Rapport de méthodes

Pondérations du SILC pilote

SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Département fédéral de l'intérieur DFI
Office fédéral de la statistique OFS

Neuchâtel, 2008

Statistik der Schweiz

Die vom Bundesamt für Statistik (BFS) herausgegebene Reihe «Statistik der Schweiz» gliedert sich in folgende Fachbereiche:

0 Statistische Grundlagen und Übersichten
1 Bevölkerung
2 Raum und Umwelt
3 Arbeit und Erwerb
4 Volkswirtschaft
5 Preise
6 Industrie und Dienstleistungen
7 Land- und Forstwirtschaft
8 Energie
9 Bau- und Wohnungswesen
10 Tourismus
11 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Informationsgesellschaft, Sport
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Kriminalität und Strafrecht
20 Wirtschaftliche und soziale Situation der Bevölkerung
21 Nachhaltige Entwicklung und Disparitäten auf regionaler und internationaler Ebene

Statistique de la Suisse

La série «Statistique de la Suisse» publiée par l'Office fédéral de la statistique (OFS) couvre les domaines suivants:

0 Bases statistiques et produits généraux
1 Population
2 Espace et environnement
3 Vie active et rémunération du travail
4 Economie nationale
5 Prix
6 Industrie et services
7 Agriculture et sylviculture
8 Energie
9 Construction et logement
10 Tourisme
11 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, société de l'information, sport
17 Politique
18 Administration et finances publiques
19 Criminalité et droit pénal
20 Situation économique et sociale de la population
21 Développement durable et disparités régionales et internationales

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Pondérations du SILC pilote

SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés

Auteur **Eric Graf**
Office fédéral de la Statistique

Editeur **Office fédéral de la statistique**

Neuchâtel, 2008

Préambule

L'enquête SILC est le fruit d'une intense coordination au niveau européen. L'objectif est d'assurer une bonne comparabilité internationale de l'estimation des revenus, des conditions de logement et des conditions de vie en général. Sous l'égide d'Eurostat (Office statistique des Communautés européennes), le contenu des questionnaires, les méthodes d'enquête statistique, mais aussi les futurs indicateurs sont coordonnés dans plus de 25 pays européens.

En Suisse l'enquête principale, qui a débuté en mars 2007, fut précédée de deux enquêtes pilotes en 2004 – 05 et 2005-06. Ces pré-tests quasi grandeur nature furent très étroitement coordonnés dans leur réalisation avec le Panel Suisse de Ménages (PSM). Les données SILC récoltées dans cette phase initiale du projet, bien que non utilisées au niveau international, sont diffusées communément avec celle du PSM. Ce rapport détaille et documente la manière dont ces données ont été pondérées. Les concepts relatifs à la pondération d'enquêtes panel et de la combinaison d'enquête panels sont assez ardus. Il se peut que le lecteur non averti rencontre quelques difficultés de compréhension dans certains passages. Si nécessaire, on trouvera dans la bibliographie une liste de documents traitant de ces sujets plus en détails.

Le travail présenté ici résulte d'une collaboration fructueuse avec Statistique Canada, en particulier avec Caroline Cauchon et Michel Latouche. Que ces personnes soient tout particulièrement remerciées ici.

Toute l'équipe du Panel Suisse de Ménage a également joué un rôle clé car elle a fait bénéficier l'enquête SILC de son expérience dans la réalisation d'une enquête panel dans les domaines les plus variés. Qu'ils soient tous remerciés.

Toute notre gratitude va aussi à Paul-André Salamin (METH) pour ses remarques avisées après une lecture approfondie de ce rapport.

Résumé

La structure du document est la suivante: la façon dont les données ont été collectées est brièvement décrite pour chacun des panels, puis sont définis les différents concepts liés à la pondération. Ensuite, les étapes de la pondération sont présentées, puis décrites une à une pour chacun des panels également. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite aussi séparément les pondérations au niveau ménage de celles au niveau individuel. Les deux dernières sections font l'objet de la pondération combinée du SILC_I vague 2 et SILC_II vague 1. Finalement, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes mentionnées que brièvement dans le texte principal.

Mots-clé

Rapport de méthode; méthode de segmentation; non réponse; calage; plan d'échantillonnage; combinaison de panel; partage des poids.

Complément d'information: Eric Graf, tél. 032 713 64 01
Eric.Graf@bfs.admin.ch

Réalisation: Service de méthodes statistiques, OFS

Diffusion: Office fédéral de la statistique
CH-2010 Neuchâtel
Tél. 032 713 60 60 / Fax 032 713 60 61
Order@bfs.admin.ch

Internet: http://www.statistik.admin.ch

Numéro de commande: 338-0051

Prix: gratuit

Série: Statistique de la Suisse

Domaine: 0 Bases statistiques et produits généraux

Langue du texte original: Français

Graphisme/Layout: OFS

Copyright: OFS, Neuchâtel 2008
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

ISBN: 978-3-303-00406-7

Tables des matières

1 INTRODUCTION : L'ENQUÊTE SILC PILOTE 5

2 STRUCTURE DES ENQUÊTES SILC_I ET SILC_II 7

3 PLAN D'ÉCHANTILLONNAGE ET DÉROULEMENT DE L'ENQUÊTE 8

3.1 PLAN D'ÉCHANTILLONNAGE ET TAILLES BRUTES DES SILC_I ET SILC_II 8
3.2 TAUX DE RÉPONSES ATTEINTS EN 2004 ET 2005 9
3.3 PANÉLISATION DU SILC_I 11
3.4 DÉROULEMENT DES ENQUÊTES SILC_I VAGUE 2 ET SILC_II VAGUE 1 12

4 CONCEPTS RELATIFS À LA PONDÉRATION 13

4.1 CONCEPTS RELATIFS À LA PRODUCTION DE POIDS LONGITUDINAUX 13
4.1.1 Population longitudinale cible et observée *13*
4.1.2 Unité longitudinale *13*
4.1.3 Cohabitants *14*
4.1.4 Éligibilité longitudinale *14*
4.2 CONCEPTS RELATIFS À LA PRODUCTION DES POIDS TRANSVERSAUX 14
4.2.1 Population transversale cible et observée *14*
4.2.2 Unité transversale *15*
4.2.3 Unité dans le champ de l'enquête transversalement *15*
4.2.4 Éligibilité transversale *15*
4.2.5 Unités répondantes *15*
4.3 ILLUSTRATION DES CONCEPTS PAR UN EXEMPLE 15
5.1 INTRODUCTION 18
5.2 TYPES DE PONDÉRATIONS PRODUITES 20

6 ÉTAPES COMMUNES POUR TOUS LES TYPES DE POIDS 23

6.1 MODÉLISATION DE LA NON RÉPONSE PAR LA MÉTHODE DE SEGMENTATION 23
6.1.2 La méthode de segmentation *23*
6.1.2 Segmentation et régression logistique *24*
6.1.3 Application dans le cadre des SILC pilotes *25*
6.2 POIDS INITIAL DU SILC_I ET POIDS INITIAL DU SILC_II 26
6.3 CLASSEMENT DES INDIVIDUS : ÉTAPE 2 28
6.4 AJUSTEMENT POUR LA NON-RÉPONSE – GRILLE : ÉTAPE 3 30

7 PONDÉRATION TRANSVERSALE INDIVIDUELLE 32

7.1 PARTAGE DES POIDS (UNIQUEMENT POUR SILC_I VAGUE 2) : ÉTAPE 4 32
7.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE INDIVIDUEL : ÉTAPE 5 33
7.3 CALAGE : ÉTAPE 6 34

8 PONDÉRATION TRANSVERSALE MÉNAGE 36

8.1 PARTAGE DES POIDS (UNIQUEMENT POUR SILC_I VAGUE 2) : ÉTAPE 7 36
8.2 AJUSTEMENT POUR LA NON-RÉPONSE – QUESTIONNAIRE MÉNAGE : ÉTAPE 8 37
8.3 CALAGE : ÉTAPE 9 38

9 PONDÉRATION TRANSVERSALE INDIVIDUELLE COMBINÉE SILC_I ET SILC_II 41

9.1 COMBINAISON TRANSVERSALE DE PANELS À PARTIR DES ÉCHANTILLONS TRANSVERSAUX DE CHACUN DES DEUX PANELS 41
9.2 COMBINAISON DES DEUX PANELS: ÉTAPE 11 43
9.3 CALAGE: ÉTAPE 12 44

10 PONDÉRATION TRANSVERSALE COMBINÉE MÉNAGE SILC_I ET SILC_II 46

10.1 COMBINAISON DES DEUX PANELS: ÉTAPE 13 46
10.2 CALAGE: ÉTAPE 14 47

11 CONCLUSIONS 49

12 BIBLIOGRAPHIE 50

ANNEXE A – FICHIERS DE DONNÉES POUR SILC_I ET SILC_II 52

ANNEXE B – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE NON-RÉPONSE POUR LES POIDS INITIAUX DU SILC_I 53

ANNEXE C - DÉRIVATION DES VARIABLES POUR LE CLASSEMENT DES INDIVIDUS 56

ANNEXE D – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE GRILLE POUR LES POIDS INITIAUX AJUSTÉS DU SILC_I (VAGUE 2) 58

ANNEXE E – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE INDIVIDUEL, PONDÉRATION TRANSVERSALE INDIVIDUELLE DU SILC_I (VAGUE 2) ET DU SILC_II (VAGUE 1) 60

ANNEXE F – MODÉLISATION DE LA NON-RÉPONSE AU QUESTIONNAIRE MÉNAGE, PONDÉRATION TRANSVERSALE MÉNAGE DU SILC_I ET SILC_II 63

ANNEXE G – CARACTÉRISTIQUES DES DISTRIBUTIONS DES DIFFÉRENTES PONDÉRATIONS PRODUITES 66

ANNEXE H – DESCRIPTION DÉTAILLÉE DU STATUT FINAL ATTRIBUÉ À CHAQUE ADRESSE, TAUX DE RÉPONSE 71

1 Introduction : l'enquête SILC PILOTE

L'enquête SILC (Statistics on Income and Living Conditions) est une enquête réalisée auprès des ménages et des personnes résidant de manière permanente en Suisse. Les questions traitées couvrent un large éventail de thématiques, telles que la composition du ménage, la formation, le travail, les revenus, la prise en charge des enfants, les conditions de logement, la santé, etc. L'enquête vise à collecter des données comparables au niveau européen et à établir des indicateurs sur les revenus, la pauvreté et les conditions de vie.
Elle est menée par la section EKL de l'OFS et réalisée en collaboration avec l'institut de recherche de marché M.I.S-TREND SA (basé à Lausanne).

L'enquête principale SILC en Suisse a débuté en mars 2007. Il s'agit d'un panel rotatif sur 4 ans qui devra satisfaire les exigences européennes dictées par EUROSTAT dans de nombreux domaines, en particulier dans la manière de calculer les pondérations (cf. les références [1] et [2]), pour les tailles minimales des échantillons de répondants (ménages et individus) (cf. [3]) et enfin dans la précision (calcul de la variance) des indicateurs clés à livrer chaque année.

Cette enquête principale a été précédée de deux pilotes réalisés en 2004 et 2005[1]. La première de ces enquêtes pilote fut partiellement panélisée sur l'année 2005. En parallèle un nouvel échantillon fut tiré et interrogé en 2005. Nous utiliserons les notations suivantes dans la suite du document :

- ***SILC_I vague 1***, l'échantillon interrogé en 2004
- ***SILC_I vague 2***, le sous-échantillon de SILC_I vague 1 réinterrogé en 2005
- ***SILC_II vague 1***, l'échantillon interrogé pour la première et unique fois en 2005

Le calcul des poids pour le SILC_I vague 1 (2004) a été réalisé par Statistique Canada (cf. [4]).

Le document dont il est question ici décrit de façon détaillée le processus de production des poids pour les enquêtes pilotes de 2005, à savoir, celles du SILC_I vague 2, du SILC_II vague 1 ainsi que la pondération combinée des échantillons SILC_I et SILC_II pour 2005. Il s'agit des jeux de poids suivant :

- Le poids transversal individuel du SILC_I
- Le poids transversal ménage du SILC_I
- Le poids transversal individuel du SILC_II
- Le poids transversal ménage du SILC_II
- Le poids transversal individuel du SILC_I et II combinés
- Le poids transversal ménage du SILC_I et II combinés

Vu la petite taille du sous-échantillon du SILC_I vague 1 qui fut panélisé, il fut décidé de ne pas produire de pondération longitudinale (même si cela est théoriquement et pratiquement possible).

Les questionnaires « grille »[2] des enquêtes SILC pilotes et PSM (Panel Suisse de Ménages) étaient identiques tandis qu'environ 60% du contenu des questionnaires ménage et individuel de ces enquêtes se recoupaient.
C'est pourquoi, les données récoltées pour les années 2004 et 2005 des enquêtes SILC pilote et PSM ont été livrées en commun avec des pondérations permettant de combiner

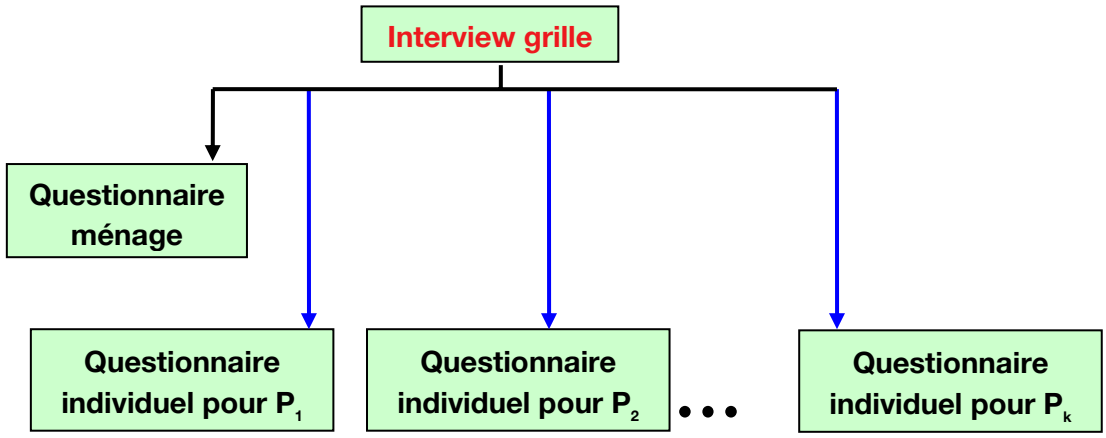
[1] Pour alléger le texte, nous dirons « réalisé en 2004 » pour l'enquête septembre 2004 à février 2005 et « réalisé en 2005 » pour celle qui dura de septembre 2005 à février 2006.
[2] Se référer au chapitre 2 pour la définition.

transversalement le PSM, SILC_I et SILC_II. On a ainsi pu gagner en précision pour des exploitations transversales.
Les pondérations décrites dans ce document sont donc, à quelques détails près, très semblables à celles qui furent effectuées par Statistique Canada pour l'année 2004 et parfaitement compatibles avec les jeux de poids calculés pour le PSM.
On documente ici le travail effectué dans le domaine de la production des pondérations pour le SILC pilote. Une bonne partie des méthodes seront reprises pour pondérer l'enquête SILC principale à partir de 2007, bien que des adaptations soient nécessaires afin de respecter les recommandations d'Eurostat et la nature d'un panel rotatif sur 4 ans.

2 Structure des enquêtes SILC_I et SILC_II

SILC est une enquête qui comporte 3 questionnaires : la grille, le questionnaire ménage et le questionnaire individuel.

Figure 1 Structure des enquêtes SILC



Le questionnaire grille consiste en une sorte de screening du ménage, il dure moins de 10 minutes et permet, entre autres, de récolter les informations sociodémographiques de base sur tous les membres du ménage ainsi que de déterminer leur statut d'éligibilité aux interviews. Ni le questionnaire ménage, ni aucun des questionnaires individuels ne peuvent être complétés avant que le questionnaire grille ne soit réalisé. En principe c'est la *personne de référence*[3] qui répond au questionnaire grille, mais si cette dernière n'est pas disponible au premier contact, un autre adulte (de 14+ ans) peut le faire.
Une fois que la grille est complétée l'interviewer peut passer aux questionnaires ménage et individuel(s). Le questionnaire ménage est de préférence rempli par la personne de référence ou par un adulte qui « connaît bien son ménage ». Toute personne de 14 ans ou plus vivant dans le ménage est éligible pour le questionnaire individuel.

A la vague 2 du SILC_I, on a tenté de contacter tous les individus qui faisaient partie des ménages qui, à la vague 1, avaient répondu au moins à la grille et dont le ménage faisait partie du sous-échantillon des ménages destinés à être panélisés. Ces individus des ménages de vague 1 sont appelés les « longitudinaux » du SILC_I. On a aussi ajouté à l'échantillon toutes les nouvelles personnes qui s'étaient jointes aux ménages des longitudinaux. Celles-ci sont appelées les « cohabitants »[4]. Les ménages qui avaient répondu au questionnaire non-réponse à la vague 1 (2004) n'ont pas été recontactés.

Une description des fichiers de données utilisés pour la pondération de la vague 2 du SILC_I et de la vague 1 du SILC_II ainsi que leur combinaison se trouve dans l'annexe A.

[3] La personne de référence doit être majeure et « bien » connaître son ménage. C'est en principe elle qui répond au questionnaire ménage année après année.
[4] Pour plus de précisions sur les concepts de longitudinaux et de cohabitants, voir au chapitre 4.

3 Plan d'échantillonnage et déroulement de l'enquête

3.1 Plan d'échantillonnage et tailles brutes des SILC_I et SILC_II

Les échantillons de SILC_I et SILC_II sont stratifiés par grande région géographique (NUTS II), proportionnellement au nombre de ménages par strate. Les adresses de l'échantillon brut furent réparties dans les proportions suivantes (recensement 2000) :

Tableau 2 Stratification de l'échantillon brut

Strates *h*	Cantons	Proportion d'adresses P_h (en %)
1: Région lémanique	VD, VS, GE	18.22
2: Espace Mitteland	BE, FR, SO, NE, JU	22.92
3: Suisse du nord-ouest	BS, BL, AG	13.86
4: Zurich	ZH	18.22
5: Suisse orientale	GL, SH, AR, AI, SG, GR, TG	13.70
6: Suisse centrale	LU, UR, SZ, OW, NW, ZG	8.75
7: Tessin	TI	4.33
	Total	100

Le tirage est donc **uniquement** proportionnel au nombre de ménages par grande région et **ne tient pas compte** du nombre moyen de personnes dans les ménages par région. On n'augmente pas la taille du Tessin.
Le cadre de sondage est le SRH[5] actualisé trimestriellement en mars, juin, septembre et décembre. Les SILC_I a été tiré dans le SRH de 02/04, les SILC_II dans le SRH 02/05. Les tailles des strates au moment du tirage pour SILC_I et SILC_II étaient les suivantes :

Tableau 3 Tailles (SRH) N_h des strates h (h = 1,..,7) au moment des deux tirages

Strates *h*	SILC_I, 2004 N_h^{04}	SILC_II, 2005 N_h^{05}
1: Région lémanique	648'590	648'176
2: Espace Mitteland	784'266	785'951
3: Suisse du nord-ouest	455'833	456'925
4: Zurich	587'850	599'145
5: Suisse orientale	493'606	495'755
6: Suisse centrale	306'605	308'626
7: Tessin	160'123	159'622
TOTAL	**3'436'873**	**3'454'200**

La base de sondage SRH est sujette aux erreurs suivantes :

- *sous-couverture* : certains ménages ne figurent pas dans l'annuaire au temps de la sélection. Par exemple les ménages dont aucun de leurs numéros sont inscrits, ou les ménages non joignables par téléphone. Ce problème peut produire des biais, c'est-à-dire des différences entre les estimations produites à partir de la population

[5] Registre pour les enquêtes auprès des ménages de l'OFS.

observée (l'enquête SILC) et celles qu'on aurait produites à partir de la population cible (l'ensemble des ménages privés résidant en Suisse à un certain moment fixé dans le temps). On estime, qu'en Suisse, 98.5% des ménages privés sont atteignables par téléphone. En 2004 et 2005, au moment des tirages des échantillons pour les deux pilotes de SILC, on estime que la base de sondage couvrait environ 93% des ménages privés atteignables par téléphone.

- *doublons :* malgré les procédures rigoureuses de nettoyage du SRH visant à ne garder qu'un numéro de téléphone par ménage, certains ménages figurent plus d'une fois dans la base de sondage. Le problème conduit à une détermination erronée de certaines des probabilités initiales de sélection. Malgré cela, on ne calcule pas de facteur de correction pour les ménages avec plusieurs lignes téléphoniques, l'information serait disponible, mais l'effet est négligeable.
- *sur-couverture* : sélection d'unités hors de la population cible (entreprises, homes, prisons, ménages collectifs, résidences secondaires, …). Notons que pour un panel, ce problème n'est encouru qu'à la vague 1 et qu'en principe ces adresses-là sont considérées comme hors échantillons (*non sample cases*).

Le tableau 4 indique la répartition de l'échantillon brut, i.e. le nombre d'adresses injectées dans l'enquête par strate, noté S_h.

Tableau 4 Tailles S_h de l'échantillon brut par strates. Les S_h représentent le nombre d'adresses injectées dans l'enquête. Ces adresses ne correspondent pas forcément à des ménages privés, mais pour la plupart c'est le cas. Nous renvoyons le lecteur à l'annexe H pour une description détaillée du statut final attribué à chaque adresse.

Strates h	SILC_I, v1 2004	SILC_I, v2 2005 S_h^I	SILC_II, v1 2005 S_h^{II}
1: Région lémanique	737	267	370
2: Espace Mitteland	915	360	467
3: Suisse du nord-ouest	574	222	282
4: Zurich	708	277	370
5: Suisse orientale	560	205	281
6: Suisse centrale	338	146	177
7: Tessin	168	73	88
TOTAL	**4'000**	**1'550**	**2'035**

Les tableaux 3 et 4 permettent de calculer les taux de sondage et les poids d'échantillonnage par strate. Pour cela, nous renvoyons le lecteur à la section 6.2.

3.2 Taux de réponses atteints en 2004 et 2005

Les tableaux 5 et 6 suivants résument les taux de réponse obtenus par rapport à l'échantillon brut, i.e. par rapport au nombre d'adresses injectées dans les enquêtes indépendamment de leur statut final (validité).

Tableau 5 Taux de réponse aux différents niveaux de collecte des données pour le SILC_I en 2004 (vague 1) et 2005 (vague 2 pour la partie panélisée (source : statistiques finales MIS-Trend SA). On peut constater l'impact négatif sur les taux de réponses en vague 2 du questionnaire papier qui fut administré à un sous-échantillon des répondants à la vague 1.

		SILC_I vague 1, 2004				SILC_I vague 2, 2005				Vague 2, 2005, sans questionnaire papier				Vague 2, 2005, avec questionnaire papier			
Niveau ménages	Echantillon brut (nombre d'adresses injectées)	4000	100%			1550	100%			743	100%			807	100%		
	Nombre de grilles réalisées	2287	57%	100%		1103	71%	100%		570	77%	100%		533	66%	100%	
	Grilles + quest. Ménage réalisés	2199	55%	96%	100%	1064	69%	96%	100%	558	75%	98%	100%	506	63%	95%	100%
	Nombre de ménages complets[6]	2116	53%	93%	96%	1042	67%	94%	98%	550	74%	96%	99%	492	61%	92%	97%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	4441	100%			2160	100%			1100	100%			1060	100%		
	Nombre de quest. Individuel réalisés	3498	79%			1683	78%			862	78%			821	77%		

[6] Un ménage est dit « complet » si les questionnaires grille, ménage et au moins un questionnaire individuel ont été réalisés.

Tableau 6 Taux de réponse aux différents niveaux de collecte des données pour le SILC_II (vague 1) en 2005 (source : statistiques finales M.I.S-Trend SA). Les taux de réponse furent similaires à ceux de la vague 1 du SILC_I l'année d'avant

		SILC_II vague 1, 2005			
Niveau ménages	Echantillon brut (nombre d'adresses injectées)	2035	100%		
	Nombre de grilles réalisées	1263	62%	100%	
	Grilles + quest. Ménage réalisés	1204	59%	95%	100%
	Nombre de ménages complets	1132	56%	90%	94%
Niveau individus	Nombre d'individus éligibles pour q. indiv.	2465	100%		
	Nombre de quest. Individuel réalisés	1797	73%		

3.3 Panélisation du SILC_I

En 2004, l'échantillon du SILC_I fut divisé en 4 sous-échantillons :

1. Un premier destiné à n'être interrogé que par CATI et seulement en 2004,
2. Un deuxième destiné à être interrogé par CATI en 2004 et auquel il fut adressé un questionnaire papier complémentaire volumineux après la fin de la période du terrain CATI (en printemps 2005)
3. Un troisième destiné à être interrogé par CATI en 2004 puis réinterrogé en 2005, donc panélisé sur deux ans.
4. Un quatrième destiné à être interrogé par CATI en 2004, auquel il fut adressé un questionnaire papier comme au deuxième sous-échantillon, et qui fut encore réinterrogé en 2005 comme le troisième sous-échantillon.

Les quatre sous-échantillons furent additionnés de paquets de réserves d'adresses nécessaires pour atteindre au moins 500 ménages interrogés par CATI dans chaque groupe en 2004.
Comme le montre le tableau 5, à la vague 2 en 2005, le quatrième sous-échantillon montra un taux de réponse bien inférieur à celui du troisième (effet négatif du questionnaire papier). En 2005, la condition fut à nouveau d'atteindre environ 500 ménages panélisés pour les sous-échantillons 3 et 4. Nous avons donc créé des paquets de réserves à partir de l'ensemble des ménages des sous-échantillons 1 et 2 qui avaient au moins complété un questionnaire grille en 2004. Les paquets de réserves étaient des sous échantillons aléatoire simple de ce dernier ensemble.
Comme les ménages des quatre sous-échantillons de 2004 n'étaient pas informés du fait d'être panélisés ou pas, l'usage des paquets de réserves ainsi construits n'a pas posé de problèmes.

3.4 Déroulement des enquêtes SILC_I vague 2 et SILC_II vague 1

L'échantillon brut des enquêtes pilotes SILC_I vague 2 et SILC_II vague 1 en 2005 fut débloqué en trois vagues[7] d'envoi afin de garantir un traitement le plus uniforme possible des adresses par l'institut de sondage. Les paquets de réserve d'adresses furent activés à la mi-novembre. Les rappels de refus commencèrent début novembre. Il n'y eut pas d'enquête non réponse en 2005 comme cela avait été le cas pour SILC_I vague 1 en 2004.
Pour une description plus précise et complète du déroulement des enquêtes pilotes, nous revoyons le lecteur aux références [5] et [6].

[7] 1er envoi fin août, 2ème envoi le 5 septembre, 3ème envoi le 19 septembre.

4 Concepts relatifs à la pondération

On a ici brièvement résumé la terminologie et la philosophie relatives à la manière de construire nos pondérations. Les lecteurs non familiers avec ces concepts peuvent se référer, par exemple, à [9], [15], [16], [17] et [18] pour étudier le sujet de manière plus approfondie.

4.1 Concepts relatifs à la production de poids longitudinaux

Aucune pondération longitudinale n'a été produite pour le SILC pilote, même si le SILC_I a été partiellement panélisé. Néanmoins les données ont été traitées dans une optique visant à produire tant des poids transversaux que longitudinaux. Comme l'enquête SILC principale est un panel rotatif, nous avons traité les pilotes comme des débuts de panel.

4.1.1 Population longitudinale cible et observée

La *population longitudinale cible* est composée de tous les individus vivant dans des ménages privés de Suisse en septembre 2004 pour le SILC_I, en septembre 2005 pour le SILC_II, date à laquelle la collecte des données a commencé. La base de sondage utilisée est le SRH (voir chapitre 3).
La *population longitudinale observée* est composée de tous les individus vivant dans des ménages privés de Suisse qui avaient un numéro de téléphone valide, inscrit et répertorié dans le SRH 02/04 pour le SILC_I, respectivement dans le SRH 02/05 pour le SILC_II, date à laquelle l'échantillon a été sélectionné.

4.1.2 Unité longitudinale

À la sélection de l'échantillon, toutes les personnes faisant partie des ménages choisis et étant âgées de 14 ans et plus lors de la sélection, deviennent des membres de l'échantillon longitudinal, indépendamment du fait qu'ils soient répondants ou non répondants au questionnaire individuel, et sont appelés « longitudinaux ». Ces personnes sont considérées comme des membres de l'échantillon longitudinal pendant la durée de vie du panel.
En pratique, la règle de suivi des longitudinaux était que leur ménage ait au moins répondu au questionnaire grille en vague 1. Par souci de simplification on a, dans les procédures de pondérations, appelé longitudinaux, les individus de l'échantillon longitudinal qui vivaient dans des ménages ayant au moins répondu à la grille en vague 1.

Si une personne longitudinale déménage, décède ou quitte la population cible, elle fera toujours partie de la population longitudinale cible et aura un poids longitudinal. En effet, les totaux de contrôle utilisés pour le calage considèrent toutes les unités de la population de la vague 1 sans égard par rapport à ce qui peut leur arriver dans le futur. On se doit donc d'être conséquent et d'assigner un poids à toutes les unités même si elles n'ont aucune utilité analytique. De plus, il est avantageux que toutes les estimations produites soient cohérentes avec celles provenant d'autres sources. De cette manière, il est possible d'évaluer la représentativité de l'échantillon au fil du temps.

Aucune autre personne[8] ne devient membre de l'échantillon longitudinal du panel. Ainsi, l'échantillon longitudinal est constitué au moment de l'introduction du panel, à la vague 1, et reste tel quel pendant toute la durée du panel. La population ciblée par l'échantillon longitudinal demeurera la même pour toute la durée de vie du panel. Du point de vue longitudinal, l'unité d'intérêt est l'individu; le ménage pouvant difficilement servir d'outil d'analyse longitudinale à cause de sa nature dynamique dans le temps. De plus, le calcul d'un poids longitudinal au niveau du ménage demanderait un développement théorique supplémentaire. En effet, un grand nombre de définitions de ménage longitudinal sont à considérer avant de faire un choix.

4.1.3 Cohabitants

On appelle *cohabitants* les individus qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante. Les cohabitants sont classés en deux catégories:

- les cohabitants *initialement présents*, qui faisaient partie de la population transversale cible au moment de la sélection du panel,
- les cohabitants *initialement absents*, qui se sont joints à la population transversale cible entre la première vague et la vague courante. Ces derniers sont des immigrants

4.1.4 Éligibilité longitudinale

Une personne de l'échantillon original est éligible pour la *pondération longitudinale* si elle faisait partie de la population cible lors de la sélection de l'échantillon original (en 2004 pour le SILC_I) et a rempli le questionnaire individuel pour l'année courante.
Notons que cette règle présente l'inconvénient pour les utilisateurs de données qu'un individu longitudinal non répondant en vague 1 reçoit un poids de zéro en vague 1, mais s'il répond à la vague 2, il aura un poids longitudinal en vague 2…

4.2 Concepts relatifs à la production des poids transversaux

4.2.1 Population transversale cible et observée

La *population transversale cible* pour la vague 2 du SILC_I est composée des ménages privés de Suisse existant en septembre 2005.
La *population transversale observée* pour la vague 2 du SILC_I est composée des ménages privés vivant en Suisse en septembre 2005 et dont au moins un membre faisait partie de la population longitudinale observée. En particulier, cette population exclut donc les ménages composés exclusivement d'immigrants qui sont arrivés en Suisse après septembre 2004 ainsi que certaines personnes qui étaient institutionnalisées ou non sélectionnables pour une autre raison en septembre 2004.

Pour la vague 1 du SILC_II, la *population transversale cible* est composée des ménages privés de Suisse existant en septembre 2005.
La *population transversale observée* est composée des ménages privés vivant en Suisse en septembre 2005 et sélectionnés dans l'échantillon à enquêter.

[8] Pour un panel d'une durée indéterminée on devrait décider quel statut donner aux enfants nés de longitudinaux après la sélection du panel. Mais cette question ne se pose pas pour un panel rotatif sur 4 ans comme le sera l'enquête SILC principale.

4.2.2 Unité transversale

Du point de vue transversal, on s'intéresse aussi bien aux individus âgés de 14 ans et plus qu'aux ménages. Le SILC interviewe toutes les personnes habitant avec au moins un individu longitudinal. L'échantillon transversal pour une année donnée est alors constitué de toutes les personnes longitudinales dans le champ de l'enquête à l'année de référence correspondante, et de toutes les personnes qui habitent avec elles au moment de l'entrevue et qui sont dans le champ de l'enquête.

4.2.3 Unité dans le champ de l'enquête transversalement

Un ménage de l'échantillon est considéré comme étant dans le champ de l'enquête pour la vague en cours s'il est situé dans une des sept grandes régions échantillonnées et fait partie de la population cible.
Toute personne longitudinale qui est dans le champ de l'enquête transversalement fera aussi partie de l'échantillon transversal.

4.2.4 Éligibilité transversale

Une personne dans l'échantillon est éligible à la *pondération transversale* si elle fait partie de la population cible pour l'année de référence traitée et a répondu au questionnaire individuel pour l'année courante. De même, un ménage dans l'échantillon est éligible à la pondération transversale s'il fait partie de la population cible pour l'année de référence traitée et a répondu au questionnaire ménage ainsi qu' (obligatoirement) à la grille pour l'année courante.

4.2.5 Unités répondantes

Une personne est considérée comme *répondante* si elle a complété le questionnaire individuel.
Un ménage répondant est considéré comme tel s'il a complété la grille et le questionnaire ménage.

Il arrive qu'un ménage n'ait complété que la grille et son questionnaire ménage, et aussi, pour un nombre restreint de cas, qu'un ménage ait complété la grille ainsi qu'un ou plusieurs questionnaires individuels sans avoir complété le questionnaire ménage.

4.3 Illustration des concepts par un exemple

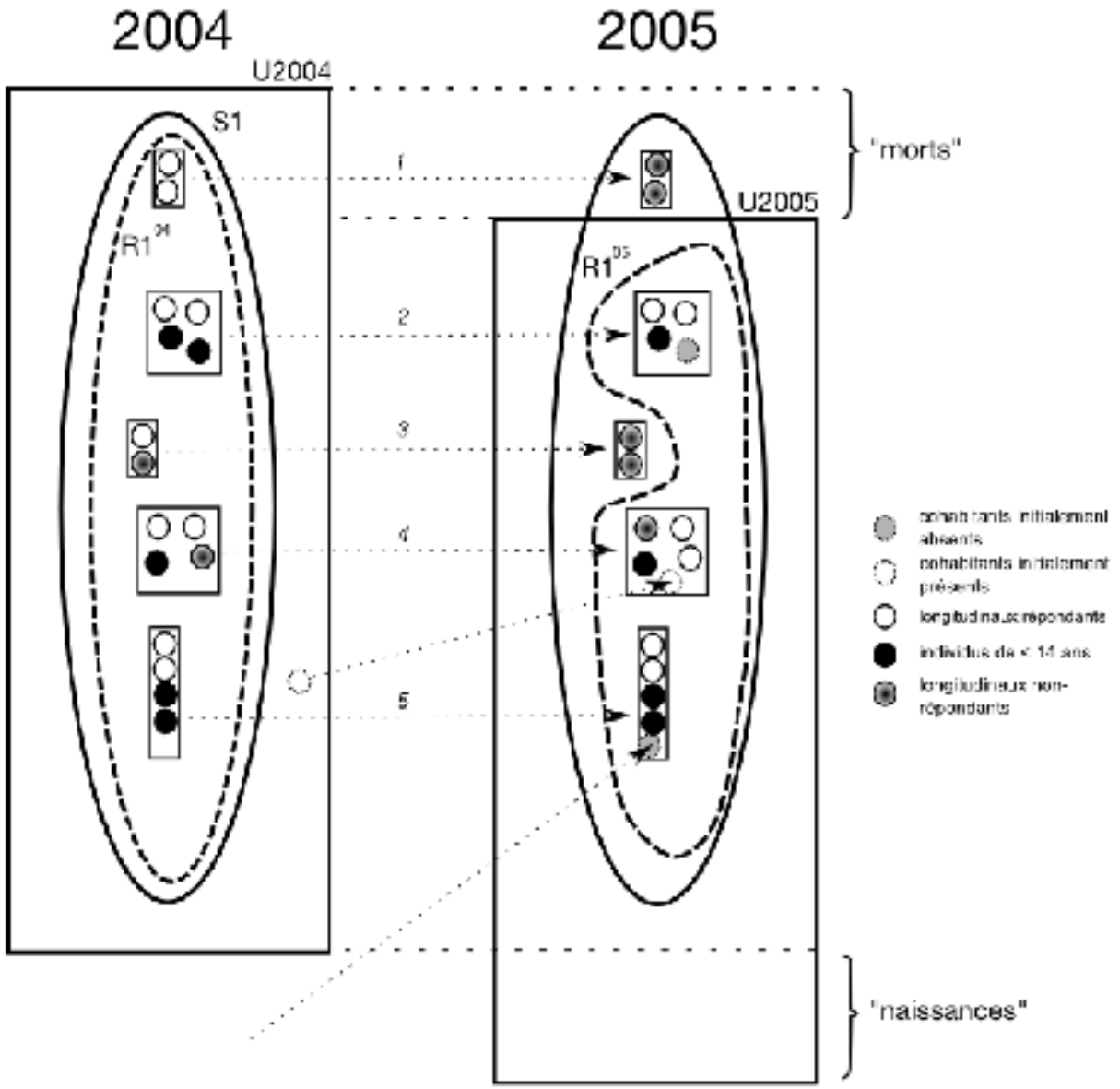
La figure 7 permet de visualiser les définitions et concepts listés dans ce chapitre. Les cercles représentent des individus et les rectangles les ménages les abritant. On a considéré 5 cas de figure courants (flèches numérotées dans la figure) de transitions entre deux années consécutives que nous commentons ci-dessous :

1. Tous les longitudinaux d'un ménage quittent la population transversale cible de l'année courante. Ce sont des « morts » au sens large car ces individus ne sont pas forcément décédés. Ces individus gardent un poids longitudinal en 2005, mais ils n'auront pas de poids transversal.
2. Un des enfants du ménage devient éligible pour l'interview individuelle. Strictement cet enfant doit être traité comme un individu initialement absent car il ne faisait pas partie de la population cible en 2004. Dans ce cas, à partir de 2005, s'il répond, cet individu est éligible pour la pondération transversale mais pas pour la pondération

longitudinale. Cependant, comme l'a fait le Panel Suisse de Ménage dès sa vague 5, on peut décider de considérer cet enfant de longitudinaux comme un longitudinal lui aussi dès qu'il entre dans l'âge d'être interviewé. Il est alors éligible pour les pondérations longitudinale et transversale. Cette stratégie n'est théoriquement pas tout à fait correcte, elle risque d'inclure un biais dans les analyses, mais elle permet de continuer de suivre des ménages dont l'enfant (en 2004) resterait le seul individu répondant du ménage à la vague courante.

3. Les individus d'un ménage répondant en vague 1 ne répondent plus aux vagues subséquentes mais ils restent dans la population transversale cible. Dans ce cas ils ne sont plus éligibles pour aucune des pondérations.
4. Un individu longitudinal non répondant en vague 1 se met à répondre à une vague subséquente. Cet individu aura un poids transversal et longitudinal.
   Un autre individu longitudinal, répondant en vague 1, ne répond pas à une vague subséquente mais il reste dans la population transversale cible. Cet individu n'aura ni poids transversal, ni poids longitudinal.
   Un individu qui ne faisait pas partie de l'échantillon S1, mais qui faisait partie de la population transversale cible en vague 1, se joint au ménage. C'est donc un cohabitant initialement présent, il aurait pu être sélectionné pour faire partie de l'enquête en vague 1. Cet individu aura un poids transversal (s'il répond) mais pas de poids longitudinal.
5. Un individu qui ne faisait pas partie de la population transversale cible en vague 1 se joint au ménage. Cet individu aura un poids transversal (seulement s'il répond) mais pas de poids longitudinal.

Figure 7 Illustration des concepts de base des pondérations longitudinale et transversale.



U2004	Population transversale cible du SILC_I en 2004 = population longitudinale cible du SILC_I en 2005
U2005	Population transversale cible du SILC_I en 2005.
S1	Échantillon brut du SILC_I
R1^{04}	Échantillon des personnes dans des ménages ayant répondu au moins au questionnaire grille en 2004, SILC_I (= échantillon net).
R1^{05}	Échantillon des personnes dans des ménages ayant répondu au moins au questionnaire grille en 2005, SILC_I (= échantillon net).
''Morts''	Individus ayant quitté la population transversale cible de l'année courante (individus décédés, partis à l'étranger, institutionnalisés,…)
''Naissances''	Individus qui ne faisaient pas partie de la population cible l'année précédente (bébés nouveaux né, immigrants venus s'établir en Suisse,…)

5 Étapes de la pondération

5.1 Introduction

Pour toutes les pondérations, les individus sont d'abord classés selon leur statut et leurs réponses aux différents questionnaires des pilotes SILC. On procède ensuite à un ajustement pour la non-réponse à la grille.

Pour la *pondération transversale individuelle*, un partage des poids est d'abord effectué, puis les poids des longitudinaux et des cohabitants[9] sont ajustés pour la non-réponse au questionnaire individuel. Enfin, un calage sur marge ajuste les poids pour que les totaux des effectifs estimés correspondent à ceux de la population de la Suisse en 2005.

Pour la *pondération transversale ménage*, un partage des poids est fait, puis les poids sont ajustés pour la non-réponse au questionnaire ménage. Enfin, un calage intégré ajuste les poids pour que le nombre total de personnes soit équivalent à l'effectif de la population de la Suisse en 2005, tout en contraignant les poids des personnes d'un même ménage à rester égaux.

Dans tous les ajustements pour la non-réponse on utilise la méthode de l'analyse par segmentation. De plus, pour la pondération individuelle aucun ajustement n'est effectué pour la non-réponse au questionnaire ménage. L'ajustement pour la non-réponse au questionnaire individuel est donc fait sur la base de tous les individus des ménages ayant répondu à la grille. Enfin, pour la pondération transversale individuelle, les longitudinaux et les cohabitants ont dû être modélisés ensemble. Les raisons de ces choix sont expliquées en détails dans les sections correspondantes du document. Précisions aussi que la distribution des poids n'a pas été tronquée, par exemple, aux 2^{e} et 98^{e} percentiles.

Les calages sur marges ont été effectués à l'aide d'une macro SAS développée par Jean-François Naud et Caroline Cauchon de Statistique Canada pour le Panel Suisse de Ménage. Plus précisément, la méthode utilisée est un calage par régression généralisée qui se base sur une fonction de distance G_k définie par (voir [10], [19]):

$$G_k(d_k, w_k) = \frac{(w_k - d_k)^2}{2\, d_k q_k}$$

où w_k est le poids après calage pour l'unité k, d_k le poids avant calage, et q_k est un facteur d'ajustement indépendant des d_k. On peut poser $q_k = 1$ pour toutes les observations. Notons que $G_k(d_k, w_k)$ peut produire des poids négatifs. On peut obtenir les mêmes résultats avec la macro CALMAR développée par l'INSEE en appliquant la méthode « linéaire ». La plupart des enquêtes de l'OFS utilisent (CALMAR avec) la méthode « raking ratio », justement pour éliminer tout risque d'obtenir des poids négatifs. Cette dernière méthode présente par contre d'autres désavantages, on a par exemple constaté empiriquement que pour SILC et PSM, elle produisait plus de poids dans l'intervalle]0,1] ainsi que des poids très élevés.

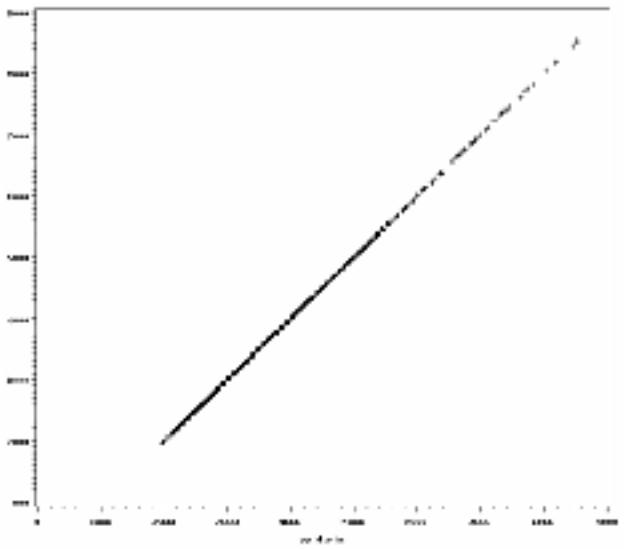
[9] Voir le chapitre 4 : «Concepts relatifs à la pondération» pour la définition des cohabitants.

Les deux méthodes sont défendables. Pour les deux enquêtes SILC pilotes, nous avons choisi de travailler avec la méthode linéaire pour les deux raisons principales suivantes :

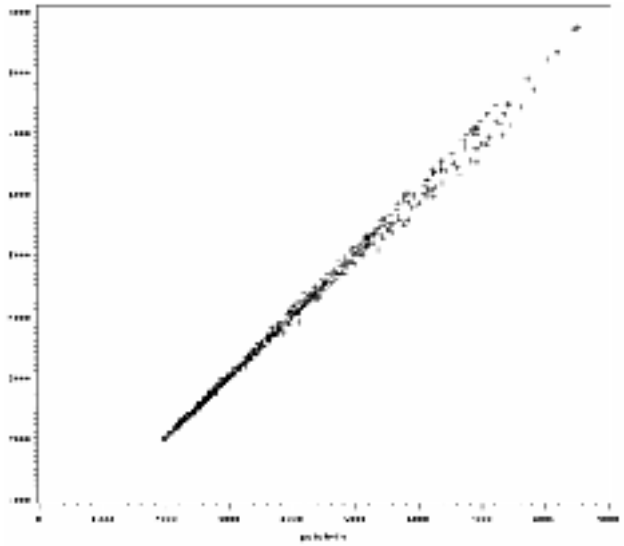
1. Afin de garantir la comparabilité et la compatibilité avec les pondérations produites pour le PSM,
2. Au niveau ménage, la macro développée par Statistique Canada permet d'effectuer facilement un calage intégré. Le temps nous a manqué pour nous assurer que CALMAR offre la même possibilité.

La figure 8 permet de se faire une idée de la différence qu'on obtient dans les poids finaux selon trois différentes fonctions de distance choisies dans CALMAR pour le poids transversal individuel du SILC_II vague 1.

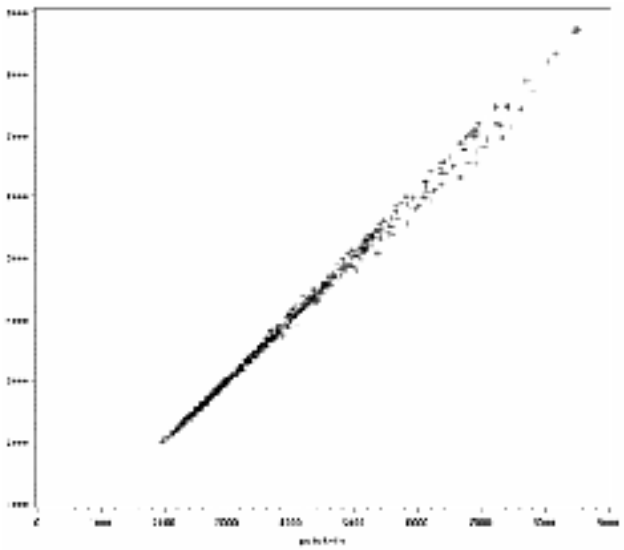
Figure 8 Comparaison de trois fonctions de distances pour le calage au niveau individuel sur les 1797 observations à pondérer du SILC_II vague 1. En abscisse dans les trois graphiques, on a le poids calé calculé avec la macro de Statistique Canada, en ordonnée le poids calé fournit par CALMAR en utilisant la méthode dite « linéaire » (graphique 1), celle du « raking ratio » (graphique 2) et la méthode « logit » (graphique 3). On vérifie que pour obtenir les mêmes poids finaux avec CALMAR qu'avec la macro de Statistique Canada, il faut appliquer la méthode « linéaire ».



1. Méthode « linéaire »



2. Méthode « raking ratio »



3. Méthode « logit »

Pour la pondération transversale combinant l'échantillon du SILC_I vague 2 avec celui du SILC_II vague 1, il y a deux manières de procéder :

1. On peut combiner les panels en une seule fois après l'ajustement à la grille[10], c'est-à-dire à partir des échantillons longitudinaux. C'est ainsi que les méthodologues de Statistique Canada ont procédé pour combiner les enquêtes PSM _I vague 6, PSM_II vague 1 et SILC_I vague 1 réalisées en 2004.
2. L'alternative est de combiner de manière transversale (à partir des échantillons transversaux), soit après l'ajustement pour la non-réponse au questionnaire individuel pour la pondération transversale individuelle et après l'ajustement pour la non-réponse au questionnaire ménage pour la pondération transversale ménage.

C'est la deuxième manière de faire qui est appliquée et décrite dans ce document. Elle est similaire à celle utilisée dans l'enquête SILC dans d'autres pays (voir [15]).
Précisons encore que les enquêtes sont combinées au niveau des grandes régions, les facteurs d'allocation variant d'une région à l'autre. On calcule donc sept différents facteurs d'allocation (un par région) au lieu d'en calculer un seul au niveau de la Suisse.

Pour une discussion plus détaillée des deux alternatives pour combiner les panels, on peut se rapporter aux références [12],[13] et [15].
Le contenu des fichiers de données utilisés pour construire les pondérations sont décrits dans l'annexe A.

5.2 Types de pondérations produites

Pour la vague 2 du SILC_I, deux pondérations ont été produites :

1 – poids transversal ménage
- *représentant la population des ménages en 2005 (vague 2)*

2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2005 (vague 2)*

Comme dit précédemment, on a renoncé à produire une pondération longitudinale au vu de la petite taille de l'échantillon panélisé.

Pour la vague 1 du SILC_II, deux types de pondérations ont également été produites :

1 – poids transversal ménage
- *représentant la population des ménages en 2005 (vague 1)*

2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2005 (vague 1)*

Pour les échantillons combinés SILC_I vague 2 & SILC_II vague 1, les deux mêmes types de pondérations ont été produites. C'est en principe ces dernières qui sont utilisées pour les analyses de données puisqu'elles permettent d'utiliser toutes les données récoltées en 2005 et donc bénéficier d'une taille d'échantillon plus grande, ce qui implique une meilleure précision.

1 – poids transversal ménage
- *représentant la population des ménages en 2005*

2 – poids transversal individuel
- *représentant la population des individus de 14+ ans en 2005*

Les trois premières étapes de la pondération sont communes aux deux types de poids, tandis que les suivantes sont spécifiques à chacun. La figure 9 résume le processus de la

[10] Statistique Canada se base sur [14] pour faire la combinaison avec les longitudinaux seulement.

pondération de la vague 2 du SILC_I et de la vague 1 du SILC_II, alors que la figure 10 traite du processus de pondération des échantillons combinés.

Figure 9 Étapes de la pondération du SILC_I vague 2. Les étapes numérotées entre parenthèses n'ont pas été réalisées pour la pondération du SILC_II vague 1 car il n'y eut d'une part pas d'enquête non réponse conduite en fin du terrain CATI en 2005, et que d'autre part, l'ajustement pour la non réponse à la grille et le partage des poids n'ont pas lieu d'être lors de la pondération d'une vague 1. Entre parenthèses, dans les cases, on a toujours le nom du jeu de poids créé à l'étape considérée.

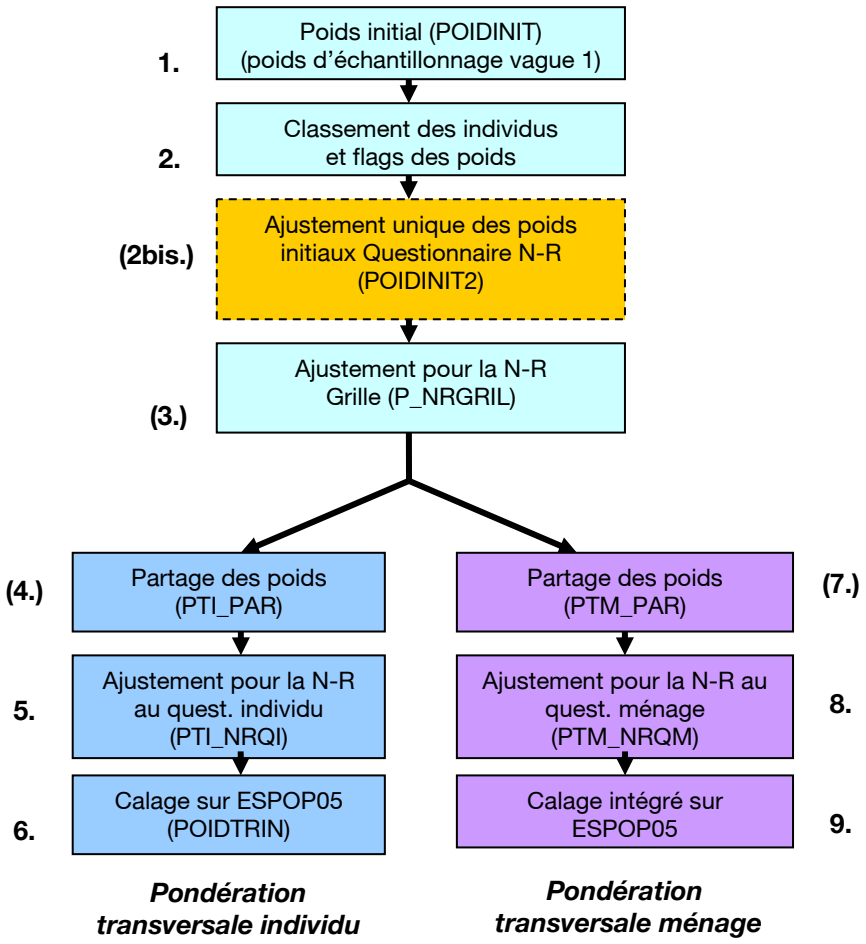
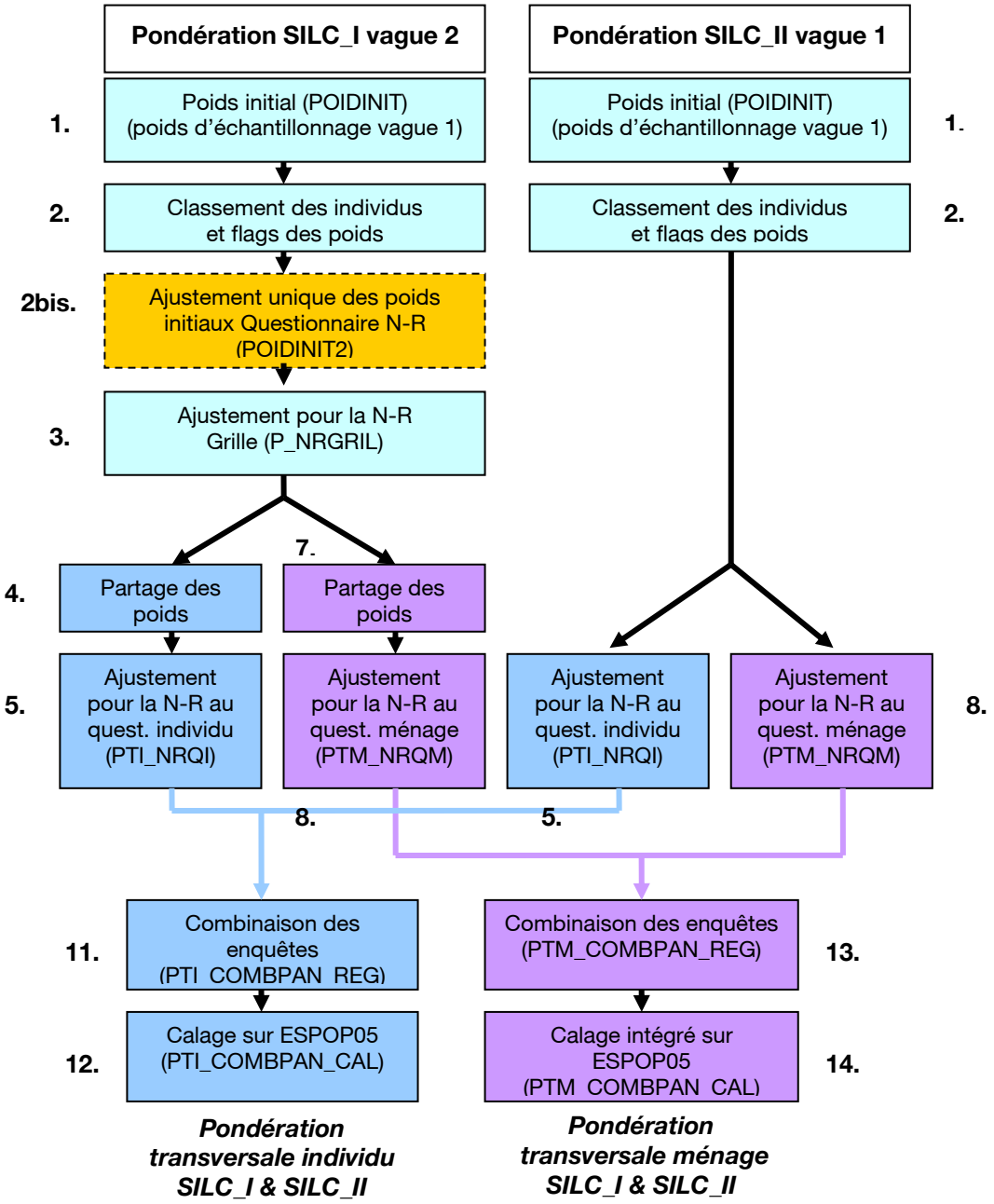


Figure 10 Étapes de la pondération permettant d'agréger l'échantillon du SILC_I vague 2 avec celui du SILC_II vague 1. Il y a une étape supplémentaire comparativement à la pondération séparée par enquête. Cette étape supplémentaire est la combinaison des échantillons[11]. Il faut combiner les échantillons pour chaque pondération (ménage et individu). Les étapes allant du calcul des poids initiaux jusqu'à l'ajustement pour la non réponse au niveau des individus (ou des ménages) sont les mêmes que celles calculées lors de la pondération séparée des deux enquêtes. Entre parenthèses, dans les cases, on a toujours le nom du jeu de poids créé à l'étape considérée.



[11] Dans l'enquête SILC principale, dès 2007, il s'agira de combiner jusqu'à quatre enquêtes panel.

6 Étapes communes pour tous les types de poids

Tout en se rapportant aux figures 9 et 10, cette section décrit les étapes 1 à 3 communes tant à la production des poids aux niveaux individu que ménage. On a procédé de la même manière pour pondérer la vague 2 du SILC_I et la vague 1 du SILC_II à la différence que le SILC_I dispose d'un ajustement supplémentaire pour la non réponse (étape 2bis). Après une brève introduction à la méthode de segmentation, on traite du calcul du poids initial, puis de la façon dont les individus sont classés selon leur statut et leur réponse et, enfin, de l'ajustement pour la non-réponse à la grille.

6.1 Modélisation de la non-réponse par la méthode de segmentation

6.1.2 La méthode de segmentation

Une fois les poids initiaux calculés, if s'agit de calculer un facteur d'ajustement pour la non-réponse (totale) de chaque unité (ménage ou individu selon les cas) considérée comme répondante. Le facteur d'ajustement pour la non-réponse appliqué à chaque unité répondante est défini comme étant l'inverse du taux de réponse du groupe homogène de réponse (GHR) dans lequel l'unité se situe. Les GHR sont formés en regroupant ensemble des unités dont les caractéristiques expliquant la non-réponse sont similaires. Si les GHR sont définis de telle sorte que la non-réponse soit complètement aléatoire à l'intérieur d'un groupe, alors le biais dû à la non-réponse est négligeable (voir [20], [21]). Pour trouver un ajustement adéquat pour gonfler les poids des unités répondantes afin que ces dernières soient représentatives de l'ensemble des unités (répondantes et des non-répondantes), on procède à la modélisation de la non-réponse.

La première étape de la modélisation de la non-réponse consiste à déterminer quelles sont les variables explicatives qui doivent être considérées dans le modèle afin de bien prédire la variable dichotomique de réponse. Le choix des variables est souvent assez restreint car on doit détenir l'information tant pour les unités répondantes que pour les non-répondantes. Il s'agit de dresser une liste des variables disponibles pouvant décrire le processus de non-réponse. La nature de cette liste est variable selon les thèmes abordés par l'enquête, le mode d'interview, le plan d'échantillonnage, etc. Tout facteur susceptible d'expliquer la non-réponse doit être pris en compte dans la mesure du possible. Les variables peuvent êtres des grandeurs mesurées par l'enquête elle-même ou des informations auxiliaires qu'on peut obtenir par d'autres biais (par exemple des registres) sur les unités répondantes et non-répondantes.
Dans une enquête panel comme celle qui nous intéresse ici, on a dès la vague deux, un choix de variables beaucoup plus grand pour modéliser la non-réponse aux questionnaires ménage (voir paragraphe 8.2) ou individuel (voir paragraphe 7.2) car on peut alors se reposer sur des informations récoltée en vague 1.
Notons encore que les variables catégoriques dont le nombre de modalités (m) est supérieur à 2 doivent être recodées en m variables dichotomiques.

La méthode utilisée pour la modélisation de la non-réponse est l'analyse par segmentation (voir [8]). Cette méthode est utilisée pour tous les ajustements nécessaires dus à la non réponse et présentés dans ce document.
La modélisation par segmentation utilise un processus itératif pour partitionner le fichier de données et ainsi former l'arbre de décision. Le premier nœud de l'arbre de décision est tout simplement l'ensemble complet des données. La première itération définit le premier embranchement de l'arbre. Ce dernier est créer en déterminant la variable qui influence le

plus significativement la variable réponse parmi les variables retenues pour la modélisation. Une mesure statistique reflétant l'importance de la variable pour discriminer la non-réponse est calculée pour chaque variable; d'après cette mesure, la variable qui est la plus importante significativement est choisie. Le premier embranchement est ainsi formé, créant un nœud à chaque extrémité. Pour chaque nœud formé, une nouvelle itération est effectuée en répétant le même mécanisme. De nouveaux embranchements et de nouveaux nœuds sont ainsi créés. Si pour un nœud donné, aucune variable n'est significative pour expliquer la non-réponse, ou qu'un des critères d'arrêt de l'algorithme est rencontré, aucun autre embranchement n'est créé.

Plusieurs méthodes peuvent être employées pour déterminer la variable qui influence le plus la réponse (voir [22]). Nous utilisons ici l'algorithme de CHAID[12], qui sélectionne la variable avec la valeur statistique du chi-deux de Pearson la plus élevée.

6.1.2 Segmentation et régression logistique

L'alternative pour modéliser la non réponse serait d'utiliser une régression logistique. La raison principale qui nous a mené à choisir la méthode de segmentation était qu'il nous fallait rester le plus compatible possible avec les pondérations du PSM qui utilise cette méthode déjà depuis 9 ans. C'est un argument important vu qu'une pondération permettant de combiner les échantillons du Panel Suisse de Ménages et des SILC pilotes a du être produite.
Par ailleurs, mentionnons que les auteurs de [7] ont montré sur deux études[13] empiriques similaires à SILC que la méthode de segmentation était meilleure que la régression logistique pour créer des groupes homogènes de réponse (GHR) et par conséquent pour réduire le biais. Effectivement, la méthode de segmentation est plus flexible que la régression logistique. La régression logistique force le modèle à être symétrique : une fois que les variables significatives sont déterminées, les GHR sont formés en prenant toutes les combinaisons possibles de l'intersection entre ces variables. En forçant le modèle à être symétrique, de petites cellules ou des cellules qui ne sont pas significatives peuvent être formées, ce qui peut affecter la fiabilité des estimations. La modélisation par segmentation quant à elle détermine à chaque embranchement de l'arbre de décision quelle variable est la plus significative, ce qui garantit que chaque GHR formé est significatif et résulte habituellement en un modèle asymétrique. Cela présente aussi l'avantage de pouvoir utiliser un plus grand nombre de variables pour modéliser la non-réponse.

Un inconvénient de la modélisation par segmentation est directement relié à l'arbre de décision lui-même. Lorsqu'un arbre de décision est formé, il peut devenir assez lourd : dès qu'il comporte plusieurs embranchements, il devient plus difficile à interpréter. Les arbres de décisions sont aussi assez instables : l'ajout, le retrait ou le changement d'une variable ou d'une unité peut avoir un effet important sur l'arbre lui-même.

Le tableau 11 résume quelques caractéristiques comparatives entre les méthodes de modélisation de la non-réponse par segmentation et par régression logistique.

[12] CHAID: Chi-Square Automatic Interaction Detection, algorithme inventé par Kass (voir [8]).
[13] SLID (*Survey of Labour and Income Dynamics*), un panel rotatif, et NLSCY (*National Longitudinal Survey of Children and Youth*) un panel simple.

Tableau 11 Comparaison entre la régression logistique et la méthode de segmentation dans le cadre de la modélisation de la non réponse

Régression logistique	Segmentation
➢ Procédure pas à pas (stepwise) pour obtenir la liste des prédicateurs les plus significatifs de la prédisposition à répondre, ➢ Les GHR sont créés sur la base de 2^k combinaisons des k prédicateurs sélectionnés, ➢ Approche qu'on peut qualifier de symétrique, ➢ On peut ajouter des contraintes pour les GHR (typiquement n>30 et taux de réponse >50% par GHR), ➢ Nombre de variables pouvant être utilisées pour modéliser la non-réponse est limité.	➢ Utilise l'algorithme de CHAID qui divise l'échantillon en plus petits sous-groupes d'après leur taux de réponse jusqu'à ce que plus aucun prédicateur significatif ne soit trouvé. Les sous-groupes finaux forment les GHR. ➢ Approche qu'on peut qualifier de non symétrique. ➢ Les contraintes doivent être satisfaites pour chaque GHR (par exemple, n>30 et taux de réponse >50%: le choix de ces contraintes a une influence directe sur le nombre final de GHR), ➢ Peut utiliser un éventail plus large de variables pour modéliser la non-réponse.

6.1.3 Application dans le cadre des SILC pilotes

On utilise une macro SAS qui fut développée par Jean-François Naud de Statistique Canada lorsque ce dernier reçut le mandat d'effectuer les pondérations du Panel Suisse de Ménages pour la première fois en 2001. La macro permet de modéliser la non- réponse pour plusieurs types d'ajustements (se reporter aux figures 9 et 10 pour la définition des étapes mentionnées):

- Ajustement unique des poids initiaux (paramètre `Type_aj=4`), étape 2 bis,
- ajustement pour la non-réponse au questionnaire Grille (paramètre `Type_aj=1`), étape 3,
- ajustement pour la non-réponse au questionnaire individuel dans le cadre de la pondération longitudinale (paramètre `Type_aj=2`),
- ajustement pour la non-réponse au questionnaire individuel dans le cadre de la pondération transversale (paramètre `Type_aj=3`), étape 5,
- ajustement pour la non-réponse au questionnaire ménage dans le cadre de la pondération transversale (paramètre `Type_aj=4`), étape 8.

Les unités répondantes de chaque GHR voient leur poids avant ajustement multiplié par un facteur correspondant à l'inverse du taux de réponse pondéré de leur GHR.

Les exigences minimales (critères d'arrêt de l'algorithme dans sa détermination d'embranchements supplémentaires, outre le fait de ne plus trouver de variable expliquant la variable réponse de manière significative), pour chaque GHR, ont été 50% de taux de réponse minimal et 30 unités au minimum.

6.2 Poids initial du SILC_I et poids initial du SILC_II

Étape 1

Pour les deux enquêtes pilotes, le poids initial (POIDINIT) pour la pondération correspond à celui calculé à l'étape 1 (voir figure 9), c'est l'inverse de la probabilité d'inclusion du ménage dans l'échantillon. On l'appelle aussi souvent le poids de sondage (ou poids de design ou encore poids d'échantillonnage) :

$$POIDINIT_h = \frac{nombre\ de\ ménages\ dans\ strate\ h}{nombre\ de\ ménages\ sélectionnés\ dans\ strate\ h} = \frac{N_h}{S_h}$$

Pour le SILC_I vague 2 on utilise N_h^{04} et S_h^{I} et pour SILC_II vague 1 on utilise N_h^{05} et S_h^{II} (cf. tableaux 3 et 4). Notons qu'à ce stade tous les poids initiaux des ménages d'une même strate *h* (grande région) sont égaux.
Le tableau 12 donne les valeurs de ces poids initiaux, respectivement des taux de sondages T_h, par strate.

Tableau 12 Poids initiaux et taux de sondage des pilotes SILC

Strates *h*	SILC_I, v1 2004 $POIDINIT_h$	(T_h)	SILC_II, v1 2005 $POIDINIT_h$	(T_h)
1: Région lémanique	880	0.11%	1752	0.06%
2: Espace Mitteland	857	0.12%	1683	0.06%
3: Suisse du nord-ouest	794	0.13%	1620	0.06%
4: Zurich	830	0.12%	1619	0.06%
5: Suisse orientale	881	0.11%	1764	0.06%
6: Suisse centrale	907	0.11%	1744	0.06%
7: Tessin	953	0.10%	1814	0.06%

Étape 2bis

Pour le SILC_I vague 2 s'ajoute l'étape 2bis (voir figure 9) où le poids initial est ajusté pour la non-réponse totale à l'enquête non-réponse. Pour ce faire, on crée un fichier ne contenant que les ménages ayant répondu à la grille ou à l'enquête non-réponse de la vague 1 avec l'information commune aux deux questionnaires.
Le type d'ajustement est identique à celui qui est utilisé pour l'étape 8 lorsqu'il s'agit de modéliser la non réponse au questionnaire ménage (donc `Type_aj=4`). Pour pouvoir utiliser la macro il a fallu créer une variable `Coderep_grille04` permettant de discerner parmi les ménages contenant au moins un individu longitudinal, lesquels

- ont un questionnaire non réponse complété (donc non-répondants au questionnaire grille) (`Coderep_grille04=3`),
- sont répondants au questionnaire grille au moins (`Coderep_grille04=4`).

On utilise intentionnellement les codes '3' et '4' pour que la macro interprète[14] correctement notre modèle. Le tableau 13 donne la répartition des ménages selon la variable `Coderep_grille04`.

[14] Pour plus de précisions à ce propos, voir le code du programme EGR_aj_poidinit_uniqueSILCI_2006-12-13.sas, et celui de la macro aj_nr.sas.

Tableau 13 Répartition des ménages ayant complété un questionnaire non réponse versus au moins un questionnaire grille.

	SILC_I vague 1
Questionnaire NR complété (Coderep_grille04=3)	**428** (15.8%)
Répondants au questionnaire grille au moins (Coderep_grille04=4).	**2287**(*) (84.2%)
TOTAL	**2715**

(*) Correspond aux nombre de grilles réalisées en 2004 (cf. tableau 5)

Nous avons imposé un minimum de 30 unités et un taux minimal de réponse de 50% dans chacun des groupes homogènes de réponse (GHR). La modélisation est faite à partir des données pondérées par le poids initial (POIDINIT).

Le modèle de non-réponse créé avec l'analyse par segmentation a conduit à la création de 34 GHR. Le poids ajusté pour la non-réponse (au questionnaire non-réponse) est POIDINIT2.

Le tableau 14 donne quelques informations sur le facteur d'ajustement ainsi calculé et appliqué aux POIDINIT pour obtenir les POIDINIT2.

Tableau 14 Statistiques descriptives du facteur d'ajustement, f_aj_poidinit, pour les 2287 ménages répondants (voir tableau 13). POIDINIT2 = f_aj_poidinit * POIDINIT.

f_aj_poidinit			
Moyenne	1.19	Min	1.00
Médiane	1.11	Max	1.93
Écart-type	0.25	Étendue	0.93

Les détails sur cette modélisation de la non-réponse se trouvent à l'annexe B.

Remarque

Les ménages répondants à l'enquête non-réponse effectuée en 2004 n'auront pas de poids (poids final = 0) contrairement à ce qui fut fait en vague 1 par Statistique Canada (voir [4]). Dans les étapes ultérieures de la pondération, ces ménages ne sont plus considérés du tout. L'information qu'ils ont délivrée est « contenue » dans les nouveaux poids initiaux (POIDINIT2).

6.3 Classement des individus : étape 2

Cette étape a pour but de créer un fichier de base pour la pondération contenant tous les individus caractérisés par des variables de classification qui servent lors des étapes subséquentes.

Le fichier contient tous les individus des ménages ayant rempli au moins le questionnaire grille (ou un questionnaire non réponse) en 2004 ou en 2005[15]. Les variables contenues dans le fichier « classement » sont les suivantes:

Nom de la variable	Contenue dans le fichier classement SILC_I v2	Contenue dans le fichier classement SILC_II v1	Libellé
IDPERS	(x)	(x)	identificateur de la personne
idhous_o	(=IDHOUS04)	(=IDHOUS05)	identificateur du ménage à la vague originale (=vague 1)
idhous_c	(=IDHOUS05)	(=IDHOUS05)	identificateur du ménage à la vague courante (2005) (= vague 2 pour SILC_I, vague 1 pour SILC_II)
f_plong	(x)	(x)	indicateur de poids longitudinal
f_ptrin	(x)	(x)	indicateur de poids transversal individuel
f_ptrme	(x)	(x)	indicateur de poids transversal ménage
agecorro	(x)	(x)	âge de l'individu à la vague originale
agecorrc	(x)	(x)	âge de l'individu à la vague courante
poidinit	(x)	(x)	poids initial (étape 1)
poidinit2	(x)		poids initial (étape 2bis)
longst	(x)	(x)	statut de l'individu (longitudinal / cohabitant)
coderep	(x)	(x)	code de réponse à la vague courante
cr_v1	(x)		code de réponse à la vague 1 du SILC_I
grid04	(x)		indique si le ménage a rempli une grille ou un quest. NR en 2004
	6424 obs.	3015 obs.	

On peut trouver les détails sur la manière dont ces variables ont été construites dans l'annexe C.

Les tableaux 15, 16 et 17 donnent les distributions des individus selon leur statut, leur code de réponse et s'ils doivent ou non avoir un poids non-nul.

En pratique, l'étape 2bis peut être réalisée avant ou après le classement des individus. Mais dans tous les cas, le POIDINIT2 doit figurer dans le fichier classement final avant de pouvoir passer à l'étape 3.

[15] A ce stade, le SILC_I est traité séparément du SILC_II. Il y a donc un classement distinct pour ces deux enquêtes.

Tableau 15 Distribution des individus du SILC_I et du SILC_II selon leur statut. Le nombre de longitudinaux adultes et enfants ne change pas d'une vague à l'autre (pour SILC_I) car on fait toujours référence aux statuts qu'avaient les individus en l'année de sélection de l'échantillon panélisé.

Statut de l'individu – LONGST	Fréquence SILC_I vague 1	Fréquence SILC_I Vague 2	Fré-quence SILC_II Vague 1
1 – Longitudinal adulte	5346	5346	2522
2 – Longitudinal enfant	987	987	493
3 – Cohabitant[16] initialement présent	0	59	0
4 – Cohabitant initialement absent	0	32	0
Total	**6333**	**6424**	**3015**

Tableau 16 Distribution des individus du SILC_I et du SILC_II selon leur code de réponse[17]

Code de réponse – CODEREP	Fréquence SILC_I Vague 1	Fréquence SILC_I Vague 2	Fréquence SILC_I Vague 2 **avec POIDINIT2>0**	Fréquence SILC_II Vague 1
0 – < 14 ans à la vague courante	987	941	805	493
1 – Hors-cible transversalement	0	16	16	0
2 – Non-répondant total	0	3224	2440	0
3 – Répondant grille seulement	940	79	78	119
4 – Répondant grille et questionnaire ménage seulement, dans un ménage où tous les individus sont non-répondants au questionnaire individuel	161	50	50	137
5 – Répondant grille et questionnaire ménage seulement, dans un ménage où au moins un individu est répondant au questionnaire individuel	747	431	404	469
6 – Répondant grille, questionnaire ménage et questionnaire individuel	3494	1678	1652	1791
7 – Répondant grille et questionnaire individuel seulement	4	5	5	6
Total	**6333**	**6424**	**5450**	**3015**

[16] Voir paragraphe 4.1.3 pour les définitions des cohabitants.
[17] Les enfants de moins de 14 ans sont tous classés hors-cible transversalement.

Tableau 17 Distribution des individus du SILC_I et du SILC_II selon leur indicateur de poids longitudinal, transversal ménage[18] et transversal individuel

Indicateur de poids	Fréquence SILC_I Vague 1	Fréquence SILC_I Vague 2	Fréquence SILC_II Vague 1
F_PLONG = 0 – Poids longitudinal = 0	-	4782	1218
F_PLONG = 1 – Poids longitudinal > 0	-	1642	1797
Total	**-**	**6424**	**3015**
F_PTRME = 0 – Poids transversal ménage = 0	1931	4265	618
F_PTRME = 1 – Poids transversal ménage > 0	4402	2159	2397
Total	**6333**	**6424**	**3015**
F_PTRIN = 0 – Poids transversal individuel = 0	2835	4741	1218
F_PTRIN = 1 – Poids transversal individuel > 0	3498	1683	1797
Total	**6333**	**6424**	**3015**

6.4 Ajustement pour la non-réponse – Grille : étape 3

Pour le SILC_I vague 2, on a tenté de retracer tous les longitudinaux qu'on avait décidé de panéliser, qu'ils vivent toujours ou non dans le même ménage qu'à la vague 1, et qu'ils aient ou non été répondants à la vague 2. Tous les ménages dans lesquels des longitudinaux vivaient à la vague 2 étaient susceptibles de répondre à la grille.
Toutefois, comme l'unité longitudinale est l'individu et non le ménage, la probabilité de réponse à la grille a été modélisée au niveau de l'individu. On modélise donc la probabilité qu'un individu fasse partie d'un ménage ayant répondu à la grille. Les longitudinaux adultes et enfants sont utilisés. Ils sont cependant divisés selon qu'ils soient ou non des enfants dès la première étape de modélisation. La somme des poids longitudinaux avant et après l'ajustement pour la non-réponse à la grille reste la même pour les enfants et pour les adultes. La modélisation est faite à partir des données ménage et individuelles de la vague 1.

Pour le SILC_II il n'y a pas d'ajustement pour la non réponse à la grille puisqu'on est en vague 1. Dès lors, pour cette enquête, on a P_NRGRIL=POIDINIT.

La modélisation de la non réponse à la grille (pour SILC_I) se fait de la même façon et avec sensiblement les mêmes variables que la modélisation des POIDINIT2.
On crée un fichier ne contenant que les longitudinaux adultes (LONGST=1) et enfants (LONGST=2).

On définit la variable :
REPGRIL = 0 Non-répondant à la grille.
(CODEREP=0, 2).
1 Répondant à la grille.
(CODEREP=3, 4, 5, 6 ou 7).

[18] Les enfants de moins de 14 ans ont F_PTRME=0 même si leur ménage a POIDTRME > 0.

Tableau 18 Distribution des longitudinaux selon la réponse à la grille. On a 5434 individus au total car on a écarté les 16 individus hors cible transversalement (voir tableau 16)

	SILC_I Vague 2		
Réponse à la grille (REPGRIL)	Nombre de longitudinaux (Proportion %)		
	Adultes	Enfants	Total
0 – Non-répondant à la grille.	2412	833	**3245 (59.7)**
1 – Répondant à la grille.	2139	50	**2189 (40.3)**
Total	**4551 (83.8)**	**883 (16.2)**	**5434**

Notons que les personnes dans les ménages non panélisés du SILC_I vague 1 ont toutes été considérées comme non répondantes, d'où la petite proportion (40.3%) de répondants. Également pour cette raison, il a fallu relaxer les contraintes du modèle de segmentation: un taux de réponse pondéré d'au moins 30 % et un effectif minimal de 50 individus ont été exigés dans chaque GHR.
Aucun ajustement pour la non-réponse n'a été fait pour les individus hors-cible transversalement (CODEREP=1). Leur poids ajusté pour la non-réponse sera identique au poids initial.

Le modèle de non-réponse créé avec l'analyse par segmentation a conduit à la création de 25 GHR. Le tableau 19 donne quelques informations sur le facteur d'ajustement ainsi calculé et appliqué aux POIDINIT2 pour obtenir les P_NRGRIL.

Tableau 19 Statistiques descriptives du facteur d'ajustement, f_aj_poidinit2, pour les 2139 individus répondants (voir tableau 18). P_NRGRIL = f_aj_poidinit2 * POIDINIT2.

f_aj_poidinit2			
Moyenne	2.11	Min	1.36
Médiane	2.00	Max	3.32
Écart-type	0.55	Étendue	1.96

Les détails sur cette modélisation de la non-réponse se trouvent à l'annexe D.

7 Pondération transversale individuelle

Dans ce chapitre on décrit les étapes 4, 5 et 6 de la figure 9. La procédure est la même pour le SILC_I vague 2 et le SILC_II vague 1. Le poids de départ de la pondération transversale individuelle est le poids ajusté pour la non-réponse à la grille des longitudinaux âgés de 14 ans ou plus à la vague courante. L'étape du partage des poids permet d'attribuer un poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante. On procède ensuite à un ajustement pour la non-réponse au questionnaire individuel et, finalement, à un calage sur marge en utilisant des totaux de contrôle pour la population résidant dans des ménages privés de Suisse en 2005.

7.1 Partage des poids (uniquement pour SILC_I vague 2) : étape 4

Après l'étape de l'ajustement pour la non-réponse à la grille, seuls les longitudinaux des ménages répondants à la grille ont un poids P_NRGRIL non-nul. L'étape du partage des poids consiste à attribuer une partie de ces poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante.
Pour les considérations et validations théoriques quant à la méthode du partage des poids, voir [9].
Notons que les enfants longitudinaux qui ont atteint 14 ans à la vague courante sont traités comme des longitudinaux, et non pas comme des cohabitants initialement absents. Ainsi on tient compte de la possibilité qu'un enfant longitudinal de plus de 14 ans soit le seul individu longitudinal d'un ménage.

Processus

On utilise tous les individus faisant partie de la population cible transversale et répondants à la grille (variable CODEREP = 3, 4, 5, 6 ou 7). Le poids de base (avant partage) est P_NRGRIL. Il est nul pour tous les cohabitants.
Le poids après le partage est PTI_PAR. **Ce poids peut être différent pour les individus d'un même ménage.**
Le poids partagé se calcule de la manière suivante :
i : individu
L : nombre de longitudinaux dans le ménage
P : nombre de cohabitants initialement présents dans le ménage

- *si le ménage comporte des cohabitants initialement présents*, le poids partagé est le même pour tous et est égal à la somme des poids des longitudinaux divisée par le nombre de longitudinaux et de cohabitants initialement présents dans le ménage.

$$PTI_PAR_i = \frac{\sum_{i=1}^{L} P_NRGRIL_i}{L+P}$$

- *s'il n'y a aucun cohabitant initialement présent dans le ménage*, alors les poids des longitudinaux demeurent les mêmes et ceux des cohabitants initialement absents sont égaux à la moyenne des poids des longitudinaux.

$$PTI_PAR_i = \begin{cases} P_NRGRIL_i & \textit{pour i longitudinaux} \\ \dfrac{\sum_{i=1}^{L} P_NRGRIL_i}{L} & \textit{pour i cohabitants initialement absent} \end{cases}$$

On utilise la macro SAS `%partage()` développée par Jean-François Naud de Statistique Canada pour effectuer le partage des poids. Pour plus de précisions quant à l'utilisation de cette macro se reporter à son code source.

7.2 Ajustement pour la non-réponse – questionnaire individuel : étape 5

Une fois que le ménage a rempli la grille, les individus peuvent répondre au questionnaire individuel, peu importe que le questionnaire ménage ait été rempli ou non. C'est pourquoi, pour être cohérent avec le processus de collecte et parce que peu d'information du questionnaire ménage s'avérerait utile pour la modélisation, nous faisons l'ajustement pour la non-réponse au questionnaire individuel directement à partir du poids partagé qui représente les individus des ménages ayant répondu à la grille.

Processus

On crée un fichier contenant les individus (longitudinaux adultes et cohabitants) dont les ménages ont répondu à la grille et faisant partie de la population cible transversale, i.e. tous ceux pour qui PTI_PAR > 0 pour le SILC_I vague 2, et respectivement tous ceux pour qui POIDINIT > 0 pour le SILC_II vague 1.

On définit : REPQI = 0 Non-répondant au questionnaire individuel (CODEREP=3, 4 ou 5)
1 Répondant au questionnaire individuel (CODEREP=6 ou 7).

Tableau 20 Distribution des adultes faisant partie de la population cible transversale et appartenant à des ménages répondants à la grille selon la réponse au questionnaire individuel.

	SILC_I Vague 1	*SILC_I Vague 2*	*SILC_II Vague 1*
Réponse au Q.I. (REPQI)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)	Nombre de personnes (proportion en %)
0 – Non-répondant au Q.I.	1'848 (34,57)	560 (24,97)	725 (28.75)
1 – Répondant au Q.I.	3'498 (65,43)	1'683 (75,03)	1'797 (71,25)
Total	**5'346**	**2'243**	**2'522**

L'ajustement se fait au niveau des individus. Étant donné que le nombre de cohabitants est très faible (voir tableau 15), il a été nécessaire de faire la modélisation des longitudinaux et des cohabitants ensemble. Les données provenant de la grille de la **vague courante** ont été utilisées pour la modélisation.
Pour le SILC_I vague 2, le modèle obtenu avec l'analyse par segmentation comporte 16 GHR (voir annexe E) et 15 GHR pour le SILC_II vague 1 (voir annexe F). Comme pour l'ajustement pour la non-réponse à la grille, un taux de réponse pondéré d'au moins 50% et un effectif minimal de 30 individus étaient exigés dans chaque GHR.

Le poids ajusté pour la non-réponse (questionnaire individuel) est PTI_NRQI.

7.3 Calage : étape 6

Le poids transversal ajusté pour la non-réponse au questionnaire individuel (PTI_NRQI) représente la population de la Suisse à la vague courante. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2005 selon l'ESPOP (de décembre 2005). Cette étape est réalisée pour chacune des deux enquêtes. Les totaux se trouvent dans le tableau 22.

Poids avant calage : PTI_NRQI

Totaux de contrôle

1. Sexe*(14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)
4. État civil (marié; pas marié)

Poids final : POIDTRIN[19]

Le tableau 21 donne quelques nombres caractérisant les calages effectués. Le rapport entre le plus grand et le plus petit poids vaut respectivement 38.3 et 4.4 pour le SILC_I et le SILC_II. Cette différence est en grande partie due au fait que, pour le SILC_I, la modélisation de la non-réponse s'est faite sur la base de tous les répondants à la vague 1, et non pas que sur ceux qu'on avait décidé de panélisé. Ceci a eu pour conséquence de créer un arbre de décision très détaillé, et donc amené beaucoup dans variations parmi les poids.[20]

Tableau 21 Caractéristiques du calage sur marge, pondération transversale individuelle. On a respectivement 1683 et 1797 observations avec un POIDTRIN>0 pour SILC_I et SILC_II.

	facteur d'ajustement du calage = POIDTRIN/PTI_NRQI		*POIDTRIN*	
	SILC_I	*SILC_II*	*SILC_I*	*SILC_II*
Moyenne	0.85	0.89	3648	3417
Médiane	0.83	0.84	2869	3141
Écart-type	0.21	0.20	2415	1126
Min	0.57	0.63	519	1961
Max	1.64	1.50	19901	8579
Étendue	1.07	0.87	19382	6617

[19] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal individuel, donc POIDTRIN, est renommée en:

- Wp05TS1P pour le SILC_I vague 2
- Wp05TS2P pour le SILC_II vague 1

[20] En principe, il aurait été méthodologiquement plus correct de modéliser la non-réponse en se basant que sur les répondants de vague 1 qu'on a tenté de panéliser, et avant cela de gonfler les poids initiaux desdits individus par un facteur reflétant le tirage aléatoire simple qui a été effectué pour choisir les ménages qu'on allait panéliser.

Tableau 22 Effectifs de population estimés avant et après calage pour le SILC_I. Poids transversal individuel. Notons que ces totaux reflètent la population de 14 ans et plus.

	Estimation avec le poids avant calage (PTI_NRQI)		*Population (et estimation avec le poids après calage POIDTRIN)*
Variable	*SILC_I vague 2*	*SILC_II vague 1*	
Homme 14-24 ans	645’729	625’315	475’589
Homme 25-34 ans	438’275	462’884	478’590
Homme 35-44 ans	617’276	664’226	611’819
Homme 45-54 ans	640’330	554’175	523’030
Homme 55 ans et plus	964’346	968’070	908’112
Femme 14-24 ans	627’348	545’748	453’603
Femme 25-34 ans	525’629	566’895	487’849
Femme 35-44 ans	740’472	691’491	612’325
Femme 45-54 ans	696’214	630’517	519’152
Femme 55 ans et plus	1’110’076	1’198’769	1’069’501
Suisse	6’094’603	5’987’739	4’878’470
Étranger	911’091	920’350	1’261’099
Lémanique, Centrale et Tessin	2’389’052	2'183’746	1’947’586
Mitteland, N-O, Zürich, Orientale.	4'616’642	4'724’343	4’191’983
Non marié	3'334’736	3’125’942	2’743’079
Marié	3'670’958	3'782’147	3’396’490
Total	**7’005’694**	**6’908’089**	**6139’569**

8 Pondération transversale ménage

Dans ce chapitre on décrit les étapes 7, 8 et 9 de la figure 9. La procédure est la même pour le SILC_I vague 2 et le SILC_II vague 1. Le poids de départ de la pondération transversale ménage est aussi le poids ajusté pour la non-réponse à la grille des longitudinaux adultes. L'étape du partage des poids (seulement pour SILC_I) permet d'attribuer un poids aux cohabitants qui se sont joints aux ménages des longitudinaux entre la première vague et la vague courante et d'obtenir un poids au niveau du ménage. Ensuite, on procède à un ajustement pour la non-réponse au questionnaire ménage et, finalement, à un calage sur marges en utilisant des totaux de contrôle pour la population résidant dans des ménages privés de Suisse en 2005.

8.1 Partage des poids (uniquement pour SILC_I vague 2) : étape 7

Après l'étape de l'ajustement pour la non-réponse à la grille, tous les longitudinaux des ménages répondants à la grille ont un poids P_NRGRIL > 0. L'étape du partage des poids consiste à attribuer une partie de ces poids aux cohabitants qui se sont joints aux ménages des longitudinaux depuis la vague 1.

Comme pour la pondération transversale individuelle, notons que les enfants longitudinaux qui ont atteint 14 ans sont traités comme des longitudinaux, et non pas comme des cohabitants initialement absents comme on pourrait aussi le faire, ceci pour tenir compte de la possibilité qu'à un enfant longitudinal de plus de 14 ans d'être le seul individu longitudinal d'un ménage.

Processus

On utilise tous les individus faisant partie de la population cible transversale et répondants à la grille (variable CODEREP = 3, 4, 5, 6 ou 7). Le poids de base (avant le partage) est P_NRGRIL. Le poids après le partage est PTM_PAR. **Ce poids est le même pour tous les individus d'un ménage.**

Le poids ménage partagé est égal à la somme des poids des longitudinaux divisée par le nombre de longitudinaux et de cohabitants initialement présents dans le ménage.

$$PTI_PAR_i = \frac{\sum_{i=1}^{L} P_NRGRIL_i}{L+P}$$

i : individu
L : nombre de longitudinaux dans le ménage
P : nombre de cohabitants initialement présents dans le ménage

Le fichier résultant est au niveau du ménage (identificateur IDHOUS05).
Comme pour la pondération transversale individus (voir paragraphe 7.1), on utilise la macro `%partage()` pour effectuer le partage des poids.

8.2 Ajustement pour la non-réponse – questionnaire ménage : étape 8

Après l'étape du partage des poids, tous les ménages ayant répondu à la grille de la vague courante ont un poids transversal. Comme seuls les ménages qui ont répondu au questionnaire ménage auront un poids transversal ménage final, il faut procéder à un autre ajustement pour la non-réponse, cette fois au niveau du ménage. Seules des données de la grille (**vague courante**) sont utilisées pour la modélisation. La modélisation est faite pour chacune des deux enquêtes (séparément).

Processus

On crée un fichier contenant les ménages qui ont répondu à la grille et qui font partie de la population cible transversale, i.e. tous ceux pour qui PTM_PAR > 0 pour SILC_I vague 2, POIDINIT > 0 pour SILC_II vague 1, et qui ont le bon code de réponse (CODEREP=3, 4, 5, 6 ou 7).

On définit: REPQM= 0 Non-répondant au questionnaire ménage
(CODEREP = 3 ou 7 pour tous les individus)
1 Répondant au questionnaire ménage
(CODEREP = 4, 5 ou 6 pour tous les individus).

Tableau 23 Distribution des ménages faisant partie de la population cible transversale et répondants à la grille selon la réponse au questionnaire ménage.

	SILC_I Vague 1	*SILC_I Vague 2*	*SILC_II Vague 1*
Réponse au QM (REPQM)	Nombre de ménages (proportion en %)	Nombre de ménages (proportion en %)	Nombre de ménages (proportion en %)
0 – Non-répondant au questionnaire ménage	516 (19.01)	39 (3,54)	59 (4,67)
1 – Répondant au questionnaire ménage	2'199 (80,99)	1'064 (96,46)	1'204 (95,33)
Total	**2'715**	**1'103**	**1'263**

La modélisation est basée sur la grille de la vague courante soit celle de la vague 2 pour le SILC_I et celle de la vague 1 pour le SILC_II. Le modèle obtenu avec l'analyse par segmentation comporte 10 GHR pour le SILC_I et 8 pour le SILC_II. Un taux de réponse pondéré d'au moins 50% et un effectif minimal de 30 individus ont été exigés dans chaque GHR.

Le poids ajusté pour la non-réponse (questionnaire ménage) est PTM_NRQM. Le tableau 24 donne quelques caractéristiques numériques de cet ajustement.

Tableau 24 Statistiques descriptives du facteur d'ajustement, f_aj_tm_nrqm, pour les respectivement 1064 et 1204 ménages répondants (voir tableau 23) des SILC_I et SILC_II.
PTM_NRQM = f_aj_tm_nrqm * PTM_PAR pour SILC_I,
PTM_NRQM = f_aj_tm_nrqm * POIDINIT pour SILC_II.

	f_aj_tm_nrqm	
	SILC_I vague 2	SILC_II vague 1
Moyenne	1.04	1.05
Médiane	1.00	1.01
Écart-type	0.12	0.08
Min	1.00	1.00
Max	1.60	1.51
Étendue	0.60	0.51

Pour plus de détails sur la modélisation, se rapporter à l'Annexe F.

8.3 Calage : étape 9

Comme pour la pondération transversale individuelle, on veut que le nombre de personnes dans les ménages représente le total de la population résidente en Suisse. Il s'agit ici de faire un calage sur marges *intégré*[21] pour le poids ménage en utilisant les comptes de population Suisse en 2005 (vague courante), d'après l'ESPOP de décembre 2005.

Le poids transversal ajusté pour la non-réponse au questionnaire ménage (PTM_NRQM) représente la population résidante en Suisse à la vague courante. L'étape du calage ajustera ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population en 2005. Cette étape est réalisée pour chacune des deux enquêtes. Les totaux se trouvent dans le tableau 26.

Poids avant calage : PTM_NRQM

Totaux de contrôle

1. Sexe*(< 14 ans; 14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (région lémanique & Suisse centrale & Tessin; espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse orientale)
4. État civil (marié; pas marié)

[21] Dans un calage intégré, on veut satisfaire les totaux de chaque marge au niveau individuel, tout en forçant les poids finaux (après calage) des individus d'un même ménage à rester égaux entre eux.

Le fait de n'avoir panélisé qu'un sous-échantillon du SILC_I vague 1 (2004), et d'avoir modéliser la non-réponse sur la base de tous les répondants en vague 1, s'est fait ressentir lors de cette étape du calage. En effet, après un premier calage, deux ménages se sont vus attribuer un poids après calage négatif, ce qui n'est pas acceptable. Comme les méthodologues de Statistique Canada le proposent (voir [10]), une manière de régler ce problème est d'effectuer un deuxième calage après avoir préalablement réattribué le poids avant premier calage (soit PTM_NRQM) aux observations qui posent problème.
Pour SILC_II un seul calage a suffit.
Les tableaux 25 et 26 donnent quelques nombres caractérisant les calages effectués pour SILC_I et SILC_II.

Poids final : POIDTRME[22]

Tableau 25 Caractéristiques du calage sur marge, pondération transversale ménage. On a respectivement 2594 individus dans 1064 ménages, et 2872 individus dans 1204 ménages avec un POIDTRME>0 pour SILC_I et SILC_II. Les caractéristiques des facteurs d'ajustement sont au niveau des individus, alors que celles des poids après calage sont au niveau des ménages.

	facteur d'ajustement du calage = poids avant / poids après				*Poids finaux*	
	SILC_I 1er calage	*SILC_I 2e calage*	*SILC_II*	*SILC_I Poids après 1er calage*	*SILC_I POIDTRME 2e calage*	*SILC_II POIDTRME*
	Niveau individus dans les ménages répondants au questionnaire ménage			*Niveau ménages répondants au questionnaire ménage*		
Moyenne	0.84	0.98	0.88	3019	3017	2592
Médiane	0.83	0.99	0.83	2653	2640	2451
Écart-type	0.34	0.03	0.24	1692	1701	655
Min	**-0.22**	0.82	0.38	**-816**	131	1032
Max	2.13	1.03	1.92	13450	13554	5958
Étendue	2.35	0.21	1.55	14266	13423	4926

[22] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal ménage, donc POIDTRME, a été renommée en:
- Wh05TS1P pour le SILC_I vague 2
- Wh05TS2P pour le SILC_II vague 1

Tableau 26 Effectifs de population estimés avant et après calage pour SILC_I et SILC_II. Poids transversal ménage. Notons que ces totaux reflètent toute la population (tous âges, y compris les moins de 14 ans).

	Estimation avec le poids avant calage (PTI_NRQI)		*Population (et estimation avec le poids après calage)*
Variable	*SILC_I vague 2 (avant 1er calage)*	*SILC_II vague 1*	
Homme < 14 ans	677'072	684'357	577'996
Homme 14-24 ans	688'955	582'102	495'263
Homme 25-34 ans	449'488	428'350	478'590
Homme 35-44 ans	622'427	688'954	611'819
Homme 45-54 ans	630'172	573'278	530'616
Homme 55 ans et plus	962'631	1'041'343	908'112
Femme < 14 ans	600'316	683'209	545'814
Femme 14-24 ans	632'936	502'903	467'359
Femme 25-34 ans	455'953	538'364	487'849
Femme 35-44 ans	735'705	691'162	619'911
Femme 45-54 ans	694'495	604'833	519'152
Femme 55 ans et plus	1'110'037	1'227'959	1'069'501
Suisse	7'237'757	7'100'078	5'782'529
Étranger	1'022'429	1'146'736	1'529'453
Lémanique, Centrale et Tessin	2'767'773	2'625'674	2'350'333
Mitteland, N-O, Zürich, Orientale.	5'492'413	5'621'141	4'961'649
Non marié	4'673'914	4'406'884	3'900'319
Marié	3'586'273	3'839'930	3'411'663
Total	**8'260'186**	**8'246'815**	**7'311'982**

9 Pondération transversale individuelle combinée SILC_I et SILC_II

9.1 Combinaison transversale de panels à partir des échantillons transversaux de chacun des deux panels

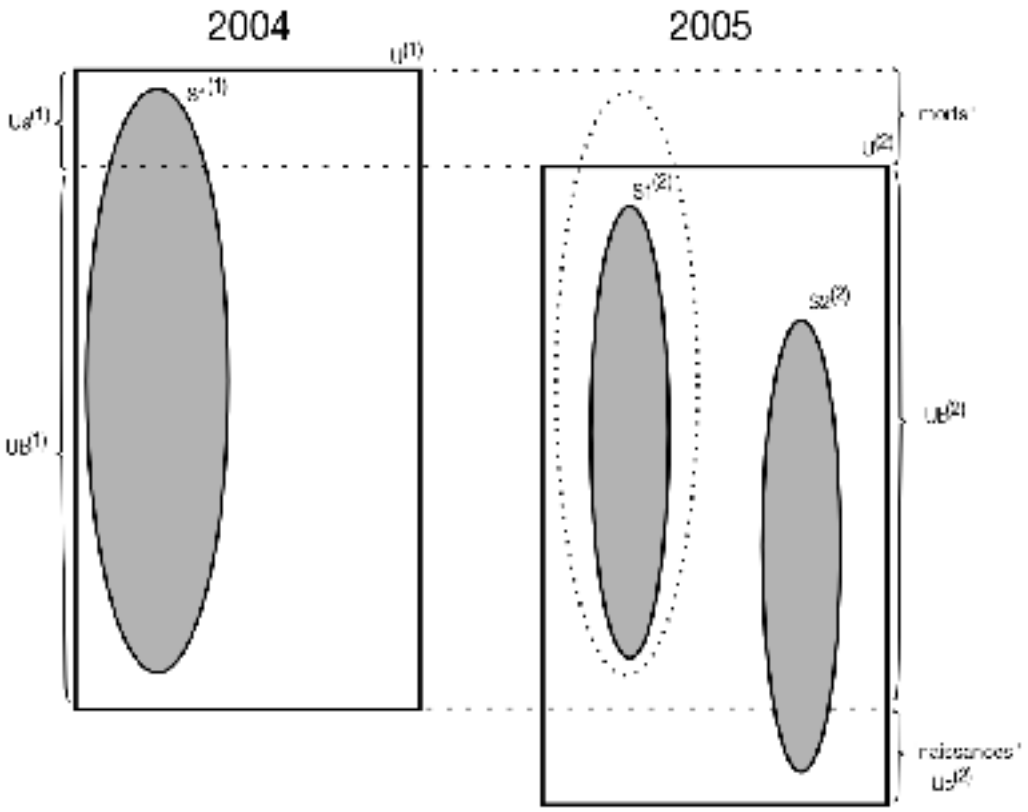
La pondération transversale combinée est faite en combinant les enquêtes (considérées comme des débuts de panels) transversalement.
Fixons les notations, soient :

t année de la vague courante : 2005,
t_{S1} année de sélection du SILC_I : 2004,
t_{S2} année de sélection du SILC_II : 2005.

En l'occurrence nous nous trouvons dans le cas particulier où t= t_{S2}.

Pour combiner transversalement le SILC_I au SILC_II, on crée d'abord des poids afin que chacun des deux échantillons soit représentatif de la population transversale au temps t. En fait, on récupère les étapes de l'ajustement pour la non-réponse à la grille, du partage des poids pour les cohabitants et de l'ajustement pour la non-réponse au questionnaire individuel déjà faites séparément pour chacune des deux enquêtes. On part pour chacun des deux SILC du poids PTI_NRQI (étape 5, voir figure 10). On obtient donc deux « panels » représentatifs de la population. Ensuite, on combine les deux échantillons transversaux.

Figure 27 Combinaison des panels.



Légende

$U^{(1)} = U_a^{(1)} \cup U_b^{(1)}$	Population cible pour le SILC_I (panel 1)
$U^{(2)} = U_b^{(2)} \cup U_c^{(2)}$	Population cible du SILC_II (panel 2)
$S1^{(1)}$	Échantillon des répondants du SILC_I ,vague 1
$S1^{(2)}$	Échantillon des répondants du SILC_I, vague 2
$S2^{(2)}$	Échantillon des répondants du SILC_II, vague 1

Notons que les individus de $U_a^{(1)}$ (voir figure 27) ayant quitté la population cible entre la sélection du SILC_I et le temps t ne font pas partie de cet échantillon combiné $S_1^{(2)} \cup S_2^{(2)}$.

Les différences entre les populations $U^{(1)}$ et $U^{(2)}$ sont dues aux personnes qui sont sorties de la population cible entre septembre 2004 et septembre 2005 (personnes décédées, institutionnalisées, parties à l'étranger, etc.) et par les nouveaux arrivés dans la population après septembre 2004 (immigrants, naissances). Une partie de l'échantillon du premier panel comprend des personnes de la population sortante $U_a^{(1)}$; les individus de cette population ne font alors plus partie de la population suisse au moment de la sélection du deuxième panel (SILC_II) et ont ainsi une probabilité nulle d'être choisis. De même, une partie du second panel est composée d'individus qui sont dans la population suisse au moment de la sélection de cet échantillon mais qui n'étaient pas présents lors de la sélection du premier panel ($U_c^{(2)}$); ces personnes n'avaient aucune chance d'être choisies au moment t_{S1}.

Afin de développer une méthodologie qui permet de combiner l'information des échantillons transversaux des deux panels, il suffit de considérer les deux panel au moment de l'introduction du deuxième panel. La population que l'on doit considérer est $U^{(2)}$, celle effective lors de l'introduction du deuxième panel. Les échantillons longitudinaux à considérer sont les échantillons originaux choisis à la vague une des panels. L'échantillon du premier panel ayant été réduit à cause des individus qui ont quitté la population et des non-répondants. La méthode présentée ici a été développée par Merkouris (voir [14] et [16]). Soit *Y* un paramètre de la population. Pour estimer *Y* il s'agit d'utiliser l'estimateur combiné suivant (qui donne une importance relative à chacun des panels selon un certain facteur *p*) :

$$\hat{Y} = p \cdot \hat{Y}_1 + (1-p) \cdot \hat{Y}_2$$

où $\hat{Y}$ est l'estimation transversale pour une variable d'intérêt, $\hat{Y}_i$ est l'estimation produite pour cette même variable d'intérêt pour le panel *i*, et *p* est le facteur d'allocation des panels. **La valeur optimale du facteur d'allocation des panels a été déterminée de façon à ce que la variance de $\hat{Y}$ soit minimale.** Le facteur d'allocation des panels est défini comme suit :

$$p = \frac{n_1}{n_1 + n_2(\frac{d_1}{d_2})}$$

où d_i désigne l'*effet de plan d'échantillonnage*[23] pour le panel *i* et n_i représente le nombre de personnes[24] (ou de ménages au chapitre suivant) du panel *i* répondantes au temps *t* (t étant l'année de référence traitée, 2005 en l'occurrence).

[23] L'effet de plan d'échantillonnage est défini comme étant le rapport entre la variance de l'estimation sous-pondérée (i.e. sans calage sur marge pour respecter les totaux de contrôle de la population) et la variance qui aurait résulté d'un échantillon aléatoire simple de même taille.

[24] Les méthodologues de Statistique Canada suggèrent de combiner les deux panels avant d'intégrer les cohabitants par la méthode du partage des poids, i.e. que les n_i représentent le nombre de répondants longitudinaux seulement. Omettre les cohabitants dans le calcul des facteurs *p* contribue à réduire l'importance du panel le plus âgé et ainsi minimiser les impacts potentiels dus au biais d'attrition.
Mais comme dans notre cas les deux panels ont été introduits à une seule année d'intervalle, l'effet des cohabitants et leur nombre sont minimes. De ce fait, nous n'avons pas distingué les cohabitants des longitudinaux pour le calcul des n_i qui représentent, dans notre cas, la somme des deux.

Comme les plans de sondage de SILC_I et SILC_II sont identiques, on suppose que les effets de plan sont sensiblement les mêmes ($d_1/d_2=1$). Pour calculer les facteurs d'allocation, on utilise le nombre de personnes (ou de ménages) répondants.

On applique donc aux poids de tous les individus un facteur d'allocation $p \leq 1$ de la manière suivante :

p : individus du SILC_I
1-p : individus du SILC_II qui habitent un ménage dont au moins un membre faisait partie de la population cible au moment t_{S1}.
1 : individus du SILC_II qui habitent dans un ménage dont aucun des membres ne faisaient partie de la population cible au moment t_{S1}.

Les échantillons sont combinés au niveau des grandes régions. Un facteur d'allocation est donc calculé pour chaque grande région. On procède ensuite au calage selon les totaux de la population résidant dans des ménages privés de Suisse en 2005 pour obtenir le poids transversal individuel combiné.

9.2 Combinaison des deux panels: étape 11

On combine les deux panels afin de former un nouveau panel qui inclut les deux échantillons. La combinaison s'effectue en ajustant les poids de chaque unité selon le facteur d'allocation des panels cité plus haut.

Processus

On utilise les variables PRES_P1 (=1 si l'individu faisait partie de la population cible à la sélection du SILC_I, 0 sinon) et PRES_P2 (=1 si l'individu faisait partie de la population cible à la sélection du SILC_II, 0 sinon) pour déterminer le poids combiné :

- On applique *p* aux individus du SILC_I qui ont PRES_P2=1.
- On applique (*1-p*) aux individus du SILC_II qui ont PRES_P1=1.
- On applique 1 aux individus du SILC_II qui habitent dans un ménage dont aucun des membres ne faisaient partie de la population cible au moment de la sélection du SILC_I.

Le tableau 28 présente les facteurs d'allocation calculés pour chaque grande région.

Remarque

Dans les faits, on a eu PRES_P1=PRES_P2=1 pour tous les individus! Le facteur d'allocation des panels s'obtient donc simplement comme suit : pour chaque grande région h, $p_h = n_{h1}/ n_h$ (voir tableau 28 pour les notations).
Le poids après combinaison s'obtient comme suit:

$$\text{PTI_COMBPAN_REG} = \begin{cases} \text{PTI_NRQI} * p_h & \text{pour les individus du SILC_I} \\ \text{PTI_NRQI} * (1- p_h) & \text{pour les individus du SILC_II} \end{cases}$$

Tableau 28 Facteurs d'allocation par grande région. Pondération transversale individuelle combinée du SILC_I et le SILC_II.

REGION05 h	*Nombre total de personnes* n_h	*Nombres de personnes SILC_ I:* n_{h1}	*Nombres de personnes SILC_II:* n_{h2}	*Facteur d'allocation des panels:* p_h
1	571	261	310	0.46
2	831	390	441	0.47
3	492	234	258	0.48
4	645	310	335	0.48
5	479	254	225	0.53
6	317	161	156	0.51
7	145	73	72	0.50
TOTAL	3480	1683	1797	

9.3 Calage: étape 12

Le poids combiné (PTI_COMBPAN_REG) représente la population de la Suisse à la vague courante. L'étape du calage ajuste ces poids pour que les totaux selon certaines variables de contrôle soient identiques à ceux de la population suisse en 2005. Les totaux se trouvent dans le tableau 30.
Poids avant calage : PTI_COMBPAN_REG.

Les totaux de contrôle sont les mêmes que pour les pondérations transversales individuelles prises pour chaque enquête séparément :

1. Sexe*(14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (Région lémanique & Suisse Centrale & Tessin; Espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse Orientale)
4. État civil (marié; pas marié)

Poids final : POIDTRIN[25].

Le tableau 29 donne quelques nombres caractérisant le calage effectué.

Tableau 29 Caractéristiques du calage sur marge, pondération transversale individuelle. On a respectivement 3480 individus avec un POIDTRIN>0 pour les SILC_I et SILC_II combinés.

	facteur d'ajustement du calage = POIDTRIN/ PTI_COMBPAN_REG	*POIDTRIN*
Moyenne	0.87	1764
Médiane	0.82	1553
Écart-type	0.20	901
Min	0.62	256
Max	1.52	9747
Étendue	0.90	9491

[25] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRIN, est renommée en wp05T4P.

Tableau 30 Effectifs de population estimés avant et après calage pour le SILC_I combiné au SILC_II. Poids transversal individuel combiné.

Variable	*Estimation avec le poids avant calage (P_COMBPAN_IND)*	*Population (et estimation avec le poids après calage POIDTRIN)*
Homme 14-24 ans	638'042	475'589
Homme 25-34 ans	451'186	478'590
Homme 35-44 ans	640'909	611'819
Homme 45-54 ans	597'346	523'030
Homme 55 ans et plus	965'549	908'112
Femme 14-24 ans	588'213	453'603
Femme 25-34 ans	549'068	487'849
Femme 35-44 ans	715'065	612'325
Femme 45-54 ans	662'166	519'152
Femme 55 ans et plus	1'157'457	1'069'501
Suisse	6'045'961	4'878'470
Étranger	919'039	1'261'099
Lémanique, centrale et Tessin	2'292'055	1'947'586
Mitteland, N-O, Zürich, orientale.	4'672'945	4'191'983
Pas Marié	3'230'940	2'743'079
Marié	3'734'060	3'396'490
Total	6'965'000	**6139'569**

10 Pondération transversale combinée ménage SILC_I et SILC_II

La pondération transversale combinée ménage se fait de la même manière que la pondération transversale combinée individuelle sauf qu'on combine deux échantillons transversaux composés de ménages et non d'individus. On combine donc les panels après l'ajustement pour la non-réponse au questionnaire ménage. Le poids avant combinaison est donc PTM_NRQM (qui a été calculé lors des pondérations transversales ménage de chacune des deux enquêtes, étape 8, voir figure 10). En dernier lieu, on procède à un calage intégré en utilisant les totaux de la population résidant dans des ménages privés de Suisse en 2005.

10.1 Combinaison des deux panels: étape 13

On combine les deux panels afin de former un nouveau panel basé sur les deux échantillons. La combinaison s'effectue en ajustant les poids de chaque unité selon le facteur d'allocation des panels.

Calcul du poids combiné:

- On applique *p* aux individus du SILC_I qui ont PRES_P2=1.
- On applique (*1-p*) aux individus du SILC_II qui ont PRES_P1=1.
- On applique 1 aux individus du SILC_II qui habitent dans un ménage dont aucun des membres ne faisaient partie de la population cible au moment de la sélection du SILC_I.

Le tableau 31 présente les facteurs d'allocation calculés pour chaque grande région de Suisse.

Tableau 31 Facteurs d'allocation par région. Pondération transversale ménage du SILC_I combiné au SILC_II.

REGION05 h	*Nombre total de personnes* n_h	*Nombres de personnes SILC_I:* n_{h1}	*Nombres de personnes SILC_II:* n_{h2}	*Facteur d'allocation des panels:* p_h
1	376	175	201	0.47
2	537	247	290	0.46
3	314	151	163	0.48
4	436	198	238	0.45
5	293	142	151	0.48
6	214	105	109	0.49
7	98	46	52	0.47
TOTAL	2268	1064	1204	

Comme pour la pondération transversale individuelle combinée, on utilise les variables PRES_P1 et PRES_P2 pour déterminer le poids combiné.
On a : PRES_P1=PRES_P2=1 pour tous les ménages. Le facteur d'allocation des panels s'obtient donc simplement comme suit : pour chaque grande région h, $p_h = n_{h1}/ n_h$ (voir tableau 31 pour les notations).

Le poids après combinaison s'obtient comme suit:

$$\text{PTM_COMBPAN_REG} = \begin{cases} \text{PTM_NRQI} * p_h & \text{pour les ménages du SILC_I} \\ \text{PTM_NRQI} * (1 - p_h) & \text{pour les ménages du SILC_II} \end{cases}$$

10.2 Calage: étape 14

Comme pour la pondération transversale ménage séparée par panel, on veut que le nombre de personnes dans les ménages représente le total de la population résidant dans des ménages privés de Suisse. On procède donc à nouveau à un calage sur marges intégré en utilisant les comptes de l'ESPOP (de décembre 2005).

Le poids transversal combiné (PTM_COMBPAN_REG) représente la population de la Suisse à la vague courante. Les totaux se trouvent dans le tableau 33.

Poids avant calage : P_ PTM_COMBPAN_REG.

Totaux de contrôle

1. Sexe*(moins de 14; 14-24 ans; 25-34 ans; 35-44 ans; 45-54 ans; 55+ ans)
2. Nationalité (suisse; étranger)
3. Région (région lémanique & Suisse centrale & Tessin; espace Mitteland & Suisse du Nord-Ouest & Zürich & Suisse orientale)
4. État civil (marié; pas marié)

Poids final : POIDTRME[26]

Comme pour la pondération transversale ménage du SILC_I (voir paragraphe 8.3), le calage du poids transversal ménage est effectué deux fois car après le premier calage, il y a un ménage dont le poids est négatif (d'une valeur de -12,9). Avant de refaire le second calage, on remet les poids négatifs à leur valeur avant le premier calage (soit P_ PTM_COMBPAN_REG). POIDTRME est le poids transversal ménage final de la pondération combinée (donc après le second calage).
Le tableau 32 résume les faits.

[26] Dans les descriptifs des pondérations livrés aux utilisateurs, la variable poids final transversal combiné, i.e. POIDTRME, est renommée en wh05T4P.

Tableau 32 Caractéristiques du calage sur marge, pondération transversale ménage combinée. On a 5466 individus dans 2268 ménages avec un POIDTRME>0 pour SILC_I combiné au SILC_II. Les caractéristiques des facteurs d'ajustement sont au niveau des individus, alors que celle des poids après calage sont au niveau des ménages.

	facteur d'ajustement du calage = poids avant / poids après			
	SILC_I & SILC_II 1er calage	*SILC_I & SILC_II 2e calage*	*SILC_I & SILC_II Après premier calage*	*Poids finaux SILC_I & SILC_II POIDTRME*
	Niveau individus dont le ménage a répondu au questionnaire ménage		*Niveau ménages répondants au questionnaire ménage*	
Moyenne	0.87	1.00	1325	1325
Médiane	0.85	1.00	1209	1208
Écart-type	0.27	0.01	620	618
Min	**-0.01**	0.95	**-13**	130
Max	2.14	1.01	5309	5291
Étendue	2.15	0.06	5322	5161

Tableau 33 Effectifs de population estimés avant et après calage pour le SILC_I combiné au SILC_II. Poids transversal ménage combiné.

Variable	*Estimation avec le poids avant calage (PTM_COMBPAN_REG)*	*Population (et estimation avec le poids après calage POIDTRME)*
Homme moins de 14 ans	684'345	577'996
Homme 14-24 ans	633'418	495'263
Homme 25-34 ans	436'597	478'590
Homme 35-44 ans	661'125	611'819
Homme 45-54 ans	603'483	530'616
Homme 55 ans et plus	997'769	908'112
Femme moins de 14 ans	646'475	545'814
Femme 14-24 ans	564'124	467'359
Femme 25-34 ans	504'717	487'849
Femme 35-44 ans	709'734	619'911
Femme 45-54 ans	649'383	519'152
Femme 55 ans et plus	1'168'582	1'069'501
Suisse	7'170'091	5'782'529
Étranger	1'089'660	1'529'453
Lémanique, centrale et Tessin	2'696'139	2'350'333
Mitteland, N-O, Zürich, orientale.	5'563'613	4'961'649
Pas Marié	4'534'763	3'900'319
Marié	3'724'989	3'411'663
Total	**8'259'752**	**7'311'982**

11 Conclusions

Les pondérations décrites dans ce document ont été calquées sur celles produites par des méthodologues de Statistique Canada (voir [11]) pour les données du Panel Suisse de Ménages (PSM) récoltées de septembre 2005 à février 2006. Ceci était un pré-requis puisque les données SILC pilotes dont il est question dans ce rapport ont été récoltées durant la même période, par le même institut de sondage, et qu'elles sont diffusées ensemble avec celles du PSM. La raison pour diffuser les données aux utilisateurs communément avec celles du PSM est que plus de la moitié des questions posées par CATI étaient communes aux deux enquêtes. De ce fait il fut souhaité que les réponses aux questions, récoltées simultanément dans les deux enquêtes, puissent être réunies et exploitées grâce à une pondération commune aux deux projets (PSM et SILC). Il a donc fallu procéder d'une manière similaire à ce qui a été fait pour le PSM afin de garantir la plus grande compatibilité possible des jeux de poids à combiner produits pour chaque enquête séparément.
La pondération combinant les enquêtes PSM et SILC n'est pas décrite ici. Elle a été faite similairement à la combinaison du SILC_I vague 2 au SILC_II vague 1 décrite dans ce document. Sa description fera l'objet d'une note séparée.
On a détaillé ici les pondérations des SILC pilotes. Pour l'enquête SILC principale, dès 2007, il s'agira de pondérer un panel rotatif sur 4 ans et donc de combiner jusqu'à quatre échantillons-panels. La manière de procéder sera fondamentalement la même que ce qui a été fait ici, bien que quelques adaptations soient nécessaires afin de tenir compte des spécificités du panel rotatif.
La combinaison des panels pour les pondérations transversales s'est faite après le partage des poids, cette option fut choisie pour sa simplicité, comme discuté au chapitre 5, mais ce n'est pas la seule façon de faire.
Le fait de n'avoir panélisé qu'une partie de l'échantillon original du SILC_I créa des difficultés dans l'élaboration des pondérations (petite taille d'échantillon), surtout au niveau du calage intégré (chapitres 8 et 10).
Concernant le calage sur marge, l'OFS ayant l'habitude d'utiliser la macro CALMAR développée par l'INSEE, on pourrait souhaiter l'appliquer aussi dans le cadre de SILC pour l'enquête principale. Ceci impliquera encore d'étudier les possibilités de CALMAR afin de s'assurer que la macro permette d'exécuter un calage intégré comme on le fait ici pour réaliser les pondérations transversales ménage. Le choix de la méthode de calage (fonction de distance) doit aussi être rediscuté.

Pour clôturer ces conclusions, et toujours avec un regard vers le futur de SILC, ajoutons encore que la modélisation de la non réponse, faite ici par la méthode segmentation (chapitre 6), pourrait dans à l'avenir aussi se faire par régression logistique. L'expérience avec le PSM a montré que la méthode de segmentation, meilleure en principe (cf. [7]) et plus flexible, présente l'inconvénient, pour une enquête panel, d'être assez variable d'une vague à l'autre. Une recommandation pour les pondérations de l'enquête SILC principale, serait de fixer une modèle par régression logistique pour les pondérations longitudinales de chacun des quatre panels en tournus, et de modéliser par la méthode de segmentation dans le cadre des pondérations transversales. Cette façon de faire aurait l'avantage de garantir le maximum de stabilité dans le temps (année après année) pour les pondérations longitudinales, et en même temps, le maximum de flexibilité pour les pondérations transversales. On pourrait aussi décider d'effectuer toutes les modélisations de non réponse par segmentation, mais de fixer, pour chaque échantillon panélisé sur quatre ans, les premières branches de l'arbre de décision. Ce procédé serait probablement idéal, mais il implique de modifier les macros effectuant la segmentation afin de pouvoir ajouter ce genres de contraintes au modèle.

12 Bibliographie

[1] EUROSTAT. *Cross-sectional weighting: from second year of the survey onwards*. EU-SILC Doc. 157/05.

[2] EUROSTAT. *Cross sectional weighting: first year of each sub-sample*. Working Group on EU-SILC, Luxembourg 29-30 March 2004. EU-SILC 134/04.

[3] Graf, E. (2006). *SILC, Sample sizes requirements for Switzerland according to the rules defined by EUROSTAT for the European Union and rotational design*. Document interne, OFS.

[4] Cauchon, C. et Latouche, M. (2006). *Pondération du SILC, SILC vague 1, PSM II et SILC combinés, PSM I, PSM II et SILC combinés; description détaillée des tâches*. Panel Suisse de Ménage.

[5] Graf, E. et Tillmann, R. (2005). *Comparaison du déroulement des enquêtes PSMI, PSMII et SILC (2004-2005).* Working Paper 4_05, Panel Suisse de Ménage.

[6] Graf, E. et Tillmann, R. (2006). *Comparaison du déroulement des enquêtes PSM et SILC (2005-2006) : l'exemple de la gestion des interviews.* Bulletin de la société suisse de statistiques, 53 (avril 2006), 9-13.

[7] J. Dufour, F. Gagnon, Y. Morin, M. Renaud, C.-E. Särndal (1998). *Measuring the impact of alternative weighting schemes for longitudinal data.* SAA proceedings, Survey Research Methods Section, pp. 552-557.

[8] Kass, G.V. (1980). *An exploratory technique for investigating large quantities of categorical data*. Applied Statistics, 29, 119-127.

[9] Lavallée, P. (2002). Le sondage indirect ou la méthode généralisée du partage des poids.

[10] Latouche, M. (2007). *Calage des poids et estimations de la variance pour le Panel Suisse de ménages.* Panel Suisse de Ménages, document interne.

[11] Cauchon, C. et Latouche, M. (2007). *Pondération du Panel Suisse de Ménages, PSM_I vague 7, PSM II vague 2, PSM_I et PSM_II combinés; description détaillée des tâches*. Panel Suisse de Ménage.

[12] Latouche, M. et Naud, J.-F. (2005). Panel suisse de ménages, stratégie de pondération en présence de plusieurs panels. Document interne du PSM.

[13] Naud, J.-F. (2005). Deux façons de combiner des panels. Document interne du PSM.

[14] Merkouris, T. (2001). *Estimation transversale dans le cas des enquêtes auprès des ménages à panels multiples.* Techniques d'enquête, vol. 27, pp. 189-200.

[15] Ardilly, P. et Lavallée, P. (2003). *La méthode du partage des poids et l'enquête européenne sur le revenu et les conditions de vie.* Recueil du Symposium de Statistique Canada sur les Défis à la réalisation d'enquêtes pour la prochaine décennie.

[16] Lévesque, I. et Franklin, S. (2000). *Pondérations longitudinale et transversale de l'Enquête sur la Dynamique du travail et du revenu, Année de référence 19*97. Statistique Canada, 75F0002MIF-00004.

[17] Naud, J.-F. (2004). *Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu, 1996-2002*. Statistique Canada, 75F0002MIF2004008.

[18] Latouche, M., Dufour, J. et Merkouris, T. (2000). *Pondération transversale : combinaison de deux panels ou plus*. Statistique Canada, 75F0002MIF-00006.

[19] Deville, J.-C. et Särndal, C. E. (1992). *Calibration Estimators in Survey Sampling.* Journal of the American Statistical Association, Vol87, No. 418 (June 1992), PP. 376-382.

[20] Tambay, J. L., Schiopu-Kratina, I., Mayda, J., Stukel, D. et Nadon, S. (1998). *Traitement de la non-réponse du cycle de deux enquêtes sur la santé de la population.* Techniques d'enquête, vol. 24, no 2, pp.159-169.

[21] Kalton, G. et Kasprzyk, D. (1986). *Le traitement des données d'enquête manquantes.* Techniques d'enquête, 12, pp. 1-17.

[22] Nakache, J.-P. et Confais, J. (2003). *Statistique explicative appliquée : analyse discriminante, modèle logistique, segmentation par arbre.* Editions Technip.

ANNEXE A – Fichiers de données pour SILC_I et SILC_II

Voici une description sommaire des fichiers de données à partir desquels les pondérations du SILC_I et du SILC_II ont été effectuées. Pour plus de détails, se rapporter aux documents de travail internes produits par l'auteur.

MÉNAGES 2004 : ce fichier contient les ménages qui ont répondu à la grille ou au questionnaire non réponse à la vague 1 du SILC_I,
Le fichier contient 2715 enregistrements.

INDIVIDUS 2004 : ce fichier contient tous les individus des ménages qui ont répondu à la grille ou au questionnaire non réponse à la vague 1 du SILC_I. Excepté des individus des ménage ayant seulement rempli le questionnaire non réponse, tous font partie de la population cible transversale. Ils peuvent être répondants ou non-répondants au questionnaire individuel.
6333 enregistrements.

MÉNAGES 2005 : ce fichier contient les ménages qui ont répondu à la grille à la vague 2 du SILC_I ou à la vague 1 du SILC_II.
Le fichier contient 2366 enregistrements.

INDIVIDUS 2005 : ce fichier contient tous les individus des ménages qui ont répondu à la grille à la vague 2 du SILC_I ou à la vague 1 du SILC_II. Ils font tous partie de la population cible transversale. Ils peuvent être répondants ou non-répondants au questionnaire individuel.
5704 enregistrements.

LONGOUT05 : (uniquement pour SILC_I) ce fichier contient tous les individus longitudinaux qui ont quitté leur ménage entre la vague 1 et la vague 2, avec la raison du départ. Ce fichier est utilisé pour identifier les longitudinaux qui ont quitté la population cible (ceux dont la variable RNH vaut 1, 2, 3 ou 7).
71 enregistrements.

PARTIS05 : (uniquement pour SILC_I) ce fichier contient tous les individus qui, à la vague courante, ne vivent plus dans les ménages dans lequel ils étaient recensés lors de la vague précédente.
91 enregistrements.

ANNEXE B – Modélisation de la non-réponse au questionnaire non-réponse pour les poids initiaux du SILC_I

Modélisation de la non-réponse – variables utilisées

Les variables servant à la modélisation sont des variables dichotomiques qui caractérisent les individus et les ménages et qui proviennent des données (questionnaires grilles et questionnaires non réponse) de la vague 1 (2004).

Voici une liste des variables qui pouvaient potentiellement faire partie du modèle de non-réponse (en **gras** celles qui ont été retenues pour le modèle (16 variables différentes)):

FRANC	Questionnaire grille ou non réponse rempli en français
ITAL	Questionnaire grille ou non réponse rempli en italien
ALLEM	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	Région Lémanique
MIDDLE	Région Mitteland
NOREST	Région Nord-Ouest
EST	Région Suisse orientale
CENTRAL	Région Suisse centrale
ZURICH	Région Zurich.
MIXED	Type de commune : agricole mixte
CENTRES	Type de commune : agglomération
SUBURB	Type de commune : banlieue
WEALTH	Type de commune : riche
PERIPH	Type de commune : périphérique
TOURIST	Type de commune : touristique
INDUS	Type de commune : industrielle ou tertiaire
RURAL	Type de commune : cité dortoir
AD_SEUL	Ménage avec un seul adulte
NO_KID	Ménage sans enfant
SEULE	Ménage d'une seule personne
PROPRIO	Propriétaire
TOUS_HOMMES	Tous les membres du ménage sont des hommes
TOUS_FEMMES	Tous les membres du ménage sont des femmes
UN_MARIE	Au moins une personne du ménage est mariée
MAX_EDUC_SUP	Le niveau d'éducation maximal du ménage est une formation supérieure
MAX_EDUC_PRO	Le niveau d'éducation maximal du ménage est une formation professionnelle
MAX_NO_EDUC	Le niveau d'éducation maximal du ménage n'est pas une formation professionnelle ni une formation supérieure
TOUS_SUIS	Tous les membres du ménage sont de nationalité suisse
TOUS_ETRANG	Tous les membres du ménage sont de nationalité étrangère (non suisse)
AGEMAX_1	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 15 et 25, 0 sinon.
AGEMAX_2	Vaut 1 si l'âge maximal du ménage est compris entre 25 et 34, 0 sinon.
AGEMAX_3	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 35 et 44, 0 sinon.
AGEMAX_4	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 45 et 54, 0 sinon.
AGEMAX_5	Vaut 1 si l'âge de la personne la plus âgée du ménage est plus grand que 55 ans, 0 sinon.

Modélisation de la non-réponse – analyse par segmentation

Le modèle de non-réponse créé avec l'analyse par segmentation a conduit à la création de 34 groupes homogènes de réponse (GHR).

Tableau B – 1 Arbre de segmentation, modélisation des poids initiaux, POIDINIT2, SILC_I. Les nombres entre parenthèses sont les effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés

split1	split2	split3	split4	split5	split6	split7	split8	rhg
max_educ_sup = 0 (1728) 81.3%	max_no_educ = 0 (1278) 77.8%	tous_suis = 0 (166) 57.7%	AD_SEUL = 0 (130) 51.8%					1
			AD_SEUL = 1 (36) 79.1%					2
		tous_suis = 1 (1112) 80.8%	agemax_5 = 0 (660) 86.8%	PROPRIO = 0 (291) 81.6%	un_marie = 0 (99) 59.8%			3
					un_marie = 1 (192) 92.7%	agemax_4 = 0 (91) 87.6%		4
						agemax_4 = 1 (101) 97.1%		5
				PROPRIO = 1 (369) 91.0%	agemax_4 = 0 (264) 94.7%	NO_KID = 0 (107) 88.6%	tous_femmes = 0 (51) 78.0%	6
							tous_femmes = 1 (56) 98.1%	7
						NO_KID = 1 (157) 98.8%		8
					agemax_4 = 1 (105) 81.9%	CENTRES = 0 (74) 75.8%	tous_hommes = 0 (37) 86.9%	9
							tous_hommes = 1 (37) 65.0%	10
						CENTRES = 1 (31) 96.5%		11
			agemax_5 = 1 (452) 72.0%	ALLEM = 0 (109) 64.0%				12
				ALLEM = 1 (343) 74.7%	PROPRIO = 0 (168) 82.4%	CENTRES = 0 (125) 88.7%		13
						CENTRES = 1 (43) 63.5%		14
					PROPRIO = 1 (175) 67.2%	tous_femmes = 0 (93) 54.7%		15
						tous_femmes = 1 (82) 81.6%	CENTRES = 0 (50) 71.6%	16
							CENTRES = 1 (32) 97.1%	17
	max_no_educ = 1 (450) 91.1%	PROPRIO = 0 (196) 86.1%	SEULE = 0 (122) 93.9%					18
			SEULE = 1 (74) 73.3%					19
		PROPRIO = 1 (254) 95.0%	ALLEM = 0 (96) 90.3%	un_marie = 0 (51) 96.2%				20
				un_marie = 1 (45) 83.7%				21
			ALLEM = 1 (158) 98.1%					22
max_educ_sup = 1 (987) 89.4%	agemax_5 = 0 (705) 91.4%	PROPRIO = 0 (253) 86.8%	CENTRES = 0 (198) 92.6%	agemax_4 = 0 (110) 89.3%				23
				agemax_4 = 1 (88) 96.8%				24
			CENTRES = 1 (55) 65.3%					25
		PROPRIO = 1 (452) 93.9%	agemax_4 = 0 (353) 96.6%	agemax_3 = 0 (177) 100.0%				26
				agemax_3 = 1 (176) 93.2%	un_marie = 0 (100) 97.0%			27
					un_marie = 1 (76) 88.1%			28
			agemax_4 = 1 (99) 84.6%	AD_SEUL = 0 (67) 78.8%				29
				AD_SEUL = 1 (32) 96.7%				30
	agemax_5 = 1	PROPRIO = 0 (162) 93.0%	MIDDLE = 0 (120) 90.6%					31
			MIDDLE = 1 (42) 100.0%					32
		PROPRIO = 1 (120) 73.2%	CENTRES = 0 (68) 61.8%					33
			CENTRES = 1 (52) 88.1%					34

Remarque

Les variables dichotomiques ALLEM, FRANC et LAKE_GE sont hautement corrélées entre elles (voir tableau ci-dessous), on ne devrait en principe pas les entrer ensemble pour lancer la segmentation. Seulement ALLEM a été retenue dans notre modèle de non réponse. Les autres variables (non listées dans le tableau B-2) utilisées pour modéliser la non-réponse ont toutes un coefficient de corrélation de Pearson variant entre 0.36 et 0.002 en absolu, il est donc légitime de toutes les proposer à l'algorithme de segmentation.

Tableau B – 2 Variables dichotomiques créées pour modéliser la non réponse au questionnaire grille ayant des coefficients de corrélations de Pearson > 0.36.

	ALLEM	**FRANC**	**LAKE_GE**
ALLEM	1	-0.87	0.78
FRANC		1	-0.68
LAKE_GE			1

ANNEXE C - Dérivation des variables pour le classement des individus

Les variables produites lors de l'étape 2 (classement des individus) ont été dérivées de la manière suivante :

LONGST =
1. Longitudinal – adulte.
   (si dans INDIVIDUS de la vague 1, AGECORRO >= 14).
2. Longitudinal – enfant.
   (si dans INDIVIDUS de la vague 1, AGECORRO < 14).
3. Cohabitant initialement présent.
   (Si pas dans INDIVIDUS de la vague 1, (HAB_CH05 $\geq$ 2004 et AGECORRO>=0)).
   Pour les HAB_CH05[27] < 0 (ne sait pas, pas de réponse), on supposera qu'ils sont initialement présents.
4. Cohabitant initialement absent.
   (Si pas dans INDIVIDUS de la vague 1, (HAB_CH05 > 2004 ou AGECORRO<0)).

Note : les enfants (< 14 ans) longitudinaux n'ont pas de poids longitudinal. Les enfants qui ont 14 ans ou plus à la vague 2 du SILC_I ont été traités comme des longitudinaux lors de la pondération transversale et ils ont un poids transversal, même si l'enfant est le seul membre longitudinal du ménage.

CODEREP =
0. AGECORRC< 14
1. Hors-cible transversalement
   (LONGST=1 ou 2 et dans LONGOUT et RNH=1, 2, 3 ou 7)
2. Non-répondant total
   (AGECORRC>= 14 et pas dans INDIVIDUS 2005 et pas dans LONGOUT et avec RNH=1, 2, 3 ou 7)
3. Répondant grille seulement
   (AGECORRC >= 14 et dans INDIVIDUS 2005 et STATUS05=1 ou 2 et STATHH05 = 0)
4. Répondant grille et questionnaire ménage seulement, dans un ménage ou tous les individus sont non-répondants au questionnaire individuel
   (AGECORRC >= 14 et dans INDIVIDUS 2005 et STATUS05=1 ou 2 et STATHH05 = 1 et Q_INDIV=0)
5. Répondant grille et questionnaire ménage seulement, dans un ménage ou au moins un individu est répondant au questionnaire individuel
   (AGECORRC >= 14 et dans INDIVIDUS 2005 et STATUS05=1 ou 2 et STATHH05 = 1 et Q_INDIV=1)
6. Répondant grille, questionnaire ménage et questionnaire individuel
   (AGECORRC >= 14 et dans INDIVIDUS 2005 et STATUS05=0 et STATHH05=1)
7. Répondant grille et questionnaire individuel seulement
   (AGECORRC >= 14 et dans INDIVIDUS 2005 et STATUS05=0 et STATHH05=0)

Notons que tous les enfants (i.e. < 14 ans à la vague courante) sont hors-cible transversalement.

[27] HAB_CH05 := durée de résidence en Suisse, depuis quelle année.

Indicateurs de poids

Ces variables indiquent, pour chaque individu, s'il a un poids positif ou nul.

Poids longitudinal (individuel) :

F_PLONG =	1	Poids longitudinal positif (LONGST=1, CODEREP=1, 6 ou 7)
	0	Poids longitudinal nul (autres)

Poids transversal ménage :

F_PTRME =	1	Poids transversal ménage positif (CODEREP=4, 5 ou 6)
	0	Poids transversal ménage nul (autres)

Poids transversal individuel :

F_PTRIN =	1	Poids transversal individu positif (CODEREP=6 ou 7)
	0	Poids transversal individu nul (autres)

*Construction des variables d'âge pour **SILC_I vague 2***

IF age04 >= 0 THEN agecorro = age04;
ELSE IF age05 >= 0 THEN agecorro = age05-1;
ELSE IF annee1 >= 1800 THEN agecorro = 2004-annee1;
ELSE IF annee2 >= 1800 THEN agecorro = 2005-annee2;
→ Les bébés nés après la première vague auront un AGECORR0 négatif

Imputation des âges manquants
Vague précédente :
IDPERS=40437102 then agecorro=46;
IDPERS=41147102 then agecorro=63;
IDPERS=41280103 then agecorro=27;
IDPERS=41799101 then agecorro=21;
IDPERS=41799102 then agecorro=21;
Vague courante:
else if agecorro=. then agecorro=9999;

agecorrc= agecorro + 1

*Construction des variables d'âge pour **SILC_II vague 1***

Imputation des âges manquants
IDPERS=50127102 then agecorro=55;
IDPERS=50084102 then agecorro=45;

agecorrc= agecorro

Note sur l'imputation des âges manquants : on impute un âge seulement aux individus adultes ayant répondu au questionnaire ménage ou au questionnaire individuel. Aux autres on attribue un âge de 9999 ans afin qu'ils soient traités comme des adultes tout en restant facilement identifiables.

ANNEXE D – Modélisation de la non-réponse au questionnaire grille pour les poids initiaux ajustés du SILC_I (vague 2)

Modélisation de la non-réponse – variables utilisées

Les variables servant à la modélisation sont des variables dichotomiques qui caractérisent les individus et les ménages et qui proviennent des données (questionnaires grille) de la vague 1 (2004).
Voici une liste des variables qui pouvaient potentiellement faire partie du modèle de non-réponse (en **gras** celles qui ont été retenues pour le modèle (20 variables différentes)):

Niveau ménage

FRANC	Questionnaire grille ou non réponse rempli en français
ITAL	Questionnaire grille ou non réponse rempli en italien
ALLEM	Questionnaire grille ou non réponse rempli en allemand
LAKE_GE	Région Lémanique
MIDDLE	Région Mitteland
NOREST	Région Nord-Ouest
EST	Région Suisse orientale
CENTRAL	Région Suisse centrale
ZURICH	Région Zurich
MIXED	Type de commune : agricole mixte
CENTRES	Type de commune : agglomération
SUBURB	Type de commune : banlieue
WEALTH	Type de commune : riche
PERIPH	Type de commune : périphérique
TOURIST	Type de commune : touristique
INDUS	Type de commune : industrielle ou tertiaire
RURAL	Type de commune : cité dortoir
AD_SEUL	Ménage avec un seul adulte
NO_KID	Ménage sans enfant
SEULE	Ménage d'une seule personne
LOY_1	Coût du logement jusqu'à 1000 Fr./mois
LOY_2	Coût du logement entre 1001-1500 Fr./mois
LOY_3	Coût du logement 1501 Fr./mois et plus
RAT1	nbr. de pièces / nbr. de personnes <= 1.34
RAT2	1.34 < nbr. de pièces / nbr. de personnes < 2
APPART	Ménage vivant en appartement
DETACH	Ménage vivant dans une maison indépendante
PROPRIO	Propriétaire

Niveau individu

AGEV1_1	15-24 ans (à la vague 1)
AGEV1_2	25 à 34 ans (à la vague 1)
AGEV1_3	35 à 44 ans (à la vague 1)
AGEV1_4	45 à 54 ans (à la vague 1)
AGEV1_5	55+ ans (à la vague 1)
CELIB	Célibataire
DIVORCE	Divorcé(e)
SEPARE	Séparé(e)
MARIE	Marié(e)
FORM_PRO	Formation professionnelle
FORM_SUP	Formation supérieure
NO_FORM	Pas de formation
HOMME	Homme
SWISS	Nationalité suisse.
TRAVAIL	Personne active professionnellement

Modélisation de la non-réponse – analyse par segmentation
Le modèle de non-réponse créé avec l'analyse par segmentation a conduit à la création de 25 groupes homogènes de réponse (GHR).

split1	split2	split3	split4	split5	split6	split7	split8	split9	split10	split11	rhg
NO_KID = 0 (1773) 52.7%	SWISS = 0 (299) 40.0%	FORM_PRO = 0 (218) 35.3%									1
		FORM_PRO = 1 (81) 50.0%									2
	SWISS = 1 (1474) 55.7%	PERIPH = 0 (1284) 57.5%	LAKE_GE = 0 (1078) 59.1%	SUBURB = 0 (725) 56.4%	RAT2 = 0 (552) 52.9%	loy_2 = 0 (413) 49.8%	MIXED = 0 (337) 52.7%	EST = 0 (228) 47.1%	RURAL = 0 (155) 41.0%	PROPRIO = 0 (57) 30%	3
										PROPRIO = 1 (98) 48%	4
									RURAL = 1 (73) 59.7%		5
								EST = 1 (109) 63.3%			6
							MIXED = 1 (76) 35.7%				7
						loy_2 = 1 (139) 62.3%	APPART = 0 (66) 73.6%				8
							APPART = 1 (73) 51.5%				9
					RAT2 = 1 (173) 68.1%						10
				SUBURB = 1 (353) 64.8%							11
			LAKE_GE = 1 (206) 50.0%								12
		PERIPH = 1 (190) 43.2%									13
NO_KID = 1 (2778) 43.3%	APPART = 0 (1098) 38.5%	loy_1 = 0 (819) 35.1%	RAT2 = 0 (587) 31.1%								14
			RAT2 = 1 (232) 45.2%	agevo_5 = 0 (150) 51.7%	loy_2 = 0 (100) 58.4%						15
					loy_2 = 1 (50) 38.9%						16
				agevo_5 = 1 (82) 34.4%							17
		loy_1 = 1 (279) 48.5%									18
	APPART = 1 (1680) 46.5%	FORM_SUP = 0 (1067) 43.9%	loy_2 = 0 (714) 47.5%	agevo_1 = 0 (644) 45.4%	NOREST = 0 (551) 43.8%						19
					NOREST = 1 (93) 55.7%						20
				agevo_1 = 1 (70) 67.3%							21
			loy_2 = 1 (353) 36.3%	SEULE = 0 (259) 32.2%							22
				SEULE = 1 (94) 48.6%							23
		FORM_SUP = 1 (613) 51.5%	RAT1 = 0 (500) 53.6%								24
			RAT1 = 1 (113) 42.8%								25

Figure D-1 Arbre de segmentation. Modélisation de la non-réponse au questionnaire grille, SILC_I vague 2. Les nombres entre parenthèses sont les effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés. Seuls les longitudinaux adultes sont modélisés ici (longst=1), les longitudinaux enfants (longst=2) n'ont pas subi d'ajustement car en trop petit nombre.

ANNEXE E – Modélisation de la non-réponse au questionnaire individuel, pondération transversale individuelle du SILC_I (vague 2) et du SILC_II (vague 1)

Les variables dichotomiques utilisées pour la modélisation de la non-réponse proviennent de la grille de la vague courante.
X = variable proposée pour la modélisation de la non réponse
R = variable proposée et retenue par l'algorithme de segmentation pour la modélisation

Niveau ménage

SILC_I vague 2	SILC_II vague 1	Nom de la variable	Description
R	X	CENTRAL	Région Suisse centrale
X	X	EST	Région Suisse orientale
R	X	LAKE_GE	Région Lémanique
X	X	MIDDLE	Région Mitteland
X	**R**	ZURICH	Région Zurich
R	X	AD_SEUL	Ménage avec un seul adulte
X	**R**	SEULE	Ménage d'une seule personne
R	**R**	NO_KID	Ménage sans enfant
R	X	NOREST	Région Nord-Ouest
X	X	ALLEM	Questionnaire ménage rempli en allemand
X	X	FRANC	Questionnaire ménage rempli en français
X		ITALIEN	Questionnaire ménage rempli en italien

Niveau individu

SILC_I vague 2	SILC_II vague 1	Nom de la variable	Description
R	X	AGEVC_1	15-24 ans à la vague courante
R	X	AGEVC_2	25 à 34 ans à la vague courante
X	X	AGEVC_3	35 à 44 ans à la vague courante
X	**R**	AGEVC_4	45 à 54 ans à la vague courante
R	X	AGEVC_5	55 ans et plus à la vague courante
X	**R**	NO_FORM	Pas de formation
R	**R**	FORM_PRO	Formation professionnelle
R	**R**	FORM_SUP	Formation supérieure
R	**R**	SWISS	Nationalité suisse
X	**R**	HOMME	Homme
X	X	MARIE	Marié(e)
X	X	SEPARE	Séparé(e)
X	**R**	CELIB	Célibataire
X	X	DIVORCE	Divorcé(e)
X		TRAVAIL	Personne active professionnellement

Synthèse
10 variables différentes ont été choisies tant pour SILC_I que SILC_II dont trois se retrouvent dans les deux modèles. Le modèle obtenu avec l'analyse par segmentation comporte 16 GHR pour le SILC_I et 15 GHR pour le SILC_II. Un taux de réponse pondéré d'au moins 50% et un effectif minimal de 30 individus ont été exigés dans chaque GHR.

Figure E-1 Modèle complet de non-réponse au questionnaire individuel : pondération transversale individuelle du SILC_I. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

split1	split2	split3	split4	split5	split6	split7	rhg
AD_SEUL = 0 (1842) 70.2%	FORM_SUP = 0 (1249) 65.8%	SWISS = 0 (138) 51.8%					1
		SWISS = 1 (1111) 68.1%	CENTRAL = 0 (971) 69.7%	FORM_PRO = 0 (359) 63.5%	agevc_5 = 0 (281) 66.9%	NOREST = 0 (234) 69.5%	2
						NOREST = 1 (47) 52.1%	3
					agevc_5 = 1 (78) 55.2%		4
				FORM_PRO = 1 (612) 73.1%	agevc_1 = 0 (547) 74.4%		5
					agevc_1 = 1 (65) 60.4%		6
			CENTRAL = 1 (140) 57.8%				7
	FORM_SUP = 1 (593) 80.6%	agevc_1 = 0 (510) 82.5%	SWISS = 0 (49) 71.9%				8
			SWISS = 1 (461) 83.9%	CENTRAL = 0 (414) 85.4%	agevc_2 = 0 (325) 83.5%	NO_KID = 0 (159) 88.4%	9
						NO_KID = 1 (166) 79.8%	10
					agevc_2 = 1 (89) 93.0%		11
				CENTRAL = 1 (47) 70.7%			12
		agevc_1 = 1 (83) 68.8%					13
AD_SEUL = 1 (401) 94.9%	LAKE_GE = 0 (319) 96.7%	FORM_PRO = 0 (187) 98.8%					14
		FORM_PRO = 1 (132) 94.1%					15
	LAKE_GE = 1 (82) 88.3%						16

Figure E-2 Modèle complet de non-réponse au questionnaire individuel : pondération transversale individuelle du SILC_II. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

split1	split2	split3	split4	split5	split6	split7	rhg
NO_FORM=0 (1937) 75.24%	SEULE=0 (1639) 72.43%	CELIB=0 (1272) 75.82%	HOMME=0 (631) 79.79%	SWISS=0 (71) 68.91%			1
				SWISS=1 (560) 81.18%	NO_KID=0 (258) 85.66%		2
					NO_KID=1 (302) 77.33%	FORM_PRO=0 (90) 85.50%	3
						FORM_PRO=1 (212) 73.86%	4
			HOMME=1 (641) 71.90%	FORM_SUP=0 (345) 66.87%			5
				FORM_SUP=1 (296) 77.77%	agevc_4=0 (213) 80.78%		6
					agevc_4=1 (83) 70.10%	NO_KID=0 (51) 79.16%	7
						NO_KID=1 (32) 55.30%	8
		CELIB=1 (367) 60.68%	FORM_PRO=0 (167) 70.88%	ZURICH=0 (131) 67.25%			9
				ZURICH=1 (36) 86.11%			10
			FORM_PRO=1 (200) 52.15%				11
	SEULE=1 (298) 90.77%	FORM_SUP=0 (169) 87.29%					12
		FORM_SUP=1 (129) 95.36%					13
NO_FORM=1 (585) 58.09%	SEULE=0 (505) 54.65%						14
	SEULE=1 (80) 79.73%						15

ANNEXE F – Modélisation de la non-réponse au questionnaire ménage, pondération transversale ménage du SILC_I et SILC_II

Modélisation de la non-réponse – Variables utilisées

Les variables servant à la modélisation sont des variables dichotomiques qui caractérisent les ménages et les individus y vivant et proviennent des données (questionnaires grille) de la vague 1 (2004) pour le SILC_I, respectivement de la vague 1 (2005) pour le SILC_II. Voici une liste des variables qui pouvaient potentiellement faire partie du modèle de non-réponse :
X = variable proposée pour la modélisation de la non réponse
R = variable proposée et retenue par l'algorithme de segmentation pour la modélisation

Niveau Ménage

SILC_I vague 2	SILC_II vague 1	Nom de la variable	Description
X	X	CENTRAL	Région Suisse centrale
X	**R**	EST	Région Suisse orientale
X	X	LAKE_GE	Région Lémanique
X	X	MIDDLE	Région Mitteland
R	X	ZURICH	Région Zurich
X	**R**	AD_SEUL	Ménage avec un seul adulte
X	X	SEULE	Ménage d'une seule personne
X	X	NO_KID	Ménage sans enfant
X	X	NOREST	Région Nord-Ouest
X	X	ALLEM	Questionnaire ménage rempli en allemand
X	X	FRANC	Questionnaire ménage rempli en français
X	**R**	ITALIEN	Questionnaire ménage rempli en italien
X	X	MIXED	Type de commune : agricole mixte
X	X	PERIPH	Type de commune : périphérique
X	X	RURAL	Type de commune : cité dortoir
X	**R**	SUBURB	Type de commune : banlieue
X	X	TOURIST	Type de commune : touristique
X	X	WEALTH	Type de commune : riche
X	X	CENTRES	Type de commune : agglomération
X	X	INDUS	Type de commune : industrielle ou tertiaire
R		APPART	Ménage vivant en appartement
X		DETACH	Ménage vivant dans une maison indépendante
R		LOY_1	Coût du logement jusqu'à 1000 francs / mois
R		LOY_2	Coût du logement entre 1001 et 1500 francs / mois
R		LOY_3	Coût du logement 1501 francs / mois et plus
R	**R**	PROPRIO	Propriétaire
X		RAT1	nombre de pièces / nombre de personnes <= 1.34
X		RAT2	nombre de pièces / nombre de personnes entre 1.35 et 2
X	X	TOUS_HOMMES	Tous les membres du ménage sont des hommes
X	X	TOUS_FEMMES	Tous les membres du ménage sont des femmes
X	X	UN_MARIE	Au moins une personne du ménage est mariée
X	X	MAX_EDUC_SUP	Un des membres du ménage possède une formation supérieure
X	X	MAX_EDUC_PRO	Le niveau d'éducation de la personne la plus éduquée du ménage est une formation

			professionnelle
X		MAX_NO_EDUC	La personne la plus éduquée du ménage est sans formation
X	**R**	TOUS_SUIS	Tous les membres du ménage sont Suisses
X	X	TOUS_ETRANG	Tous les membres du ménage ne sont pas Suisses
X	X	AGEMAX_1	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 15 et 25, 0 sinon
X	X	AGEMAX_2	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 25 et 34, 0 sinon
X	X	AGEMAX_3	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 35 et 44, 0 sinon
X	X	AGEMAX_4	Vaut 1 si l'âge de la personne la plus âgée du ménage est compris entre 45 et 54, 0 sinon
R	**R**	AGEMAX_5	Vaut 1 si l'âge de la personne la plus âgée du ménage plus grand que 55 ans, 0 sinon

Modélisation de la non-réponse – analyse par segmentation

Le modèle obtenu avec l'analyse par segmentation comporte 10 GHR pour le SILC_I et 8 GHR pour le SILC_II. Comme pour l'ajustement pour la non-réponse à la grille, un taux de réponse pondéré d'au moins 30 % et un effectif minimal de 50 individus étaient exigés dans chaque GHR.

Figure F-1 Modèle complet de non-réponse au questionnaire ménage : pondération transversale ménage du SILC_I. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés

<table>
<tr><th>split1</th><th>split2</th><th>split3</th><th>split4</th><th>split5</th><th>GHR</th></tr>
<tr><td rowspan="8">PROPRIO=0
(490)
92.0%</td><td rowspan="4">APPART=0
(323)
95.9%</td><td rowspan="3">loy_3=0
(196)
93.7%</td><td rowspan="2">loy_1=0
(131)
90.8%</td><td>loy_2=0
(84) 85.8%</td><td>1</td></tr>
<tr><td>loy_2=1
(47) 100%</td><td>2</td></tr>
<tr><td>loy_1=1
(65) 100%</td><td></td><td>3</td></tr>
<tr><td>loy_3=1
(127) 100%</td><td></td><td></td><td>4</td></tr>
<tr><td rowspan="4">APPART=1
(167)
83.6%</td><td rowspan="3">loy_3=0
(121)
78.2%</td><td>agemax_5=0
(59) 62.4%</td><td></td><td>5</td></tr>
<tr><td rowspan="2">agemax_5=1
(62)
92.7%</td><td>loy_1=0
(32) 86.1%</td><td>6</td></tr>
<tr><td>loy_1=1
(30) 100%</td><td>7</td></tr>
<tr><td>loy_3=1
(46) 100%</td><td></td><td></td><td>8</td></tr>
<tr><td rowspan="2">PROPRIO=1
(613)
99.8%</td><td>ZURICH=0
(479) 100%</td><td></td><td></td><td></td><td>9</td></tr>
<tr><td>ZURICH=1
(134) 99.1%</td><td></td><td></td><td></td><td>10</td></tr>
</table>

Figure F-2 Modèle complet de non-réponse au questionnaire ménage : pondération transversale ménage du SILC_II. Les nombres entre parenthèses sont des effectifs non pondérés. Les pourcentages représentent les taux de réponse pondérés.

<table>
<tr><th>split1</th><th>split2</th><th>split3</th><th>split4</th><th>GHR</th></tr>
<tr><td rowspan="5">PROPRIO=0
(602) 90.79%</td><td rowspan="3">AD_SEUL=0
(475)
92.96%</td><td>tous_suis=0
(45) 82.16%</td><td></td><td>1</td></tr>
<tr><td rowspan="2">tous_suis=1
(430) 94.13%</td><td>EST=0
(371)
92.99%</td><td>2</td></tr>
<tr><td>EST=1
(59) 100%</td><td>3</td></tr>
<tr><td rowspan="2">AD_SEUL=1
(127)
82.69%</td><td>agemax_5=0
(32) 66.21%</td><td></td><td>4</td></tr>
<tr><td>agemax_5=1
(95) 88.20%</td><td></td><td>5</td></tr>
<tr><td rowspan="3">PROPRIO=1
(661) 99.56 %</td><td rowspan="2">ITAL=0
(631)
99.68%</td><td>SUBURB=0
(421) 100%</td><td></td><td>6</td></tr>
<tr><td>SUBURB=1
(210) 99.04%</td><td></td><td>7</td></tr>
<tr><td>ITAL=1
(30) 97.07%</td><td></td><td></td><td>8</td></tr>
</table>

ANNEXE G – Caractéristiques des distributions des différentes pondérations produites

Pondération transversale individuelle finale SILC_II vague 1 (2005)
POIDTRIN>0

Variable	POIDTRIN	Quantile	Estimation
N	1797	100% Max	8578.88
Somme des poids	6139569	99%	7265.62
Moyenne	3417	95%	5977.48
Ecart-type	1126	90%	4993.73
Variance	1267068	75% Q3	3772.18
Etendue	6617	50% Médiane	3141.37
Intervalle interquartile	1122	25% Q1	2650.32
Rapport Max/Min	4.37	10%	2353.00
		5%	2242.25
		1%	2042.29
		0% Min	1961.44

Poids corrigés pour la non réponse au quest. individuel SILC_II vague 1 (2005)
PTI_NRQI>0

Variable	PTI_NRQI	Quantile	Estimation
N	1797	100% Max	6212.98
Somme des poids	6908089	99%	5928.54
Moyenne	3844	95%	5510.48
Ecart-type	857	90%	4946.87
Variance	733619	75% Q3	4426.12
Etendue	3659	50% Médiane	3663.74
Intervalle interquartile	1270	25% Q1	3156.30
Rapport Max/Min	2.43	10%	2843.26
		5%	2790.13
		1%	2553.98
		0% Min	2553.98

Poids initiaux (= poids d'échantillonnage) SILC_II vague 1 (2005)
POIDINIT>0

Variable	POIDINIT	Quantile	Estimation
N	1263	100% Max	3240.23
Somme des poids	3454200	99%	3240.23
Moyenne	2735	95%	3240.23
Ecart-type	263	90%	3240.23
Variance	68967	75% Q3	2959.71
Etendue	805	50% Médiane	2683.70
Intervalle interquartile	408	25% Q1	2551.79
Rapport Max/Min	1.33	10%	2435.55
		5%	2435.55
		1%	2435.55
		0% Min	2435.55

Pondération transversale ménage finale SILC_II vague 1 (2005)
POIDTRME>0

Variable	POIDTRME	Quantile	Estimation
N	1204	100% Max	5958.42
Somme des poids	3120971	99%	4871.07
Moyenne	2592	95%	3944.75
Ecart-type	655	90%	3411.50
Variance	429256	75% Q3	2839.92
Etendue	4926	50% Médiane	2451.47
Intervalle interquartile	660	25% Q1	2179.94
Rapport Max/Min	5.77	10%	1961.60
		5%	1825.04
		1%	1491.03
		0% Min	1032.39

Poids corrigés pour la non réponse au quest. ménage SILC_II vague 1 (2005)
PTM_NRQM>0

Variable	PTM_NRQM	Quantile	Estimation
N	1204	100% Max	4893.58
Somme des poids	3454200	99%	4053.08
Moyenne	2869	95%	3355.53
Ecart-type	352	90%	3240.23
Variance	123622	75% Q3	3102.73
Etendue	2458	50% Médiane	2744.13
Intervalle interquartile	484	25% Q1	2619.13
Rapport Max/Min	2.00	10%	2459.09
		5%	2435.55
		1%	2435.55
		0% Min	2435.55

Pondération transversale individuelle finale SILC_I vague 2 (2005)
POIDTRIN>0

Variable	POIDTRIN	Quantile	Estimation
N	1683	100% Max	19900.87
Somme des poids	6139569	99%	14371.72
Moyenne	3648	95%	8289.69
Ecart-type	2415	90%	6074.21
Variance	5831112	75% Q3	4292.30
Etendue	19382	50% Médiane	2869.36
Intervalle interquartile	2084	25% Q1	2207.88
Rapport Max/Min	38.38	10%	1842.02
		5%	1600.30
		1%	1267.74
		0% Min	519.14

Poids corrigés pour la non réponse au quest. individuel SILC_I vague 2 (2005)
PTI_NRQI>0

Variable	PTI_NRQI	Quantile	Estimation
N	1683	100% Max	14957.56
Somme des poids	7005694	99%	10206.20
Moyenne	4163	95%	8074.99
Ecart-type	1923	90%	6645.64
Variance	3699808	75% Q3	5041.78
Etendue	14177	50% Médiane	3603.13
Intervalle interquartile	2223	25% Q1	2818.70
Rapport Max/Min	19.15	10%	2360.21
		5%	2176.58
		1%	1551.16
		0% Min	780.99

Pondération transversale ménage finale SILC_I vague 2 (2005)
POIDTRME>0

Variable	POIDTRME	Quantile	Estimation
N	1064	100% Max	13553.90
Somme des poids	3209819	99%	9037.26
Moyenne	3017	95%	6615.50
Ecart-type	1701	90%	5136.47
Variance	2894783	75% Q3	3715.10
Etendue	13423	50% Médiane	2639.95
Intervalle interquartile	1830	25% Q1	1884.70
Rapport Max/Min	103.18	10%	1310.61
		5%	1087.86
		1%	640.26
		0% Min	131.36

Poids corrigés pour la non réponse au quest. ménage SILC_I vague 2 (2005)
PTM_NRQM>0

Variable	PTM_NRQM	Quantile	Estimation
N	1064	100% Max	9487.56
Somme des poids	3503342	99%	7641.51
Moyenne	3293	95%	5775.29
Ecart-type	1275	90%	4956.65
Variance	1624904	75% Q3	3934.42
Etendue	8945	50% Médiane	2921.51
Intervalle interquartile	1494	25% Q1	2440.10
Rapport Max/Min	17.49	10%	2073.81
		5%	1877.38
		1%	1221.58
		0% Min	542.46

Poids corrigés pour la non réponse au quest. grille SILC_I vague 2 (2005)
P_NRGRIL>0

Variable	P_NRGRIL	Quantile	Estimation
N	1100	100% Max	7745.18
Somme des poids	3566395	99%	6926.16
Moyenne	3242	95%	5362.00
Ecart-type	1122	90%	4804.82
Variance	1258301	75% Q3	3881.99
Etendue	6057	50% Médiane	2899.03
Intervalle interquartile	1459	25% Q1	2423.11
Rapport Max/Min	4.59	10%	2133.81
		5%	1989.87
		1%	1806.07
		0% Min	1688.21

Poids initiaux corrigé pour la non réponse au questionnaire non réponse SILC_I vague 2 (2005)
POIDINIT2>0

Variable	POIDINIT2	Quantile	Estimation
N	1104	100% Max	2863.47
Somme des poids	1655186	99%	2497.88
Moyenne	1499	95%	2230.38
Ecart-type	330	90%	2086.17
Variance	109147	75% Q3	1579.68
Etendue	1711	50% Médiane	1387.55
Intervalle interquartile	294	25% Q1	1285.61
Rapport Max/Min	2.48	10%	1188.57
		5%	1174.72
		1%	1152.65
		0% Min	1152.65

Poids initiaux (= poids d'échantillonnage) SILC_I vague 2 (2005)
POIDINIT>0

Variable	POIDINIT	Quantile	Estimation
N	1104	100% Max	1482.62
Somme des poids	1398548	99%	1482.62
Moyenne	1267	95%	1367.33
Ecart-type	92	90%	1367.33
Variance	8433	75% Q3	1334.55
Etendue	330	50% Médiane	1248.83
Intervalle interquartile	169	25% Q1	1165.81
Rapport Max/Min	1.28	10%	1152.65
		5%	1152.65
		1%	1152.65
		0% Min	1152.65

Pondération transversale individuelle finale SILC_I & SILC_II combinés (2005)
POIDTRIN>0

Variable	POIDTRIN	Quantile	Estimation
N	3480	100% Max	9746.64
Somme des poids	6139569	99%	5569.43
Moyenne	1764	95%	3384.48
Ecart-type	901	90%	2737.43
Variance	811157	75% Q3	1975.00
Etendue	9491	50% Médiane	1552.98
Intervalle interquartile	730	25% Q1	1244.73
Rapport Max/Min	38.06	10%	1025.17
		5%	908.79
		1%	712.34
		0% Min	256.06

Pondération transversale ménage finale SILC_I & SILC_II combinés (2005)
POIDTRME>0

Variable	POIDTRME	Quantile	Estimation
N	2268	100% Max	5291.25
Somme des poids	3162427	99%	3327.73
Moyenne	1394	95%	2465.76
Ecart-type	561	90%	2101.21
Variance	315180	75% Q3	1565.90
Etendue	5161	50% Médiane	1302.14
Intervalle interquartile	501	25% Q1	1065.04
Rapport Max/Min	40.75	10%	828.21
		5%	689.24
		1%	507.86
		0% Min	129.84

ANNEXE H – Description détaillée du statut final attribué à chaque adresse, taux de réponse

Figure H-1 SILC_II vague 1.

Echantillon brut	**2035**	**100%**				
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	76	4%				
Hors échantillon (numéro invalide, modem, fax, décédé)	54	2%				
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	92	5%				
Adresses éligibles (statut d'éligibilité connu)	1813	89%		**100%**		
	Rendez-vous grille en cours		43	2%		
	Refus, problèmes niveau grille		507	28%		
	Grilles réalisées		**1263**	**70%**		**100%**
			Refus, problèmes niveau ménage		59	5%
			Grille + quest. ménage réalisés		**1204**	**95%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel	2465		100%
	Rendez-vous individuel en cours	13	0%
	Refus, problèmes niveau individuel	655	27%
	Questionnaires individuel complétés	**1797**	**73%**

Figure H-2 SILC_I vague 2.

Echantillon brut	**1550**	**100%**				
Statut d'éligibilité inconnu (personne ne décroche, adresse contactée, mais statut d'éligibilité inconnu)	18	1%				
Hors échantillon (numéro invalide, modem, fax, décédé)	58	4%				
Adresses non éligibles (résidences secondaires home, institution, EMS, prison, adresses professionnelles)	6	0%				
Adresses éligibles (statut d'éligibilité connu)	1648	95%		**100%**		
	Rendez-vous grille en cours		40	3%		
	Refus, problèmes niveau grille		325	22%		
	Grilles réalisées		**1103**	**75%**		**100%**
			Refus, problèmes niveau ménage		39	3%
			Grille + quest. ménage réalisés		**1064**	**97%**

Niveau des individus:

Nombre d'individus dans les ménages, éligibles pour quest. individuel	2160		100%
	Rendez-vous individuel en cours	6	0%
	Refus, problèmes niveau individuel	471	22%
	Questionnaires individuel complétés	**1683**	**78%**

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodology reports published by the SFSO's Statistical Methods Unit

Graf, E. (2008). Pondérations du SILC pilote: SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro de commande: 338-0051

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050

Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049

Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048

Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze.Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046

Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045

Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044

Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043

Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042

Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041

Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040

Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039

Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037

Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036

Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035

Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033

Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032

Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031

Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030

Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029

Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028

Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027

Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026

Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025

Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024

Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023

Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022

Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021

Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020

Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019

Hulliger, B. (2003). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0018

Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017

Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer:
338-0016

Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande : 338-0015

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer:
338-0014

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013

Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d'échantillonnage. Numéro de commande: 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011

Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010

Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009

Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008

Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007

Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006

Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005

Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003

Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active: rapport de pondération pour 1996.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse: établissement du plan d'échantillonnage.

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Hulliger, B. (1996). Haushalterhebung Familie 1994: Stichprobenplan, Stichprobenziehung und Reservestichproben.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994: plan d'échantillonnage et pondération.

Hulliger, B. (1996). Gütertransportstatistik 1993: Schätzverfahren mit Kompensation der Antwortausfälle.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Peters, R. (1995). Enquête de structure sur les loyers: établissement d'un plan d'échantillonnage stratifié.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung: Stichprobenplan und Schätzverfahren.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Peters, R., Hulliger, B. (1994). La technique de pondération des données: application à l'enquête suisse sur la santé.

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse: stratégie d'échantillonnage pour le secteur privé.

Publikationsprogramm BFS

Das Bundesamt für Statistik (BFS) hat – als zentrale Statistikstelle des Bundes – die Aufgabe, statistische Informationen breiten Benutzerkreisen zur Verfügung zu stellen.

Die Verbreitung der statistischen Information geschieht gegliedert nach Fachbereichen (vgl. Umschlagseite 2) und mit verschiedenen Mitteln

Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

Diffusionsmittel	*Kontakt* *N° à composer*	*Moyen de diffusion*
Individuelle Auskünfte	032 713 60 11 info@bfs.admin.ch	Service de renseignements individuels
Das BFS im Internet	www.statistik.admin.ch	L'OFS sur Internet
Medienmitteilungen zur raschen Information der Öffentlichkeit über die neusten Ergebnisse	www.news-stat.admin.ch	Communiqués de presse: information rapide concernant les résultats les plus récents
Publikationen zur vertieften Information (zum Teil auch als Diskette/CD-Rom)	032 713 60 60 order@bfs.admin.ch	Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)
Online-Datenbank	032 713 60 86 www.statweb.admin.ch	Banque de données (accessible en ligne)

Nähere Angaben zu den verschiedenen Diffusionsmitteln liefert das laufend nachgeführte Publikationsverzeichnis im Internet unter der Adresse www.statistik.admin.ch→Aktuell→Publikationen.

La liste des publications, mise à jour régulièrement, donne davantage de détails sur les divers moyens de diffusion. Elle se trouve sur Internet à l'adresse www.statistique.admin.ch →Actualités→Publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodology Reports by the Statistical Methods Unit

Die Methodenberichte beschreiben die mathematischen und statistischen Methoden, die den Resultaten und Analysen der öffentlichen Statistik zu Grunde liegen. Sie enthalten ausserdem die Evaluation und Entwicklung von neuen Methoden im Hinblick auf eine zukünftige Anwendung. Diese Publikationen sollen einerseits die verwendeten Methoden dokumentieren, um Transparenz und Wissenschaftlichkeit sicher zu stellen, und sie sollen andererseits die Zusammenarbeit mit den Hochschulen und der Wissenschaft fördern.

Zur Illustration der beschriebenen mathematischen Konzepte, werden im Bericht numerische Resultate aufgeführt. Diese sind allerdings nicht als offizielle Resultate der betreffenden Erhebungen zu verstehen. Ebenfalls können die tatsächlich angewendeten Methoden leicht von den hier beschriebenen abweichen.

Die Methodenberichte sind auf der Internetseite des BFS in elektronischer Form verfügbar.

Les rapports de méthodes décrivent les méthodes mathématiques et statistiques à la base des résultats et des analyses de la statistique publique. Ils présentent également l'évaluation et le développement de nouvelles méthodes en vue d'une application future. Ces publications visent d'une part à documenter les méthodes utilisées ou envisagées dans un souci de transparence et de rigueur scientifique, et d'autre part à favoriser la collaboration avec le monde scientifique et universitaire.

Les résultats numériques présentés dans les rapports de méthodes illustrent les concepts mathématiques décrits, mais ne sont pas des résultats officiels des enquêtes concernées. De même, les méthodes réellement appliquées peuvent différer légèrement de celles décrites dans ces rapports.

Les rapports de méthodes sont disponibles sous forme électronique sur le site internet de l'OFS.

La structure du document est la suivante: la façon dont les données ont été collectées est brièvement décrite pour chacun des panels, puis sont définis les différents concepts liés à la pondération. Ensuite, les étapes de la pondération sont présentées, puis décrites une à une pour chacun des panels également. On distingue les étapes communes à toutes les pondérations des étapes propres à chacune d'elles. On traite aussi séparément les pondérations au niveau ménage de celles au niveau individuel. Les deux dernières sections font l'objet de la pondération combinée du SILC_I vague 2 et SILC_II vague 1. Enfin, le lecteur intéressé trouvera toute une série d'annexes détaillant certaines étapes qui ne sont mentionnées que brièvement dans le texte principal.

N° de commande: 338-0051
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 978-3-303-00406-7