$$

où on rappelle que SRP désigne le seuil de risque de pauvreté et où MED^p est le revenu médian calculé parmi les personnes ayant un revenu inférieur au seuil de pauvreté. La linéarisée ℓ_j intervenant dans l'approximation de la variance de l'estimateur du ERM RP donnée dans la formule (4.1) s'écrit, pour chaque individu j :

$$\ell_j = - \frac{\text{SRP} \cdot Y_j - \text{MED}^p \cdot W_j}{(\text{SRP})^2}, \quad (\text{A.17})$$

où W_j est défini par

$$W_j = - \frac{0,6}{F'_K(\text{MED})} \frac{1}{J} \left[\mathbf{1}_{\{R_j \leq \text{MED}\}} - 0,5 \right] \quad (\text{A.18})$$

et Y_j vérifie l'équation

$$F'_K(\text{MED}^p) Y_j = \frac{1}{2} \left\{ \frac{1}{J} \left[\mathbf{1}_{\{R_j \leq \text{SRP}\}} - F(\text{SRP}) \right] + F'_K(\text{SRP}) \cdot W_j \right\} - \frac{1}{J} \left[\mathbf{1}_{\{R_j \leq \text{MED}^p\}} - F(\text{MED}^p) \right]. \quad (\text{A.19})$$

Annexe B

Démonstration de la proposition 1

La preuve est composée de quatre parties.

1. Définition de $\widehat{\text{cov}}_{jj'}$

Pour tous longitudinaux j et j' appartenant $s_p^{A_2, t_1}$ et faisant partie respectivement des ménages k et k' lors de l'année d'enquête, on pose :

$$\widehat{\text{cov}}_{jj'} = \begin{cases} 0 & \text{si } k \neq k' \\ \frac{1}{(L_k + P_k)^2} (e'_k)^2 \frac{1 - q_k^b q_k^c}{(q_k^b q_k^c)^2} \mathbf{1}_{\{k \in s_m^{B,\tau}\}} & \text{si } k = k', \end{cases} \quad (\text{B.1})$$

de sorte que

$$\mathbb{E} \left[\widehat{\text{cov}}_{jj'} \mid s_p^{A_2, t_\tau} \right] = \text{cov}_{jj'}. \quad (\text{B.2})$$

2. Estimation de $V_{A_2}^\tau$

Si les Z_j étaient connus sur $s_p^{A_2, t_\tau}$, on pourrait estimer sans biais $V_{A_2}^\tau$ par $\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots)$. Comme ce n'est pas le cas, on estime sans biais $V_{A_2}^\tau$ par :

$$\hat{V}_{A_2}^\tau = \sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{\pi_{jj'}^{A_2} - \pi_j^{A_2} \pi_{j'}^{A_2}}{\pi_j^{A_2} \pi_{j'}^{A_2}} \frac{1}{\pi_{jj'}^{A_2}} \left[\hat{Z}_j \hat{Z}_{j'} - \widehat{\text{cov}}_{jj'} \right].$$

En effet, pour tous longitudinaux j et j' appartenant à $s_p^{A_2, t_\tau}$, on a :

$$\mathbb{E} \left[\hat{Z}_j \hat{Z}_{j'} - \widehat{\text{cov}}_{jj'} \mid s_p^{A_2, t_\tau} \right] = Z_j Z_{j'},$$

ce qui implique

$$\mathbb{E}(\hat{V}_{A_2}^\tau) = \mathbb{E}_{s_p^{A_2, t_\tau}} \left[\mathbb{E} \left(\hat{V}_{A_2}^\tau \mid s_p^{A_2, t_\tau} \right) \right] = \mathbb{E}_{s_p^{A_2, t_\tau}} \left[\tilde{V}_{A_2}^\tau(Z_1, Z_2, \dots) \right] = V_{A_2}^\tau.$$

3. Estimation de V_{NRG}^τ

On peut estimer V_{NRG}^τ par :

$$\hat{V}_{\text{NRG}}^\tau = \sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{1}{\pi_j^{A_2}} \frac{1}{\pi_{j'}^{A_2}} \widehat{\text{cov}}_{jj'}.$$

4. Formule finale

On estime $\text{var}(\hat{T}_\tau)$ par :

$$\widehat{\text{var}}(\hat{T}_\tau) = \hat{V}_{A_2}^\tau + \hat{V}_{\text{NRG}}^\tau.$$

Après quelques simplifications, il vient :

$$\widehat{\text{var}}(\hat{T}_\tau) = \tilde{V}_{A_2}^\tau(\hat{Z}_1, \hat{Z}_2, \dots) + \sum_{j \in s_p^{A_2, t_\tau}} \sum_{j' \in s_p^{A_2, t_\tau}} \frac{1}{\pi_{jj'}^{A_2}} \widehat{\text{cov}}_{jj'}. \quad (\text{B.3})$$

En utilisant (B.1) on peut simplifier le deuxième terme du membre de droite, de façon à obtenir la formule (5.19) donnée dans la proposition 1.

Bibliographie

- Ardilly, P., et Lavallée, P. (2007). Pondération dans les échantillons rotatifs : le cas de l'Enquête SILC en France. *Techniques d'enquête*, 33, 2, 149-156.
- Caron, N., Deville, J. et Sautory, O. (1998). Estimation de données issues d'enquêtes : document méthodologique sur le logiciel POULPE. Rapport technique 9806, INSEE, Paris.
- Clemenceau, A., et Museux, J. (2006). EU-SILC (Community Statistics on Income and Living Conditions): General presentation of the instrument. Dans *Proceedings of the EU-SILC Conference* (Helsinki, 6 au 8 novembre 2006).
- Deville, J.-C. (1999). Estimation de variance pour des statistiques et des estimateurs complexes : linéarisation et techniques des résidus. *Techniques d'enquête*, 25, 2, 219-230.
- Deville, J.-C., et Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- Eurostat (2003). *Règlement (CE) No 1177/2003 du parlement européen et du conseil du 16 juin 2003 relatif aux statistiques communautaires sur le revenu et les conditions de vie (EU-SILC)*. Luxembourg, Journal officiel de l'Union européenne.
- Eurostat (2010). 2008 comparative EU intermediate quality report. Rapport technique, European Commission, Luxembourg. http://circa.europa.eu/Public/irc/dsis/eusilc/library?l=/quality_assessment/comparative_quality_1/intermediate_version/_EN_1.0_&a=d.
- Graf, E. (2008). Pondérations du SILC pilote, SILC i vague 2, SILC ii vague 1, SILC i et SILC ii combinés. Rapport technique 338-0051, OFS, Neuchâtel. http://www.bfs.admin.ch/bfs/portal/de/index/infothek/erhebungen_quellen/methodenberichte.Document.105401.pdf.
- Langel, M., et Tillé, Y. (2011). Statistical inference for the quintile share ratio. *Journal of Statistical Planning and Inference*, 141, 2976-2985.
- Lavallée, P. (2002). *Le sondage indirect ou la méthode généralisée du partage des poids*. Paris : Ellipses.
- Merkouris, T. (2001). Estimation transversale dans le cas des enquêtes auprès des ménages à panels multiples. *Techniques d'enquête*, 27, 2, 189-200.
- OFS (2010). *Les conditions de vie en Suisse en 2009, Résultats de l'enquête sur les revenus et les conditions de vie (SILC)*. No de commande 1192-0900. <http://www.bfs.admin.ch/bfs/portal/fr/index/news/publikationen.Document.138922.pdf>.
- Osier, G. (2009). Variance estimation for complex indicators of poverty and inequality using linearization techniques. *Survey Research Methods*, 3, 167-195.
- Särndal, C.-E., et Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York : John Wiley & Sons, Inc.