

2018

00

Bases
statistiques
et généralités

Neuchâtel 2023

Relevé structurel de la population suisse 2018

Révision de la pondération et sélections répétées



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

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

Rapports de méthodes de la section Méthodes statistiques

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.

Relevé structurel de la population suisse 2018

Révision de la pondération et sélections répétées

Rédaction Lionel Qualité, OFS
Éditeur Office fédéral de la statistique (OFS)

Neuchâtel 2023

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Lionel Qualité, OFS, tél. +41 58 467 24 54
Lionel.qualite@bfs.admin.ch

Rédaction: Lionel Qualité, OFS

Contenu: Section méthodes statistiques, OFS

Série: Statistique de la Suisse

Domaine: 00 Bases statistiques et généralités

Langue du texte original: français

Mise en page: Publishing et diffusion PUB, OFS
Ce document a été produit automatiquement à partir de banques de données. Il ne répond donc pas aux normes typographiques des publications de l'OFS.

En ligne: www.statistique.ch

Imprimés: www.statistique.ch
Office fédéral de la statistique, CH-2010 Neuchâtel,
order@bfs.admin.ch, tél. +41 58 463 60 60
Impression réalisée en Suisse

Copyright: OFS, Neuchâtel 2023
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.

Numéro OFS: 338-0087

ISBN: 978-3-303-00728-0

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Brève description de la pondération jusqu'en 2017	7
2.1 Étapes du calcul	7
2.2 Données auxiliaires	9
3 Nouvelles sources de données	10
3.1 Données des caisses de compensation (AVS)	10
3.2 Registre des entreprises et établissements (REE)	12
3.3 Registre fédéral des bâtiments et logements (RegBL)	13
4 Modèles de réponse	13
4.1 Taux de réponse selon les caractéristiques enregistrées dans STATPOP	14
4.2 Nouvelles variables explicatives	17
4.2.1 Utilisation pour modéliser la non-réponse	17
4.2.2 Utilisation pour modéliser l'élimination des ménages différents de STATPOP	20
4.3 Effet de l'introduction des nouvelles variables sur les poids des personnes avant calage	23
4.4 Effet de l'introduction des nouvelles variables sur les poids des ménages avant calage	25
5 Calage	27
5.1 Nouvelles variables de calage	27
5.2 Effet de l'introduction des nouvelles variables sur les poids calés	29
6 Effet du changement de pondération sur quelques statistiques d'intérêt	31
7 Sélections répétées de personnes et de ménages dans l'échantillon du relevé structurel	33
7.1 Objectifs de la coordination des échantillons	33
7.2 Historique du système de coordination	34
7.2.1 En 2010 - prototype	34
7.2.2 De 2010 à 2012	35
7.2.3 De 2013 à 2018	35
7.2.4 Depuis fin 2018	36
7.3 Sélections répétées de personnes	36
7.4 Sélections répétées de ménages	37
8 Développements en cours et conclusions	42
Annexes	44
A Approximation des probabilités d'inclusion des ménages	44
B Taux de réponse des personnes et ménages selon les caractéristiques STATPOP	44
Bibliographie	46
Rapports de méthodes de la section méthodes statistiques de l'OFS	47

Préambule

L'Office Fédéral de la Statistique (OFS) évalue en permanence les méthodes et pratiques appliquées pour la sélection des échantillons et leur utilisation. Ces méthodes et pratiques sont révisées périodiquement. La mise à disposition, la préparation et l'utilisation de données nouvelles, par exemple des sources administratives, permettent d'améliorer la précision des résultats d'enquête et de réduire la charge associée. L'intégration des données sur les revenus fournies par les caisses de compensation constitue une avancée notable pour les statistiques portant sur la population et les ménages. Le niveau de revenu est en effet lié à une grande part des sujets d'intérêt couverts par les enquêtes de l'OFS. Cette variable vient utilement compléter les informations démographiques utilisées traditionnellement, telles que le sexe, l'âge, la nationalité, le lieu de résidence, qui n'ont pas toujours un très grand pouvoir explicatif sur les variables d'intérêt des différents thèmes abordés dans les enquêtes.

Le relevé structurel est un élément du recensement fédéral de la population qui a lieu chaque année depuis 2010. Il complète les informations des registres par des statistiques supplémentaires sur la structure de la population. La section Méthodes statistiques (METH) a révisé la pondération de l'échantillon du relevé structurel 2018 pour intégrer les données des caisses de compensation, ainsi que des informations sur le nombre de pièces des logements et des informations tirées du registre des entreprises et établissements. Ce document décrit les travaux réalisés et les résultats obtenus. Une seconde partie est consacrée aux sélections répétées de personnes et de ménages dans les échantillons produits par l'OFS et en particulier dans ceux du relevé structurel.

L'auteur remercie la section Population (POP) qui est responsable pour la production de l'enquête, en particulier S. Byrne, C. Freymond, J.-P. Kauthen, D. Mersch, M. Spörri et J. Voronkova pour la très bonne et agréable collaboration. Des remerciements supplémentaires vont à l'équipe de direction de METH (J.-P. Renfer, D. Kilchmann, D. Assoulin) ainsi qu'à C. Chevalier et J. Potterat pour la relecture attentive de ce rapport.

Résumé

En 2018, l'Office Fédéral de la Statistique (OFS) a amélioré la pondération de l'échantillon du relevé structurel grâce à l'utilisation de nouvelles données administratives dans son calcul. Les caisses de compensation de l'assurance-vieillesse et survivants (AVS) livrent à l'OFS des informations sur les personnes qui leur sont affiliées. Ces données sont désormais utilisées à différentes étapes du calcul des poids d'extrapolation du relevé structurel. En outre des données provenant du registre des bâtiments et logements (RegBL) et du registre des entreprises et établissements (REE) sont également utilisées. Jusqu'en 2017, la pondération du relevé structurel utilisait uniquement des données issues de la statistique de la population et des ménages STATPOP [voir à ce sujet [Potterat et col., 2019](#)].

Ce rapport présente les traitements introduits pour le relevé structurel 2018, les analyses qui ont conduit à cette adaptation et une étude de l'impact de la modification de la pondération sur les résultats. À la fin de ce document, la question des sélections répétées de personnes et de ménages dans l'échantillon du relevé structurel est également abordée.

1 Introduction

En 2018, l'Office Fédéral de la Statistique (OFS) a amélioré la pondération de l'échantillon du relevé structurel (Strukturerhebung - SE) grâce à l'utilisation de nouvelles données administratives dans son calcul. Les caisses de compensation de l'assurance-vieillesse et survivants (AVS) livrent à l'OFS des informations sur les personnes qui leur sont affiliées. Ces données sont désormais utilisées à différentes étapes du calcul des poids d'extrapolation du relevé structurel. En outre, des données provenant du registre des bâtiments et logements (RegBL) et du registre des entreprises et établissements (REE) sont également utilisées. Jusqu'en 2017, la pondération du relevé structurel utilisait uniquement des données issues de la statistique de la population et des ménages STATPOP.

L'analyse des résultats de l'enquête montre un taux de réponse plus élevé chez les personnes avec des revenus AVS dans le haut de la distribution. L'effet principal de la modification de la pondération est de réajuster les poids en fonction des revenus AVS. La correction conduit à diminuer le poids des répondants ayant un revenu dans le haut de la distribution et à augmenter le poids de ceux ayant un revenu dans le bas de la distribution.

Par ailleurs, plusieurs nouvelles variables liées aux thèmes de l'enquête sont utilisées lors du calage des poids d'extrapolation. Par exemple, le nombre de pièces du logement selon le RegBL est utilisé comme variable de calage : les données de STATPOP et de l'échantillon sont appariées au RegBL et la répartition de l'échantillon selon le nombre de pièces est ajustée sur la répartition de STATPOP selon ce même nombre de pièces.

Le rééquilibrage de la distribution des revenus vers le bas peut être rapproché des effets observés du changement de pondération sur les résultats de la SE2018 : une augmentation du nombre de logements de petite taille au détriment des logements de grande taille, une diminution de la proportion de ménages propriétaires de leur logement et une légère diminution des loyers moyens.

La section 7, à la fin du document, aborde la question des sélections répétées de ménages et de personnes dans l'échantillon du relevé structurel. Il s'avère, en effet, que certaines personnes sont sélectionnées plusieurs fois, parfois même lors de deux ou plusieurs années consécutives, dans l'échantillon du relevé structurel. L'OFS utilise toutefois un système de coordination d'échantillons pour réduire le nombre de ces cas. Ce nombre reste extrêmement faible par rapport à ce qui serait observé si les échantillons étaient sélectionnés sans un tel système de coordination.

2 Brève description de la pondération jusqu'en 2017

2.1 Étapes du calcul

Les détails de la procédure de pondération du relevé structurel sont donnés dans [Potterat et col. \[2019\]](#). Le calcul des poids attribués aux personnes répondantes est réalisé en trois grandes étapes rappelées ci-après. Des poids sont également calculés pour les ménages retenus dans l'échantillon final du relevé structurel. Le calcul de ces poids fait intervenir une étape supplémentaire.

1. Dans un premier temps les probabilités de sélection des unités activées sont calculées sur la base de celles du plan de sondage. Dans le cas du relevé structurel, l'objectif est d'obtenir un poids non seulement pour les personnes cibles de l'enquête mais également

pour les ménages observés. En conséquence, une probabilité de sélection π_k est retenue pour chaque personne cible k et une probabilité de sélection π_m pour chaque ménage m .

2. La non-réponse est modélisée comme le résultat de choix aléatoires et indépendants réalisés par les unités sélectionnées. La probabilité de réponse de chaque unité k est estimée par une valeur p_k . L'utilisation de cette probabilité de réponse dans le calcul des pondérations permet de «redresser» les poids pour corriger des défauts liés au manque de réponse d'une partie de l'échantillon.
3. Enfin, une procédure de calage des poids des personnes cibles et des poids des ménages sur les totaux connus de certaines variables dans la population permet d'augmenter la précision des estimations qui portent sur des caractéristiques liées aux variables de calage.
4. Une particularité du relevé structurel est que seuls les ménages dont la composition déclarée à l'enquête coïncide avec celle enregistrée dans les fichiers de la statistique de la population et des ménages STATPOP sont considérés comme des ménages répondants et conservés dans l'échantillon net des ménages. En effet, les personnes sélectionnées pour l'enquête reçoivent pour instruction de déclarer comme faisant partie de leur ménage les personnes qui y sont enregistrées au 31 décembre précédant l'enquête, ce qui correspond au concept utilisé pour STATPOP. Il s'avère cependant qu'une part non négligeable, de l'ordre de 10 à 15%, des ménages déclarés au relevé structurel ont une composition différente de celle enregistrée dans STATPOP. Dans la majorité des cas, le ménage observé à l'enquête est plus petit que le ménage enregistré dans STATPOP. Ceci s'explique au moins partiellement par des réponses incomplètes. Cette élimination des ménages qui ne sont pas conformes à STATPOP est également considérée comme le résultat d'événements aléatoires indépendants. En conséquence, une «probabilité de déclarer exactement la même composition de ménage que STATPOP», notée s_k , est estimée pour chaque personne sélectionnable. Ces probabilités estimées sont ensuite utilisées lors du calcul des poids d'extrapolation pour les ménages, avant l'étape de calage.

Il y a ainsi deux échantillons qui sont pondérés. Le premier échantillon est constitué de toutes les personnes cibles qui ont répondu au questionnaire les concernant. Pour cet échantillon, les poids des personnes cibles sont calculés, mais aucun poids n'est attribué aux ménages puisque 10 à 15% d'entre eux sont considérés comme incorrects. Le deuxième échantillon pondéré est celui des ménages et des personnes cibles des ménages qui sont identiques à la composition de STATPOP. Pour cet échantillon, un poids est calculé pour les personnes cibles, les ménages et les personnes composant les ménages (leur poids est égal à celui du ménage).

Le détail du calcul des poids pour les personnes et les ménages est donné ci-après. Soit π_k la probabilité de sélection d'une personne k dans le relevé structurel, p_k sa probabilité de réponse estimée, s_k la probabilité estimée que la personne k déclare exactement la même composition de ménage que celle enregistrée dans STATPOP et g_k un facteur de calage appliqué à l'unité k , alors le poids attribué à la personne cible k est :

$$w_k = \frac{g_k}{\pi_k p_k},$$

pour le premier échantillon et

$$w_k = \frac{g_k}{\pi_k p_k s_k},$$

pour le deuxième échantillon. La valeur de g_k peut être différente dans ces deux échantillons.

Le poids attribué à un ménage m est quant à lui donné par :

$$w_m = \frac{g_m}{\pi_m p_m s_m},$$

où :

$$\pi_m = \sum_{k \in m} \pi_k, \quad (1)$$

$$p_m = \pi_m^{-1} \sum_{k \in m} \pi_k p_k, \quad (2)$$

$$s_m = (\pi_m p_m)^{-1} \sum_{k \in m} \pi_k p_k s_k, \quad (3)$$

et g_m est un facteur de calage. Les poids ménages ainsi calculés sont le résultat d'une approximation, discutée dans l'annexe A.

Les facteurs de calage g_k (resp. g_m) sont des facteurs d'ajustement les plus proches possible de 1 qui permettent d'obtenir l'égalité

$$\sum_{k \in S} w_k \mathbf{x}_k = \mathbf{X}, \quad \left(\text{resp.} \quad \sum_{m \in S_M} w_m \mathbf{x}_m = \mathbf{X} \right),$$

pour des variables auxiliaires $\mathbf{x}_k$ concernant les personnes cibles k de l'échantillon S (resp. les ménages m de l'échantillon de ménages S_M), dont les totaux $\mathbf{X}$ sur la population sont connus [voir [Deville et Särndal, 1992](#)]. Lorsque les totaux de calage des poids ménages et personnes sont compatibles, il est possible d'imposer en outre dans le deuxième échantillon $g_k = g_m$ si k est dans le ménage m [voir [Le Guennec et Sautory, 2002](#)]. Cela n'est pas le cas actuellement, mais fait partie des développements futurs possibles.

2.2 Données auxiliaires

Les données utilisées pour calculer les poids d'extrapolation des relevés structurels jusqu'en 2017 proviennent essentiellement des fichiers de la Statistique de la population et des ménages STATPOP. Pour une petite partie des unités, 2386 en 2018, soit 0.7% de l'échantillon brut, le numéro AVS n'a pas été retrouvé dans STATPOP. Les informations utilisées pour celles-ci proviennent du cadre de sondage SRPH (Stichprobenrahmen für Personen und Haushalte). Ces fichiers sont construits en utilisant les données des registres de personnes des communes, des cantons et de la Confédération. Les caractères pris en compte portent sur le lieu de résidence principale, l'âge, le sexe, l'état-civil, la nationalité et le type de permis de séjour ainsi que sur la composition du ménage.

Les données de STATPOP à disposition pour la pondération de la SE2018 concernent 8 356 225 personnes résidentes permanentes en Suisse, dont 7 084 068 sont âgées de 15 ans ou plus. Ces personnes se répartissent dans 3 755 689 ménages (respectivement 3 754 954 ménages pour les seules personnes âgées de 15 ans ou plus). La pyramide des âges de la population observée dans STATPOP est donnée dans la figure 1.

Les valeurs des paramètres p_k et s_k sont obtenues en effectuant des prédictions au moyen de procédures de régression logistique [voir [Potterat et col., 2019](#), p.13 et 17]. Les prédictions sont basées sur plusieurs caractéristiques des personnes et du ménage. La modélisation des probabilités de réponse p_k fait intervenir la classe d'âge, le sexe, l'état civil, la nationalité, le permis

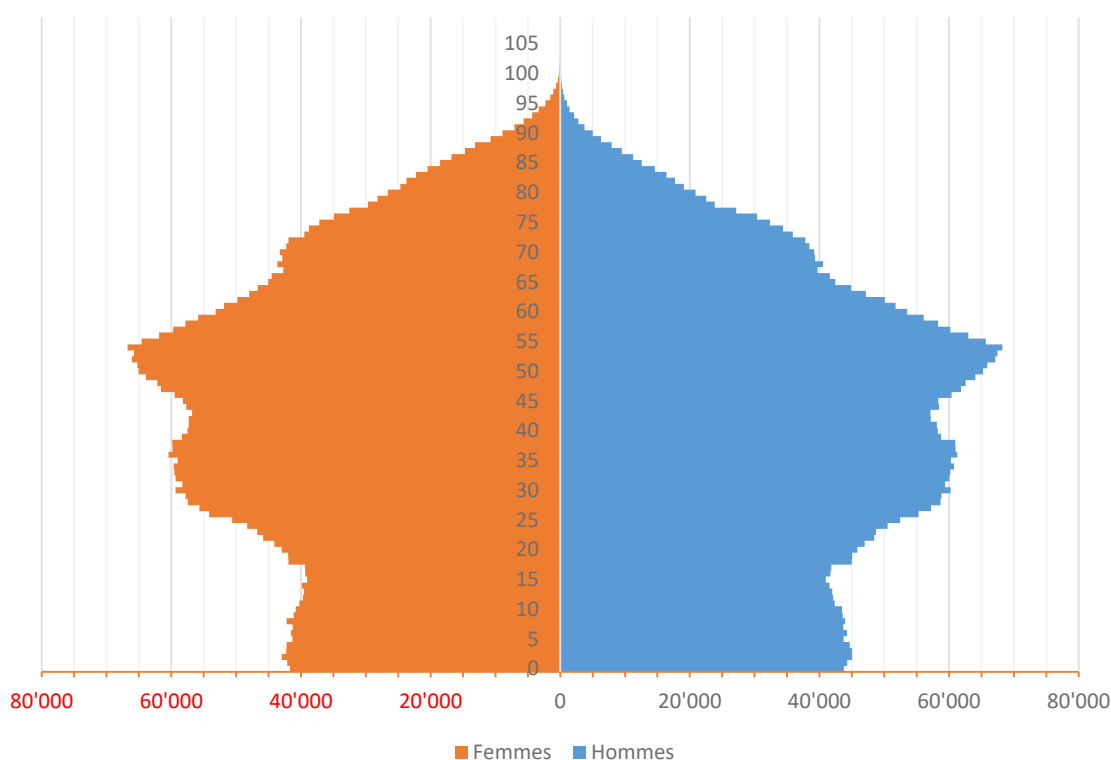


FIGURE 1 Pyramide des âges, STATPOP 2018

de séjour, le type de ménage, la taille du ménage et la zone géographique. Les probabilités s_k sont prédites à l'aide du type de ménage, du permis de séjour et de la zone géographique. Les paramètres des modèles de régression logistique sont ajustés sur l'échantillon de personnes cibles puis utilisés pour prédire p_k et s_k pour tous les membres des ménages sélectionnés. Ces valeurs servent ensuite à calculer les valeurs de p_m et s_m pour les ménages enquêtés.

Les facteurs de calage g_k et g_m sont calculés en utilisant également des variables de calage construites à partir des informations relevées dans STATPOP. La description complète de ces variables de calage se trouve en pages 14, 15 et 21 de [Potterat et col. \[2019\]](#). Là encore interviennent l'âge, le sexe, la nationalité, l'état-civil, la zone géographique, le type et la taille du ménage.

3 Nouvelles sources de données

Cette section présente les trois sources de données nouvellement utilisées pour la pondération du relevé structurel à partir de 2018.

3.1 Données des caisses de compensation (AVS)

Les caisses de compensation AVS mettent à disposition de l'OFS des données sur les revenus assurés des personnes employées en Suisse. Ces données sont traitées et gérées par la section URD (Unternehmensregisterdaten) de l'OFS. Elles permettent, pour chaque employé, de connaître un numéro d'identification de l'employeur, le revenu sur lequel est basé le calcul de la

cotisation pour le premier pilier de la prévoyance vieillesse et les mois d'activité concernés. Ces données sont en outre munies du numéro d'assuré qui est également disponible dans STATPOP. Ainsi, le fichier utilisé permet un appariement partiel entre le fichier STATPOP et le registre des entreprises et établissements.

Les données utilisables par l'OFS ont généralement une ancienneté de 18 mois environ. Les données concernant l'année de cotisation 2017 sont les plus récentes qui ont été utilisées pour pondérer l'enquête SE2018, celle-ci s'étant déroulée au premier semestre 2019. Le fichier à disposition contient des revenus associés à 5 267 746 numéros AVS. Un même numéro AVS peut apparaître plusieurs fois dans le fichier, à priori en lien avec des employeurs différents. La répartition des numéros AVS en fonction du nombre de lignes correspondantes est donnée dans la table 1.

TABLE 1 Nombre de revenus AVS enregistrés selon la fréquence d'apparition dans le fichier des données AVS

Fréquence	Effectif	Proportion (%)
1	4 044 800	76.8
2	910 952	17.3
3	219 232	4.2
4	60 735	1.1
5	19 106	0.4
6 ou plus	12 921	0.2
Total	5 267 746	100.0

Parmi les numéros AVS présents dans les données des caisses de compensation, 689 119 ne sont pas retrouvés dans les fichiers de STATPOP. En effet, outre le décalage temporel entre ces jeux de données, les cotisations au premier pilier ne concernent pas uniquement les résidents permanents mais également par exemple les personnes au bénéfice d'un permis de courte durée ou bien les travailleurs frontaliers.

De la même manière, tous les numéros AVS contenus dans le fichier STATPOP n'ont pas été retrouvés dans les données des caisses de compensation. En effet, sur les 7 084 068 personnes âgées de 15 ans ou plus, seulement 4 578 627 (64.6%) ont été identifiés dans ces données. Les personnes sans activité professionnelle ou avec un très bas revenu sont en principe absentes des données des caisses de compensation, ainsi naturellement que les personnes qui ne sont pas en âge de cotiser. Parmi les hommes âgés de 19 à 66 ans selon STATPOP, qui avaient donc entre 18 et 65 ans en 2017, 87% ont été retrouvés dans les données AVS et parmi les femmes âgées de 19 à 65 ans, 79.6%. La figure 2 illustre les proportions (en %) de personnes présentes dans le fichier STATPOP et retrouvées dans les données AVS selon le sexe et l'âge.

Les revenus enregistrés dans les données des caisses de compensation sont agrégés par numéro AVS. La variable obtenue présente une distribution très asymétrique, avec quelques valeurs largement plus grandes que les autres. La valeur moyenne de ces revenus est de 68 585 CHF. Quelques quantiles de la distribution figurent dans la table 2.

Cette variable ne peut directement être comparée avec les autres données disponibles à l'OFS sur les salaires et revenus des personnes. Elle ne tient en effet pas compte du taux d'occupation des personnes concernées et ne donne aucune information sur les transferts entre personnes. Par ailleurs certaines franchises sont applicables et la cotisation est facultative en dessous d'un certain seuil.

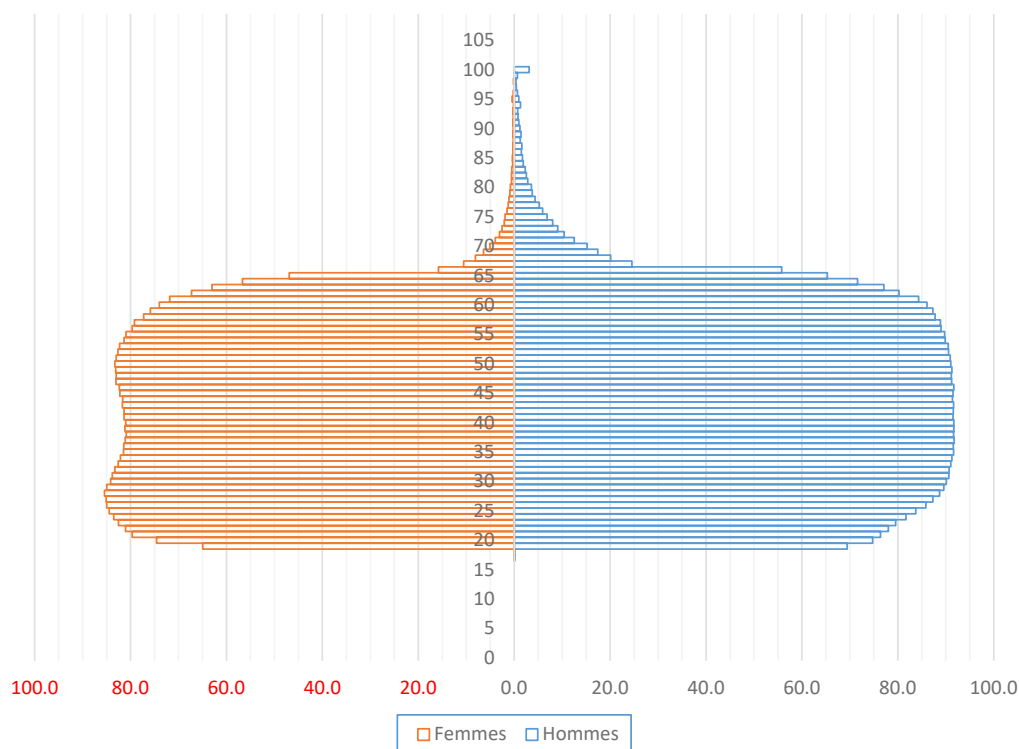


FIGURE 2 Proportion de la population STATPOP 2018 retrouvée dans les données AVS par sexe et âge

TABLE 2 Quantiles de la distribution des revenus sommés par numéros AVS

Quantile	Revenu
1%	860
5%	4 418
10%	9 547
25%	27 240
50%	58 060
75%	87 058
90%	126 414
95%	162 150
99%	311 758

Plus de 99% des numéros AVS présents dans les données des caisses de compensation y sont associés à au moins un identifiant d'entreprise. C'est le cas de 5 221 173 numéros AVS (99.1%) parmi les 5 267 746 du fichier complet et 4 544 023 (99.2%) parmi les 4 578 627 numéros retrouvés dans STATPOP.

3.2 Registre des entreprises et établissements (REE)

Le registre des entreprises et établissements contient des informations sur «tous les établissements de droit privé et public établis et générant une activité économique sur le territoire Suisse» (source : OFS). Il permet entre autres de connaître leur localisation, leur forme juridique et leur catégorie d'activité économique. Le REE est alimenté en continu à partir de multiples

sources administratives et statistiques.

L'extrait du registre des entreprises utilisé pour la pondération de l'enquête SE2018 date du 30 novembre 2018 et contient 877 373 enregistrements. L'appariement de ce fichier avec les données AVS permet d'identifier au moins un employeur dans le registre pour 4 407 907 numéros AVS parmi les 4 578 627 numéros retrouvés dans STATPOP, soit 96.3% de cette sous-population.

3.3 Registre fédéral des bâtiments et logements (RegBL)

Le registre fédéral des bâtiments et logements, géré par l'OFS, contient entre autres des informations sur les bâtiments et logements occupés par la population résidente en Suisse. Les ménages de STATPOP sont accompagnés d'identifiants de leur résidence principale, ce qui permet de faire un lien entre les ménages de STATPOP et les bâtiments et logements du RegBL. Les données du RegBL étaient jusqu'alors utilisées lors de la préparation statistique des données du relevé structurel, mais pas directement dans le calcul des poids d'extrapolation.

L'extrait de registre utilisé pour la pondération contient, en 2018, 4 817 285 logements identifiés par un numéro de bâtiment (EGID) et un numéro de logement (EWID). Parmi ces enregistrements, 3 749 807 (77.8%) sont désignés comme étant des résidences principales. Les informations disponibles dans le registre sont notamment l'année de construction du bâtiment, l'étage, le nombre de pièces et la superficie du logement.

Le lien entre les ménages d'intérêt de STATPOP et les logements enregistrés dans le RegBL a pu être établi pour 99.5% d'entre eux, soit 3 736 453 parmi 3 755 689 ménages au total. La répartition de ces ménages en fonction du nombre de pièces du logement figure dans la table 3. Le nombre de pièces considéré ici est celui enregistré dans le RegBL. Il ne coïncide pas nécessairement avec le nombre de pièces déclaré lors de l'enquête ou avec les conventions locales sur la manière de désigner les appartements.

TABLE 3 Répartition des ménages selon le nombre de pièces du logement dans le RegBL (2018)

Nombre de pièces	Effectif	Proportion (%)
manquant	9	0.0
1	180 387	4.8
2	479 536	12.8
3	986 998	26.4
4	1 081 189	28.9
5	619 575	16.6
6	256 872	6.9
7	84 018	2.3
8 ou plus	47 869	1.3
Total	3 736 453	100.0

4 Modèles de réponse

Les données disponibles pour l'ensemble de l'échantillon brut peuvent être utilisées pour construire un modèle et permettre une estimation des probabilités de réponse des unités sélectionnées. Le modèle retenu pour le relevé structurel est un modèle logistique, selon lequel la probabilité

ρ_k qu'une unité k réponde à l'enquête est strictement comprise entre 0 et 1 et est définie par la relation fonctionnelle :

$$\text{Log} \left(\frac{\rho_k}{1 - \rho_k} \right) = \sum_{i=1}^p \beta_i x_{i,k},$$

où les $x_{i,k}$, $i = 1, \dots, p$ sont les valeurs de p variables *explicatives* pour l'unité k et les β_i , $i = 1, \dots, p$ sont les coefficients du modèle. Les valeurs des β_i sont a priori inconnues. Lorsque les prédicteurs x_i sont spécifiés, les valeurs des paramètres β_i peuvent être estimées en utilisant les indicatrices de réponse des unités de l'échantillon et, par exemple, la méthode du maximum de vraisemblance. Une fois les estimations $\hat{\beta}_i$ calculées, les probabilités de réponses prédites p_k sont obtenues comme solutions de l'équation

$$\text{Log} \left(\frac{p_k}{1 - p_k} \right) = \sum_{i=1}^p \hat{\beta}_i x_{i,k}.$$

La même démarche est utilisée pour prédire une probabilité s_k qu'une personne ayant répondu à l'enquête déclare une composition de ménage identique à celle enregistrée dans STATPOP. Le ménage est alors *conservé* dans l'échantillon net. Les prédicteurs utilisés jusqu'en 2017 dans le «modèle de conservation» correspondant sont évoqués dans la section 2.2 et détaillés dans le rapport [Potterat et col. \[2019\]](#).

La théorie indique que seuls les poids calculés à partir des «vraies probabilités de réponse» permettent d'obtenir des estimateurs sans biais des totaux de toutes les variables d'intérêt possibles. À défaut de connaître ces probabilités de réponse (en supposant que la décision de répondre ou de déclarer une composition de ménage résulte effectivement d'un mécanisme aléatoire), il serait légitime d'essayer d'en obtenir la meilleure prédiction possible. Toutefois, [Haziza et col. \[2022\]](#) notent que l'utilisation d'un modèle de réponse complexe conduit à des probabilités prédites dispersées. Cela a tendance à faire augmenter la variance des estimateurs par rapport à une prédiction moins détaillée. Selon les auteurs, si les comportements de réponse ne sont pas du tout liés aux variables d'intérêt de l'enquête, une estimation simple avec une correction de non-réponse uniforme ne présente pas de biais et conduit à des estimateurs plus précis qu'une estimation avec une correction complète de la non-réponse. Ainsi, les prédicteurs utilisés dans le modèle de réponse devraient être liés à la fois au comportement de réponse et aux variables d'intérêt de l'enquête afin de minimiser le risque de biais et d'une possible augmentation de la variance des estimateurs.

Le caractère généraliste du questionnaire du relevé structurel explique que quasiment toutes les informations dont l'OFS pourrait disposer sur la population sont liées aux variables d'intérêt de l'enquête. C'est le cas des données de base sur les personnes et ménages disponibles dans STATPOP, mais aussi des informations, même imparfaites ou faisant référence à une période passée qui sont décrites dans la section 3. Le lien entre ces variables et la propension à répondre des unités sélectionnées est illustré dans les sections 4.1 et 4.2.

4.1 Taux de réponse selon les caractéristiques enregistrées dans STATPOP

Les tailles d'échantillon brut, nombre de répondants et taux de réponses observés dans le relevé structurel 2018 selon les principales caractéristiques individuelles utilisées dans la pondération avant 2018 sont donnés dans les tables 4, 5 et dans les tables 25, 26, 27, et 28 de l'annexe B. Ces tables mettent en évidence les taux de réponse sensiblement différents selon les caractéristiques considérées. Cela justifie leur utilisation dans le modèle de prédiction des probabilités de réponse.

TABLE 4 Échantillon selon le permis de résidence

Permis	Éch. brut	Éch. net	Tx. réponse (%)
Suisse	234 943	205 585	87.5
C	51 156	40 413	79.0
B et autres	28 056	20 539	73.2
Total	314 155	266 537	84.8

TABLE 5 Échantillon selon l'âge

Classe d'âge	Éch. brut	Éch. net	Tx. réponse (%)
15 - 19	18 379	15 014	81.7
20 - 24	20 475	15 602	76.2
25 - 29	24 456	18 938	77.4
30 - 34	26 441	21 412	81.0
35 - 39	26 640	22 183	83.3
40 - 44	25 857	21 834	84.4
45 - 49	27 279	23 296	85.4
50 - 54	29 238	25 243	86.3
55 - 59	26 676	23 408	87.7
60 - 64	21 940	19 531	89.0
65 - 69	18 633	16 986	91.2
70 - 74	17 548	16 225	92.5
75 - 79	13 193	12 063	91.4
80 - 84	9 218	8 176	88.7
85 ans ou plus	8 182	6 626	81.0
Total	314 155	266 537	84.8

Concernant la proportion des ménages qui sont identiques dans le relevé structurel et dans STATPOP, un examen attentif montre des disparités sensibles entre certains cantons et communes. La composition du ménage enregistrée dans STATPOP (taille ou nombre d'adultes et d'enfants) joue aussi un rôle. Les tailles d'échantillon avant et après élimination des ménages, ainsi que les taux correspondants sont donnés dans les tables 6 et 7. La table 6 fait apparaître que le taux de conservation est beaucoup plus élevé pour les ménages d'une personne (96.4%) que pour les ménages de six personnes ou plus (seulement 49.2%).

TABLE 6 Échantillon de ménages conservés selon la taille du ménage dans STATPOP

Taille du ménage	Éch. net	Éch. conservé	Tx. conserv. (%)
1 pers.	49 361	47 562	96.4
2 pers.	93 107	82 733	88.9
3 pers.	46 486	38 022	81.8
4 pers.	50 229	43 965	87.5
5 pers.	18 699	14 601	78.1
6 pers. ou plus	8 655	4 262	49.2
Total	266 537	231 145	86.7

TABLE 7 Échantillon de ménages conservés selon le canton de résidence

Canton	Éch. net	Éch. conservé	Tx. conserv. (%)
Argovie	16 321	14 726	90.2
Appenzell Rh.-Int.	360	308	85.6
Appenzell Rh.-Ext.	1 336	1 140	85.3
Berne	29 053	24 812	85.4
<i>dont ville de Berne</i>	<i>6 454</i>	<i>4 769</i>	<i>73.9</i>
<i>dont ville de Bienne</i>	<i>1 288</i>	<i>990</i>	<i>76.9</i>
Bâle campagne	7 130	6 345	89.0
Bâle ville	4 368	3 673	84.1
Fribourg	7 871	6 915	87.9
Genève	21 146	15 418	72.9
Glaris	906	816	90.1
Grisons	5 072	4 361	86.0
Jura	1 888	1 686	89.3
Lucerne	19 167	17 028	88.8
Neuchâtel	8 841	7 955	90.0
Nidwald	1 169	1 002	85.7
Obwald	1 003	870	86.7
Saint-Gall	12 354	10 960	88.7
Schaffouse	2 034	1 810	89.0
Soleure	6 723	5 943	88.4
Schwytz	3 945	3 468	87.9
Thurgovie	6 713	5 986	89.2
Tessin	18 060	16 060	88.9
Uri	913	799	87.5
Vaud	39 056	34 118	87.4
Valais	8 578	7 316	85.3
Zoug	6 403	5 796	90.5
Zürich	36 127	31 834	88.1
Total	266 537	231 145	86.7

4.2 Nouvelles variables explicatives

Les nouvelles données disponibles permettent d'améliorer la prédiction des probabilités de réponse des personnes sélectionnées et des probabilités d'observer un ménage conforme à la composition enregistrée dans STATPOP. Ces deux points sont développés dans les sections 4.2.1 et 4.2.2.

4.2.1 Utilisation pour modéliser la non-réponse

L'analyse des résultats montre un taux de réponse sensiblement plus élevé chez les personnes avec des revenus AVS dans le haut de la distribution que chez les personnes avec des revenus dans le bas de la distribution (voir table 8).

TABLE 8 Échantillon selon le quantile du revenu AVS

Quantile du revenu AVS (%)				Éch. brut	Éch. net	Tx. réponse (%)
00 - 05	(≤	4 417	CHF)	8 471	6 714	79.3
05 - 10	(≤	9 545	CHF)	9 302	7 477	80.4
10 - 15	(≤	14 613	CHF)	9 672	7 698	79.6
15 - 20	(≤	20 525	CHF)	9 705	7 858	81.0
20 - 25	(≤	27 240	CHF)	10 005	8 203	82.0
25 - 30	(≤	33 772	CHF)	9 953	8 196	82.3
30 - 35	(≤	40 334	CHF)	10 008	8 312	83.1
35 - 40	(≤	46 800	CHF)	10 059	8 266	82.2
40 - 45	(≤	52 645	CHF)	10 028	8 301	82.8
45 - 50	(≤	58 059	CHF)	10 130	8 452	83.4
50 - 55	(≤	63 200	CHF)	10 091	8 464	83.9
55 - 60	(≤	68 480	CHF)	10 137	8 520	84.0
60 - 65	(≤	73 991	CHF)	10 184	8 754	86.0
65 - 70	(≤	80 000	CHF)	10 345	9 028	87.3
70 - 75	(≤	87 055	CHF)	10 336	9 191	88.9
75 - 80	(≤	95 922	CHF)	10 310	9 249	89.7
80 - 85	(≤	108 000	CHF)	10 481	9 475	90.4
85 - 90	(≤	126 414	CHF)	10 545	9 594	91.0
90 - 95	(≤	162 145	CHF)	10 712	9 808	91.6
95 - 100	(>	162 145	CHF)	10 697	9 702	90.7
Homme 19 - 66 ans hors AVS				16 416	11 896	72.5
Femme 19 - 65 ans hors AVS				24 766	19 895	80.3
Autre hors AVS				71 802	63 484	88.4
Total				266 537	231 145	86.7

Ce comportement n'a pas été entièrement détecté par le modèle de réponse utilisé jusqu'ici, qui se basait sur les informations disponibles dans STATPOP et le SRPH. Pour compléter le modèle, une variable catégorielle a été créée avec les modalités suivantes :

- (A) : personne hors données AVS qui est soit âgée de moins de 19 ans¹, soit un homme de plus de 66 ans, soit une femme de plus de 65 ans,
- (B) : homme de 19 à 66 ans hors AVS,
- (C) : femme de 19 à 65 ans hors AVS,

1. Les limites d'âge ont été choisies pour tenir compte du décalage d'un an entre les données AVS et le déroulement de l'enquête.

- 20 groupes de personnes dont les revenus AVS totaux sont respectivement inférieur au cinquième percentile de la distribution des revenus AVS, compris entre le cinquième et le dixième percentile, le dixième et le quinzième, etc.

Cette variable catégorielle a été utilisée comme prédicteur supplémentaire dans le modèle de réponse.

Outre le revenu, il a été constaté que la forme juridique de l'entreprise (public/privé) et la catégorie d'activité sont également liées aux taux de réponse observés. Ces caractéristiques n'étaient pas compensées par l'ancien modèle de réponse utilisé pour pondérer le relevé structurel. Ajouter la variable catégorielle calculée à partir du revenu ne suffit pas non plus à compenser ces taux de réponse différents.

Dans le cas où un même numéro AVS apparaît plusieurs fois dans les données des caisses de compensation, un employeur de référence a été défini arbitrairement en choisissant, parmi les employeurs chez lequel l'employé était actif lors de son dernier mois d'activité dans l'année, celui qui correspond au plus grand revenu annuel.

Deux variables catégorielles supplémentaires ont été utilisées pour améliorer le modèle de réponse. La première est construite à partir de la forme juridique de l'employeur de référence de la personne cible. Elle indique si aucun employeur de référence n'a été déterminé ou bien si celui-ci appartient au secteur privé ou au secteur public. La seconde contient la catégorie d'activité NOGA au niveau 2 [divisions, voir [OFS, 2008](#)] pour les NOGA ayant un effectif dans la population au moins égal à 20 000. La taille de la NOGA dans la population est calculée en utilisant le fichier des revenus AVS apparié aux données de STATPOP.

Le choix d'une limite de taille pour les NOGA utilisées est nécessaire afin d'assurer une certaine stabilité des estimations et du calage, certaines NOGA au niveau 2 ayant de très petits effectifs. Le niveau supérieur de la nomenclature (sections NOGA) ne permet pas de résoudre ce problème et les taux de réponse ne sont pas homogènes à l'intérieur des sections. Avec une taille limite dans la population de 20 000 et en considérant les taux de sondage utilisés pour le relevé structurel, les NOGA retenues ont un effectif généralement supérieur à 800 (4%) dans l'échantillon.

Les taux de réponse constatés en fonction de la position du revenu dans la distribution, les taux prédits selon l'ancien modèle et prédits avec le nouveau modèle utilisant les trois nouveaux prédicteurs sont représentés dans la figure 3. Chaque barre dans ce graphique représente soit la proportion de répondants observée (barres bleues), soit la moyenne des probabilités de réponse prédites sur l'échantillon brut pour la catégorie de population concernée (barres grises pour le nouveau modèle et oranges pour l'ancien modèle). Le nouveau modèle de réponse corrige complètement les écarts entre taux de réponse observés et taux de réponse prédits en fonction du revenu (voir figures 3 et 4).

Une analyse séparée par sexe a par ailleurs été réalisée et a montré que le nouveau modèle de réponse prédit correctement les taux de réponse différents des hommes et des femmes en fonction du revenu AVS. Elle met en évidence que les taux de réponse des femmes varient de 82% pour les revenus du bas de la distribution à 91% pour les revenus du haut de la distribution alors que les taux de réponse des hommes varient de 75% pour les revenus du bas de la distribution à 91% pour les revenus du haut de la distribution.

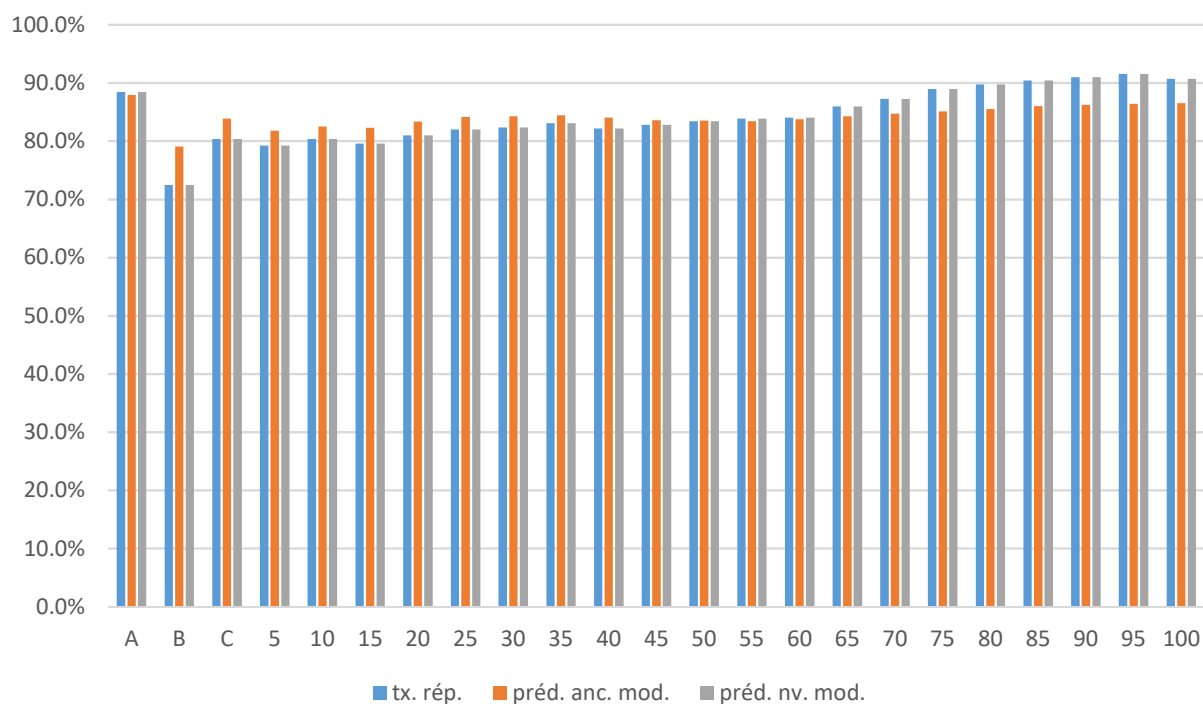


FIGURE 3 Taux de réponse observés et estimés selon le percentile de revenu AVS

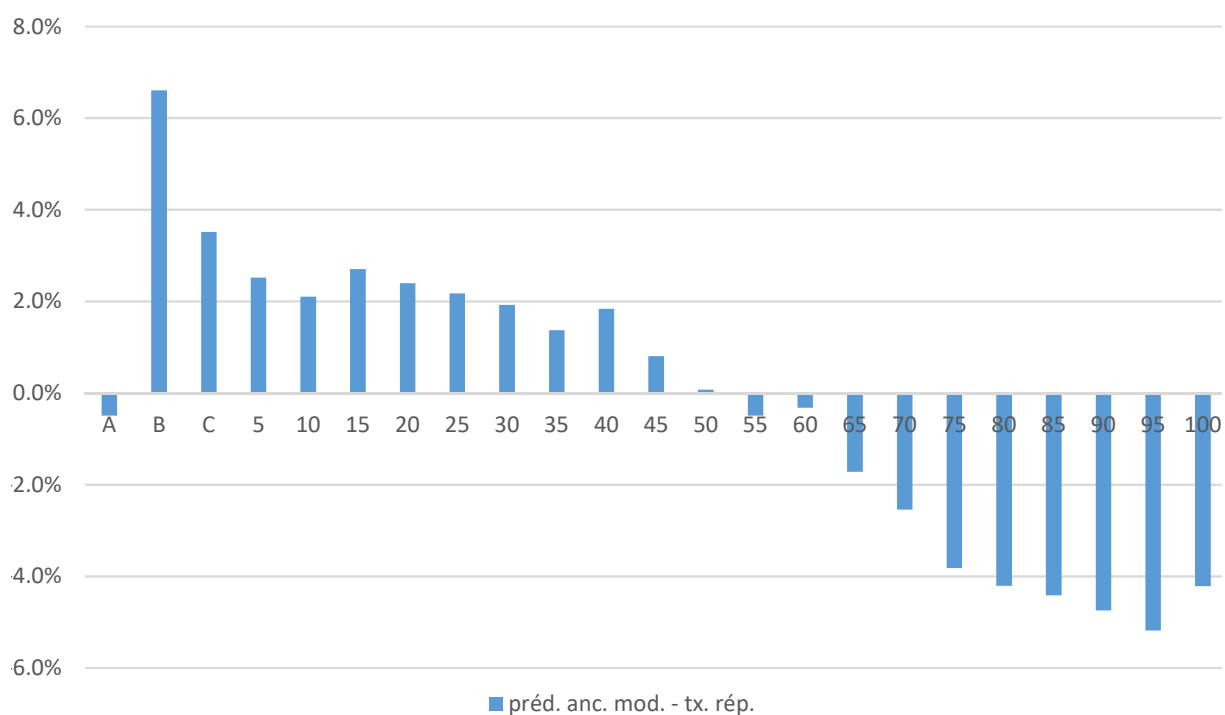


FIGURE 4 Erreur de prédiction du taux de réponse avec l'ancien modèle en fonction du revenu AVS

La table 9 présente, en fonction de la forme juridique, la taille d'échantillon avant non-réponse, le taux de réponse observé et les taux de réponse prédits en utilisant l'ancien et le nouveau modèle de prédiction. Ce dernier utilise les trois nouveaux prédicteurs décrits ci-avant. La table 10 présente les mêmes statistiques détaillées par catégorie d'activité NOGA. Elle met en évidence la disparité entre les taux de réponse selon les catégories d'activité NOGA, et permet de remarquer que le nouveau modèle de réponse produit un bon ajustement des taux de réponse lorsque la taille de l'échantillon dans la NOGA est suffisamment élevée. Les taux de réponse prédits avec le nouveau modèle se rapprochent en effet nettement des taux observés, sans nécessairement leur être parfaitement égaux.

TABLE 9 Taux de réponse observé et estimé par forme juridique

Forme juridique	éch. brut	tx. rép.	anc. mod.	nv. mod.
Sans f. jur.	114 615	84.3	85.7	84.3
Privé	165 168	84.1	84.0	84.1
Public	34 372	90.1	85.9	90.1
Total	314 155	84.8	84.8	84.8

4.2.2 Utilisation pour modéliser l'élimination des ménages différents de STATPOP

Les revenus AVS totaux ont été agrégés au sein des ménages, selon la composition enregistrée dans STATPOP de ces derniers. Ce revenu a ensuite été divisé par un facteur qui rend compte de la composition du ménage (échelle d'équivalence de l'Organisation de Coopération et de Développement Économiques - OCDE : 1 pour la personne la plus âgée, augmenté de 0.5 pour chaque personne supplémentaire âgée de 14 ans ou plus et 0.3 pour chaque personne supplémentaire âgée de moins de 14 ans). Le résultat de cette opération est appelé par la suite revenu AVS équivalent du ménage. Alors que les revenus AVS totaux sont utilisés dans les étapes de modélisation et de calage qui concernent la pondération des personnes cibles, les revenus AVS équivalents sont eux utilisés dans les étapes de modélisation et de calage qui concernent la pondération des ménages.

L'analyse des résultats montre un taux de concordance entre STATPOP et la SE légèrement moins élevé pour les ménages dont le revenu équivalent est dans le bas de la distribution. Pour en tenir compte, une variable catégorielle supplémentaire est introduite dans le modèle de prédiction des s_k . Les modalités de cette variable sont :

- aucun revenu AVS identifié dans le ménage et tous les membres du ménage âgés de moins de 19 ans ou de plus de 66 ans pour les hommes et plus de 65 ans pour les femmes,
- aucun revenu AVS identifié dans le ménage et au moins un homme entre 19 et 66 ans ou une femme entre 19 et 65 ans,
- 10 modalités pour la position du revenu AVS équivalent du ménage dans la distribution (inférieur au premier décile, entre le premier et le deuxième décile, etc.).

En outre, l'analyse des résultats montre un taux de conservation dans l'échantillon relativement plus bas pour les ménages dont le logement n'est pas retrouvé dans le RegBL ou bien pour les ménages dont le logement comporte soit une seule pièce soit au moins huit pièces selon le RegBL. Pour 99.5% des ménages de STATPOP, un logement est retrouvé dans le RegBL. Le nombre de pièces selon le RegBL est corrélé au nombre de pièces déclaré à l'enquête. En conséquence, une variable catégorielle indiquant le nombre de pièces selon le RegBL, limité à 8, ou encore le fait que cette valeur est manquante ou bien que le logement n'est pas retrouvé dans le RegBL est introduite dans le modèle de prédiction des s_k .

TABLE 10 Taux de réponse observé et estimé par NOGA

NOGA	éch. brut	tx. rép.	anc. mod.	nv. mod.	NOGA	éch. brut	tx. rép.	anc. mod.	nv. mod.
1 *	3 155	82.9	85.7	82.9	52 *	1 573	84.1	83.6	84.1
2	161	82.0	85.8	85.8	53 *	1 720	85.6	85.1	85.6
3	24	91.7	83.3	81.9	55 *	2 657	80.4	81.4	80.4
8	155	85.8	86.0	87.3	56 *	6 778	75.7	80.8	75.7
9	5	100.0	86.0	88.5	58	635	85.4	85.2	85.4
10 *	3 315	82.0	84.0	82.0	59	375	83.2	83.2	81.7
11	246	82.5	84.8	84.8	60	515	88.9	85.1	86.2
12	113	94.7	85.5	88.4	61 *	1 099	87.4	84.9	87.4
13	257	85.6	84.0	84.0	62 *	3 444	87.9	83.9	87.9
14	165	83.6	85.4	84.6	63	358	88.0	82.0	84.2
15	66	89.4	83.3	83.0	64 *	5 627	90.0	84.9	90.0
16 *	1 548	82.6	84.7	82.6	65 *	2 232	89.5	86.2	89.5
17	282	84.8	83.8	84.9	66 *	2 766	87.4	85.1	87.4
18	722	86.3	85.6	85.6	68 *	2 498	85.5	85.6	85.5
19	38	97.4	85.5	87.4	69 *	3 436	88.2	85.8	88.2
20 *	1 131	86.8	84.1	86.8	70 *	4 248	86.6	84.2	86.6
21 *	1 081	88.0	83.8	88.0	71 *	4 989	89.2	84.7	89.2
22 *	824	82.2	82.9	82.2	72 *	846	88.1	83.7	88.1
23	650	81.7	83.9	85.0	73 *	1 004	83.6	84.2	83.6
24	421	81.0	82.8	84.3	74 *	1 297	84.6	84.9	84.6
25 *	3 073	84.0	83.8	84.0	75	228	90.4	85.8	85.8
26 *	3 947	87.6	84.6	87.6	77	394	79.4	83.0	83.2
27 *	1 200	85.8	84.3	85.7	78 *	5 061	71.5	78.9	71.5
28 *	2 870	87.7	84.6	87.7	79 *	638	87.1	84.5	87.1
29	132	80.3	82.6	83.4	80 *	897	83.4	82.7	83.4
30	458	87.6	83.9	85.6	81 *	4 432	76.9	81.1	76.9
31	369	85.4	85.1	85.5	82 *	737	80.1	83.4	80.1
32 *	939	87.1	84.2	87.1	84 *	22 147	91.1	86.8	91.1
33	522	84.3	85.0	85.7	85 *	6 156	87.4	83.6	87.4
35 *	1 063	93.4	86.3	93.4	86 *	15 267	88.4	85.6	88.4
36	36	88.9	87.3	87.8	87 *	7 577	85.1	85.1	85.1
37	140	79.3	84.4	85.3	88 *	4 667	84.2	84.6	84.2
38	380	82.1	84.7	85.1	90 *	868	82.0	83.1	82.0
39	11	54.5	79.5	78.5	91	248	90.3	85.6	84.9
41 *	3 548	79.7	82.6	79.7	92	125	85.6	82.4	83.0
42 *	747	81.7	82.7	81.7	93 *	1 449	84.1	83.0	84.1
43 *	9 271	79.5	82.4	79.5	94 *	3 348	89.4	85.7	89.4
45 *	3 276	82.7	83.6	82.7	95	197	85.8	84.3	83.2
46 *	8 964	85.7	84.4	85.7	96 *	2 673	81.3	84.9	81.3
47 *	13 748	83.5	84.5	83.5	97	15	80.0	86.3	84.6
49 *	4 548	85.1	84.8	85.1	99	249	73.1	83.8	82.5
50	111	89.2	83.4	84.9	- *	1 631	84.4	83.8	84.4
51	358	89.9	84.0	84.8	- *	112 984	84.3	85.7	84.3
					Total	314 155	84.8	84.8	84.8

* : NOGA incluse comme prédicteur dans le modèle de réponse, - : sans NOGA, – : hors AVS

Les nombres de ménages répondants, taux de conservation observés, taux prédits avec le modèle de prédiction utilisé pour la SE2018 et prédits avec le nouveau modèle selon la position dans la distribution des revenus AVS équivalents sont donnés dans la table 11. Les taux de conservation constatés et prédits en fonction du nombre de pièces du logement selon le RegBL sont donnés dans la table 12.

TABLE 11 Taux de conservation observé et estimé par décile de revenu AVS équivalent du ménage

Décile	éch. net	tx. csv.	anc. mod.	nv. mod.
1er (≤ 18 790 CHF)	18 387	82.5	84.4	82.6
2e (≤ 33 257 CHF)	20 849	81.8	83.9	81.8
3e (≤ 44 167 CHF)	22 007	83.2	84.4	83.2
4e (≤ 53 613 CHF)	22 150	84.7	84.9	84.7
5e (≤ 62 603 CHF)	21 583	86.0	85.3	86.0
6e (≤ 72 150 CHF)	21 191	85.8	85.5	85.8
7e (≤ 83 146 CHF)	21 027	86.1	85.8	86.1
8e (≤ 97 892 CHF)	21 079	86.7	86.3	86.7
9e (≤ 124 269 CHF)	21 230	86.6	86.7	86.6
10e (> 124 269 CHF)	21 521	86.4	86.8	86.4
Hors AVS, <19 ou >66	40 879	95.5	93.2	95.5
Hors AVS, 19-66 ans	14 634	86.8	87.8	86.8
Total	266 537	86.7	86.7	86.7

TABLE 12 Taux de conservation observé et estimé par nombre de pièces du logement selon RegBL

Nb. pièces RegBL	éch. net	tx. csv.	anc. mod.	nv. mod.
Hors RegBL ou manquant	617	64.8	86.5	64.8
1	7 808	83.9	86.2	83.9
2	24 016	86.0	88.2	86.0
3	62 046	86.0	87.6	86.0
4	83 692	87.5	86.7	87.5
5	52 422	88.4	86.4	88.4
6	23 431	87.2	85.5	87.2
7	7 668	84.6	84.4	84.6
8 ou plus	4 837	76.9	82.1	76.9
Total	266 537	86.7	86.7	86.7

4.3 Effet de l'introduction des nouvelles variables sur les poids des personnes avant calage

L'utilisation de modèles de réponse plus complexes, avec des prédicteurs supplémentaires, a pour effet d'augmenter la dispersion des probabilités de réponse estimées et le risque d'apparition de poids d'extrapolation élevés. Il en est de même pour les probabilités estimées d'observer des ménages identiques à ceux de STATPOP et pour les poids ménages avant calage.

Les probabilités de réponse estimées ne sont qu'assez légèrement modifiées par l'utilisation des nouvelles variables. En effet, pour près des trois quarts (72.7%) des unités de l'échantillon, l'écart relatif entre les probabilités estimées à l'aide de l'*ancien* modèle et les probabilités estimées avec le *nouveau* modèle est inférieur à 5%. Les distributions des probabilités de réponse estimées avec l'ancien et le nouveau modèle sont représentées dans la figure 5.

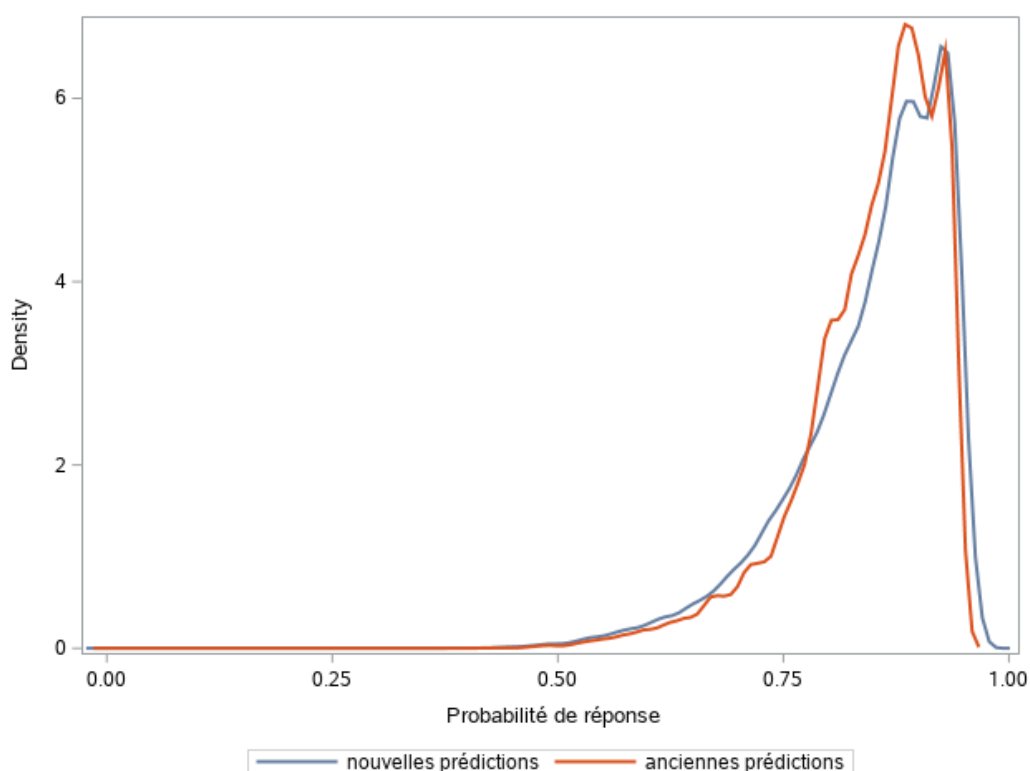


FIGURE 5 Distribution des probabilités de réponse estimées - échantillon brut.

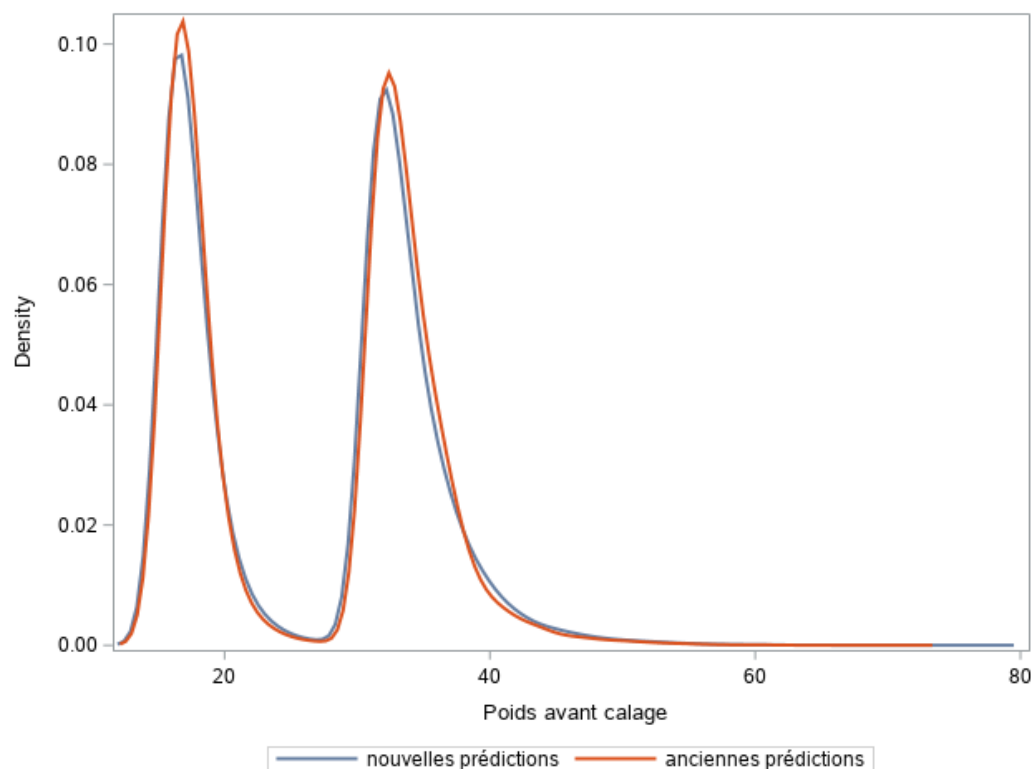
Seules les probabilités de réponse estimées pour les unités qui ont répondu à l'enquête interviennent effectivement dans les estimations par la suite. Les principaux quantiles de ces distributions, estimées selon l'ancien et le nouveau modèle, figurent dans la table 13. Cette table permet d'observer que la distribution des probabilités estimées avec le nouveau modèle, qui intègre les variables supplémentaires décrites ci-avant, est un peu plus étalée que celle des probabilités estimées avec l'ancien modèle.

Les distributions des poids avant calage ne sont quant à elles que très peu modifiées par l'utilisation des nouvelles variables pour prédire les probabilités de réponse des unités. Elles sont représentées dans la figure 6. Les deux modes apparents sur cette figure sont dus au fait que certains cantons ou communes densifient leur échantillon alors que d'autres ne le font pas. Les

TABLE 13 Quantiles des probabilités de réponse estimées

Quantile	anc. mod.	nv. mod.
0% (minimum)	0.386	0.351
1%	0.612	0.596
5%	0.714	0.699
10%	0.763	0.747
25% Q1	0.819	0.817
50% (médiane)	0.872	0.877
75% Q3	0.908	0.918
90%	0.930	0.938
95%	0.937	0.945
99%	0.945	0.958
100% (maximum)	0.952	0.978

unités répondantes dans les cantons ou communes qui densifient ont naturellement des poids d'extrapolation environ deux fois plus petits que les unités des cantons ou communes qui ne densifient pas, puisque les densifications correspondent approximativement à un doublement de l'échantillon [voir à ce sujet [Potterat et col., 2019](#)].

**FIGURE 6** Distribution des poids avant calage.

La figure 7 représente la distribution du quotient entre les poids avant calage calculés selon la nouvelle méthode et les poids avant calage calculés selon l'ancienne méthode, par classe d'âge. Elle permet de constater que ces quotients restent globalement autour de 1 pour toutes les classes d'âge. Une étude plus détaillée montre que les distributions des poids restent glo-

bablement inchangées à l'intérieur des classes d'âges et des autres caractéristiques démographiques relevées dans STATPOP.

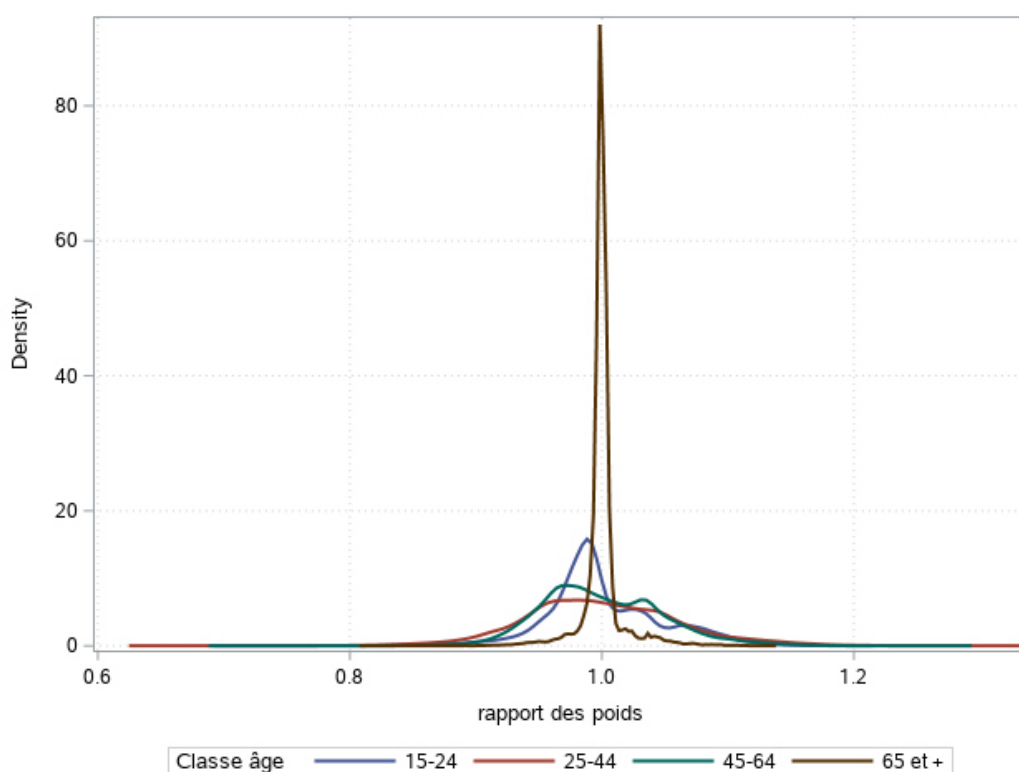


FIGURE 7 Quotient des poids avant calage, par classe d'âge.

Il apparaît par contre sur la figure 8, qui présente ces mêmes quotients groupés par quantiles de revenus AVS (les notations sont identiques à celles de la table 8), que les distributions des poids sont modifiées à l'intérieur de ces classes de population. Les quotients ne sont pas tous regroupés autour de 1, ce qui illustre que les poids des unités de certains quantiles sont plutôt augmentés par l'utilisation des nouvelles variables (essentiellement les quantiles de revenus plus faibles) alors que les poids des unités d'autres quantiles (les revenus les plus élevés) sont diminués.

4.4 Effet de l'introduction des nouvelles variables sur les poids des ménages avant calage

Des probabilités estimées que les ménages observés dans le relevé structurel aient la même composition que dans STATPOP ont été calculées en utilisant l'*ancien* modèle, décrit dans [Potterat et col. \[2019\]](#), et le *nouveau* modèle, qui utilise le nombre de pièces du logement ainsi que le quantile du revenu équivalent du ménage. Les distributions de ces probabilités estimées et des poids avant calage qui en découlent sont très similaires. Les principaux quantiles des probabilités estimées avec l'ancien et le nouveau modèle figurent dans la table 14. L'analyse de cette table montre que les probabilités estimées sont légèrement plus dispersées avec le nouveau modèle, plus complet que l'ancien. C'est également le cas des poids avant calage, qui sont également légèrement plus dispersés, avec un poids maximum qui passe à 125.2 alors qu'il était de 106.4 avec l'ancien modèle (voir table 17 dans la section 5.2).

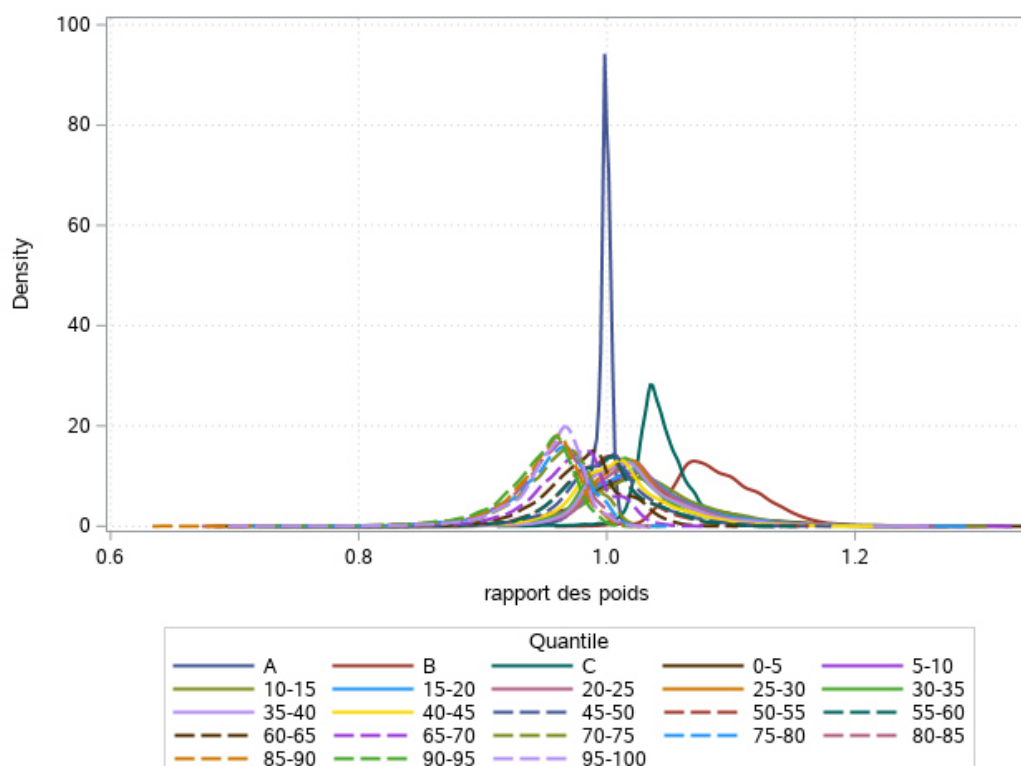


FIGURE 8 Quotient des poids avant calage, par quantile de revenu AVS.

TABLE 14 Quantiles des probabilités estimées de conservation des ménages

Quantile	anc. mod.	nv. mod.
0% (minimum)	0.233	0.141
1%	0.508	0.496
5%	0.707	0.683
10%	0.750	0.744
25% Q1	0.836	0.827
50% (médiane)	0.938	0.932
75% Q3	0.946	0.954
90%	0.957	0.965
95%	0.960	0.969
99%	0.964	0.972
100% (maximum)	0.969	0.979

Le rapport des poids avant calage obtenus avec la nouvelle et l'ancienne procédure est illustré dans la figure 9 en fonction de la taille du ménage et dans la figure 10 en fonction du quantile du revenu AVS équivalent du ménage.

Ces deux figures permettent de constater que les distributions de poids ne sont globalement pas modifiées à l'intérieur des sous-populations définies par la taille du ménage (les rapports de poids restent autour de 1). Par contre, les distributions sont légèrement modifiées en fonction du quantile du revenu AVS équivalent du ménage (les rapports de poids ne sont pas tous groupés autour de 1).

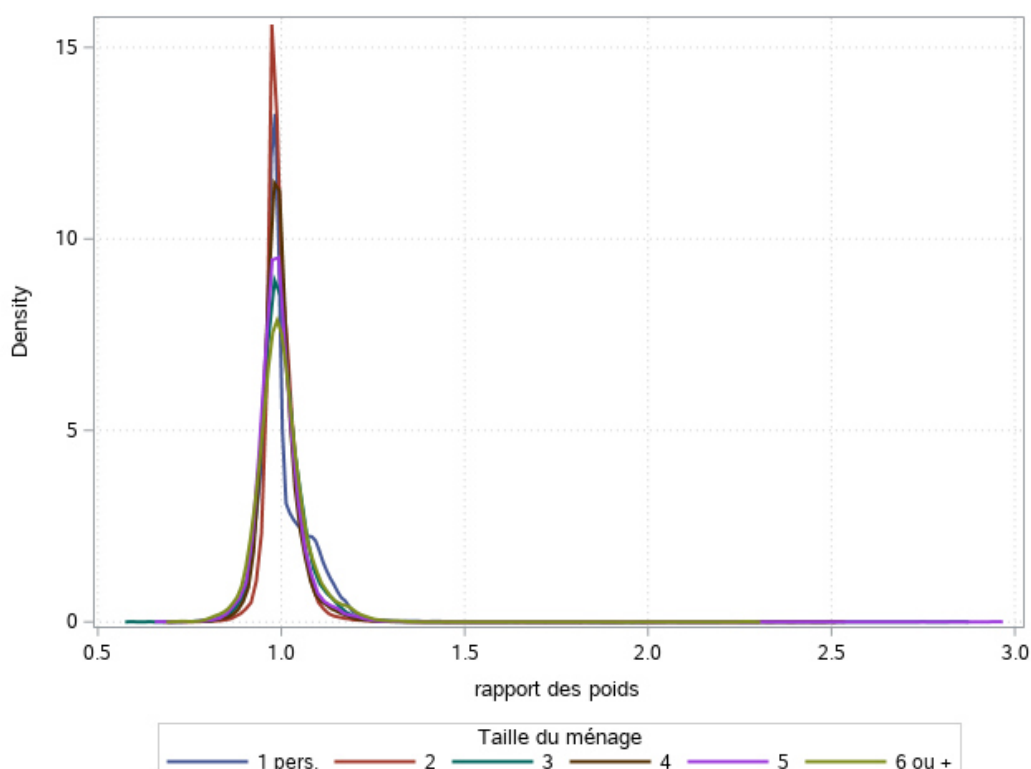


FIGURE 9 Rapport des poids avant calage, par taille de ménage.

5 Calage

Les variables de calage sont choisies soit pour assurer une cohérence entre différentes publications (ici STATPOP et relevé structurel), soit pour améliorer la précision d'estimations pour des variables qui sont liées aux variables de calage. L'échantillon du relevé structurel bénéficie depuis qu'il existe d'un calage étendu [voir [Potterat et col., 2019](#)] grâce à l'utilisation des données de STATPOP.

5.1 Nouvelles variables de calage

Pour améliorer la précision des estimations portant sur le loyer et le nombre de pièces, deux variables de calage supplémentaires ont été introduites pour la pondération des ménages. Il s'agit pour la première de la variable catégorielle créée à partir des revenus AVS équivalents et pour la seconde du croisement entre le canton et le nombre de pièces du logement selon le RegBL. La taille d'échantillon étant faible dans certains de ces croisements, les plus petits cantons ont été regroupés (Nidwald avec Obwald, Appenzell Rh.-Int. avec Appenzell Rh.-Ext. et Glaris avec Uri). De plus, les ménages hors RegBL et les ménages avec un logement de une pièce ne sont pas ventilés par canton. La ventilation des données STATPOP par nombre de pièces selon le RegBL et les extrapolations sur l'échantillon de la SE2018 avec les poids selon l'ancienne pondération et les nouveaux poids figurent dans la table [15](#).

La distribution des revenus AVS des personnes cibles répondantes reste très proche de sa distribution dans STATPOP sans qu'il n'ait été nécessaire de l'ajouter dans les variables de calage de la pondération des personnes cibles (voir table [18](#)).

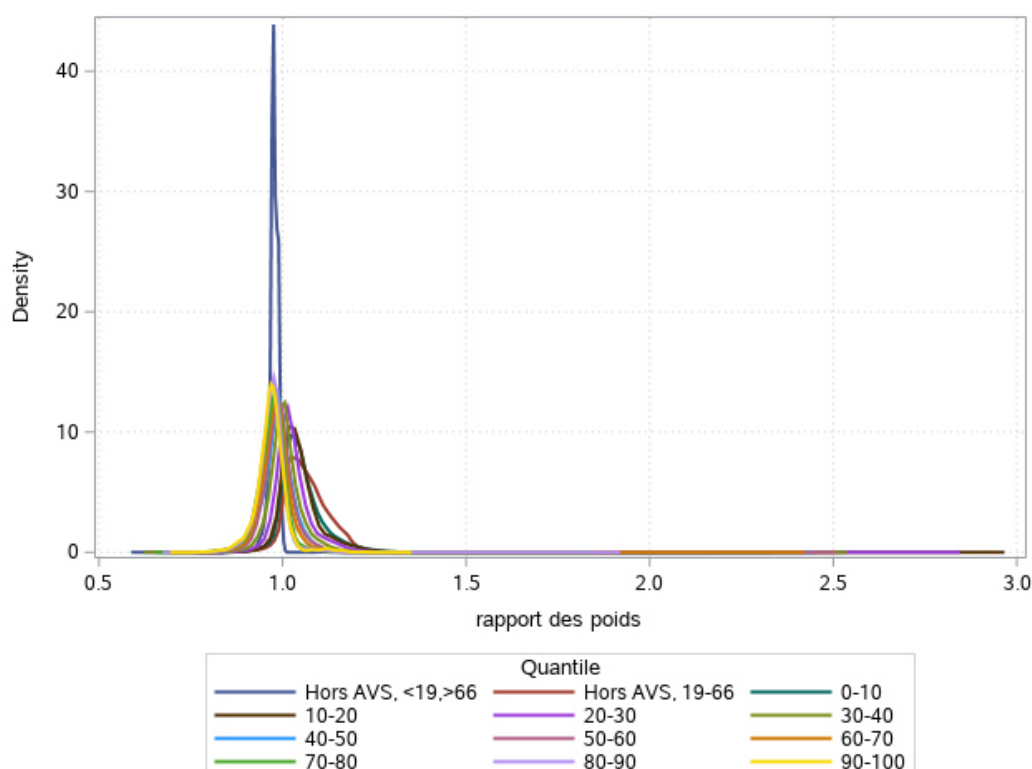


FIGURE 10 Rapport des poids avant calage, par quantile du revenu AVS équivalent du ménage.

TABLE 15 Répartition des ménages de STATPOP et répartition estimée dans la SE2018 ancienne et nouvelle pondération, selon le nombre de pièces enregistré dans le RegBL

Nb. pièces RegBL	STATPOP	anc. poids	nv. poids
Hors RegBL/manquant	19 245	9 629	19 245
1	180 387	166 125	180 387
2	479 536	467 604	479 536
3	986 998	973 249	986 998
4	1 081 189	1 102 119	1 081 189
5	619 575	635 924	619 575
6 ou plus	388 759	401 039	388 759
Total	3 755 689	3 755 689	3 755 689

Deux variables de calage supplémentaires pour les poids des personnes cibles ont été créées à partir de la commune et du canton de l'employeur de référence lorsqu'il était disponible. L'utilisation de ces variables a pour but d'améliorer la précision des estimations qui portent sur les déplacements domicile-travail. Les variables créées ont les modalités suivantes : pour la première variable,

- aucun employeur de référence trouvé,
- employeur dans la même commune que la commune de résidence,
- employeur dans une autre commune,

et pour la seconde variable,

- aucun employeur de référence trouvé,
- employeur dans le même canton que le canton de résidence,

- employeur dans un canton différent.

Il faut noter que les données des caisses de compensation permettent seulement d'identifier l'entreprise qui emploie un salarié. Dans le cas d'entreprises avec plusieurs établissements, il n'est pas possible de déterminer à quel établissement l'employé est rattaché. Seule une étude détaillée des précisions des différents estimateurs permettrait d'apprécier les gains apportés par l'utilisation de ces informations supplémentaires.

5.2 Effet de l'introduction des nouvelles variables sur les poids calés

L'utilisation de modèles de réponse ou de conservation plus complexes et l'ajout de variables de calage supplémentaires peuvent augmenter la dispersion des poids d'extrapolation. La présence de poids élevés peut conduire à une plus grande variabilité des estimateurs. Les modifications proposées ont très peu d'influence sur la distribution des poids d'extrapolation pour les personnes cibles de l'enquête. La table 16 donne des éléments sur la distribution des poids à différentes étapes de leur calcul (sélection de l'échantillon, non-réponse, calage) dans l'ancienne et dans la nouvelle pondération, ainsi que sur les distributions des probabilités de réponse prédites. Les distributions des poids calés sont représentées dans la figure 11.

TABLE 16 Distribution des poids d'extrapolation et des probabilités de réponse des personnes cibles

	Pds. tirage	après nr.		calé		pr. rép.	
		anc.	nv.	anc.	nv.	anc.	nv.
Moyenne	22.6	26.7	26.7	26.6	26.6	85.5	85.7
Minimum	14.4	15.1	14.8	12.1	11.3	38.6	35.1
1er percentile	14.4	15.3	15.3	14.4	14.2	61.2	59.6
5e percentile	14.4	15.7	15.5	15.1	14.9	71.4	69.9
1er décile	14.4	16.1	15.9	15.5	15.4	76.3	74.7
1er quartile	14.5	17.2	17.1	16.8	16.8	82.0	81.7
Médiane	28.9	31.1	31.0	30.5	30.2	87.2	87.7
3e quartile	29.1	33.5	33.3	33.7	33.6	90.8	91.8
9e décile	29.1	36.3	36.5	37.1	37.2	93.0	93.8
95e percentile	29.1	38.3	39.0	39.2	39.9	93.7	94.5
99e percentile	29.1	44.5	45.9	46.3	47.4	94.5	95.8
Maximum	29.1	70.3	76.6	83.9	90.8	95.2	97.8

Les poids des ménages sont quant à eux légèrement impactés par les modifications proposées avec notamment l'augmentation du poids maximum qui passe de 188.1 à 234.3 (voir table 17). L'augmentation de variance due à la plus grande dispersion des poids devrait toutefois rester limitée. Les distributions des poids calés des ménages sont représentées dans la figure 12.

Si les distributions globales des poids d'extrapolation des personnes et des ménages sont relativement peu modifiées par le changement de pondération, il faut tout de même noter un impact sensible dans certaines catégories de population. Concernant les personnes cibles, la nouvelle pondération conduit à augmenter le poids des personnes ou ménages en âge de cotiser et non retrouvées dans les données AVS ainsi que ceux des personnes dans le bas de la distribution des revenus AVS. En contrepartie, le poids des personnes et ménages dans le haut de la distribution des revenus AVS est diminué. Les effectifs de résidents permanents âgés de 15 ans ou plus dans STATPOP et estimés dans le relevé structurel avec l'ancienne et la nouvelle pondération figurent dans la table 18. Ils y sont répartis par quantile de revenu AVS. La dernière colonne «modif. poids» contient la différence relative moyenne entre l'ancienne et la nouvelle

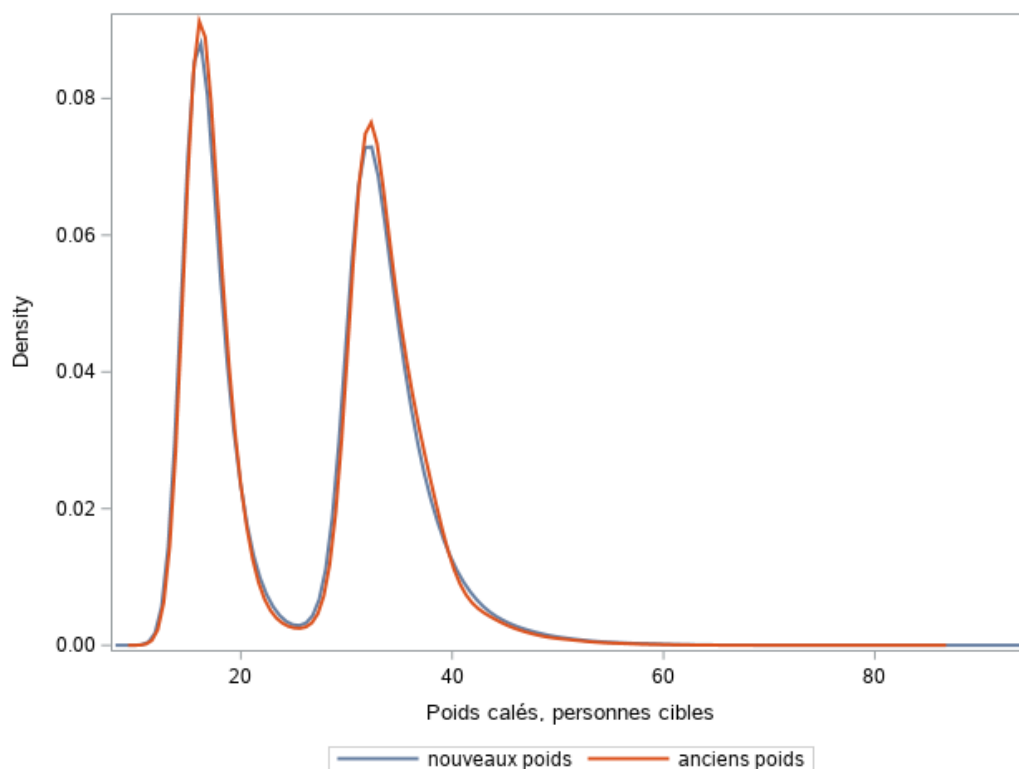


FIGURE 11 Distribution des poids calés des personnes cibles, ancienne et nouvelle pondération.

TABLE 17 Distribution des poids d'extrapolation des ménages

	Pds. tirage ($1/\pi_m$)	après nr. ($1/\pi_m p_m$)		après cons. ($1/\pi_m p_m s_m$)		calé (w_m)	
		anc.	nv.	anc.	nv.	anc.	nv.
Moyenne	12.5	14.7	14.7	16.4	16.4	16.2	16.2
Minimum	1.6	2.1	2.1	2.8	2.8	1.3	1.4
1er percentile	2.9	3.9	3.9	5.0	4.9	4.2	4.1
5e percentile	3.9	4.7	4.7	6.1	6.0	5.7	5.6
1er décile	4.9	6.0	6.0	8.1	7.9	7.3	7.1
1er quartile	7.3	8.3	8.2	9.1	9.0	8.9	8.9
Médiane	14.4	15.5	15.4	16.4	16.1	16.0	15.5
3e quartile	14.6	16.9	16.8	18.1	18.2	18.9	18.9
9e décile	29.0	32.9	32.5	34.9	34.3	32.3	31.2
95e percentile	29.1	34.6	34.3	36.9	37.1	35.9	35.8
99e percentile	29.1	39.4	39.7	43.4	44.6	43.6	46.0
Maximum	29.1	61.0	68.5	106.4	125.2	188.1	234.3

pondération dans chaque groupe de population. Il apparaît que les augmentations de poids les plus fortes concernent les personnes en âge de cotiser à l'AVS mais pour lesquelles aucun revenu AVS n'a été retrouvé. Les diminutions de poids sont, quant à elles, les plus fortes chez les personnes avec un revenu AVS dans le quart supérieur de la distribution. Les mêmes constats peuvent être réalisés à propos de la pondération des ménages, avec des différences plus marquées encore, dans la table 19.

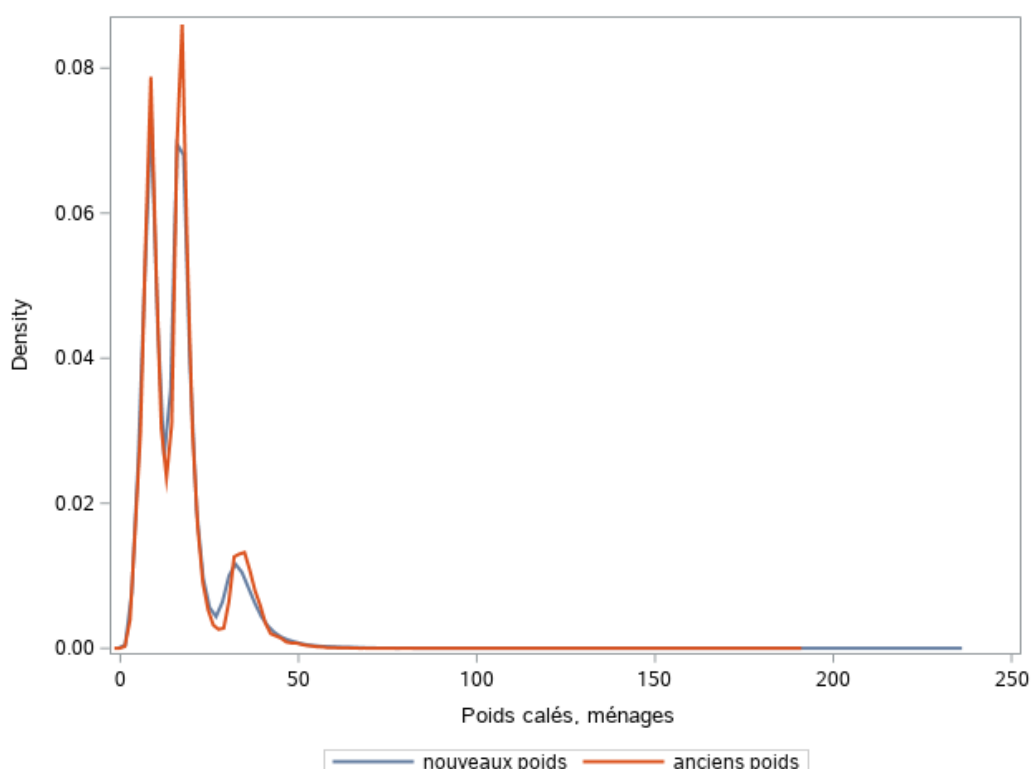


FIGURE 12 Distribution des poids calés des ménages, ancienne et nouvelle pondération.

6 Effet du changement de pondération sur quelques statistiques d'intérêt

L'effet principal de la modification de la pondération est de réajuster les poids en fonction des revenus AVS des répondants. La correction conduit à diminuer le poids des répondants ayant un revenu dans le haut de la distribution et à augmenter le poids de ceux ayant un revenu dans le bas de la distribution. L'échantillon ménage est calé sur la ventilation des ménages de STATPOP en fonction du nombre de pièces selon le RegBL. La répartition des ménages selon le nombre de pièces déclaré dans le relevé structurel, la proportion de logements occupés par leur propriétaire et le loyer mensuel net moyen des ménages locataires sont donnés dans la table 20 pour l'ancienne et la nouvelle pondération de la SE2018.

Le rééquilibrage de la distribution des revenus vers le bas peut être rapproché des changements constatés dans la table 20 : une augmentation du nombre de logements de petite taille au détriment des logements de grande taille, une diminution de la proportion de ménages propriétaires de leur logement et une légère diminution des loyers moyens.

Les modifications de pondération ont été appliquées sur les données de la SE2017 en utilisant les mêmes données des caisses de compensation de 2017. Les résultats obtenus sont donnés dans la table 21. Les modifications de résultats qui découlent du changement de pondération sont tout à fait comparables à ce qui est observé sur la SE2018 : elle conduit à une diminution du loyer mensuel net moyen estimé de l'ordre de quelques francs. Les évolutions entre 2017 et 2018 sont sensiblement les mêmes selon que l'ancienne ou la nouvelle pondération soit utilisée pour les deux enquêtes.

TABLE 18 Répartition de STATPOP et de l'échantillon de l'enquête SE2018 en fonction du quantile de revenu AVS de la personne cible

Quantile du revenu AVS (%)	STATPOP	anc. poids	nv. poids	modif. poids (%)
00 - 05 ($\leq$ 4 417 CHF)	189 506	183 609	187 329	2.03
05 - 10 ($\leq$ 9 545 CHF)	207 251	206 710	209 778	1.48
10 - 15 ($\leq$ 14 613 CHF)	219 839	215 788	220 935	2.39
15 - 20 ($\leq$ 20 525 CHF)	222 049	216 511	220 340	1.77
20 - 25 ($\leq$ 27 240 CHF)	226 564	223 026	226 533	1.57
25 - 30 ($\leq$ 33 772 CHF)	229 403	224 130	226 933	1.25
30 - 35 ($\leq$ 40 334 CHF)	228 603	226 024	227 885	0.82
35 - 40 ($\leq$ 46 800 CHF)	229 683	224 706	228 020	1.47
40 - 45 ($\leq$ 52 645 CHF)	228 366	226 773	227 368	0.26
45 - 50 ($\leq$ 58 059 CHF)	230 967	232 672	231 468	-0.52
50 - 55 ($\leq$ 63 200 CHF)	232 422	235 325	232 662	-1.13
55 - 60 ($\leq$ 68 480 CHF)	232 514	234 665	233 065	-0.68
60 - 65 ($\leq$ 73 991 CHF)	234 063	238 544	233 767	-2.00
65 - 70 ($\leq$ 80 000 CHF)	238 752	243 831	236 937	-2.83
70 - 75 ($\leq$ 87 055 CHF)	234 555	247 842	237 776	-4.06
75 - 80 ($\leq$ 95 922 CHF)	237 502	248 720	238 032	-4.30
80 - 85 ($\leq$ 108 000 CHF)	238 836	251 599	240 591	-4.38
85 - 90 ($\leq$ 126 414 CHF)	239 217	252 812	241 096	-4.63
90 - 95 ($\leq$ 162 145 CHF)	240 522	255 608	242 819	-5.00
95 - 100 ($>$ 162 145 CHF)	238 013	245 292	235 292	-4.08
Homme 19 - 66 ans hors AVS	355 202	321 303	352 549	9.72
Femme 19 - 65 ans hors AVS	540 488	516 759	543 676	5.21
Autre hors AVS	1 609 751	1 611 820	1 609 217	-0.16
Total	7 084 068	7 084 068	7 084 068	0.00

TABLE 19 Répartition de STATPOP et de l'échantillon de l'enquête SE2018 en fonction du quantile de revenu AVS équivalent du ménage

Décile du revenu AVS équivalent	STATPOP	anc. poids	nv. poids	modif. poids (%)
1er ($\leq$ 18 790 CHF)	275 756	247 254	275 756	11.53
2e ($\leq$ 33 257 CHF)	275 755	257 263	275 755	7.19
3e ($\leq$ 44 167 CHF)	275 755	265 222	275 755	3.97
4e ($\leq$ 53 613 CHF)	275 757	273 359	275 757	0.88
5e ($\leq$ 62 603 CHF)	275 753	279 899	275 753	-1.48
6e ($\leq$ 72 150 CHF)	275 915	282 181	275 915	-2.22
7e ($\leq$ 83 146 CHF)	275 596	289 366	275 596	-4.76
8e ($\leq$ 97 892 CHF)	275 755	293 439	275 755	-6.03
9e ($\leq$ 124 269 CHF)	275 755	296 032	275 755	-6.85
10e ($>$ 124 269 CHF)	275 755	292 955	275 755	-5.87
Hors AVS, <19 ou >66 ans	717 035	739 443	717 035	-3.03
Hors AVS, 19-66 ans	281 102	239 276	281 102	17.48
Total	3 755 689	3 755 689	3 755 689	0.00

TABLE 20 Nombre de logements, proportion de propriétaires occupants et loyers moyens des ménages locataires selon le nombre de pièces déclarées, estimations selon la SE2018

SE2018	Poids anc. méthode			Poids nv. méthode		
Nb. pièces	nb. logements	tx. propr.	loyer	nb. logements	tx. propr.	loyer
1	165 275	5.1	789	181 454	5.0	782
2	493 096	8.0	1 093	505 303	7.9	1 085
3	1 017 590	17.5	1 303	1 025 267	17.3	1 297
4	1 110 145	39.3	1 565	1 094 837	39.3	1 558
5	586 357	70.0	1 914	573 678	70.0	1 905
6 ou plus	383 226	84.3	2 422	375 150	84.4	2 411
Total	3 755 689	37.2	1 363	3 755 689	36.6	1 348

TABLE 21 Nombre de logements, proportion de propriétaires occupants et loyers moyens des ménages locataires selon le nombre de pièces déclarées, estimations selon la SE2017

SE2017	Poids anc. méthode			Poids nv. méthode		
Nb. pièces	nb. logements	tx. propr.	loyer	nb. logements	tx. propr.	loyer
1	154 334	5.0	752	170 106	4.8	746
2	477 680	7.7	1 061	491 449	7.7	1 053
3	993 987	17.6	1 270	1 000 120	17.5	1 264
4	1 095 660	39.6	1 519	1 083 556	39.5	1 512
5	599 067	70.3	1 855	583 791	70.3	1 845
6 ou plus	395 203	85.7	2 323	386 908	85.6	2 314
Total	3 715 931	38.0	1 329	3 715 931	37.4	1 314

7 Sélections répétées de personnes et de ménages dans l'échantillon du relevé structurel

Les échantillons des enquêtes de l'OFS, dont ceux du relevé structurel de la population suisse, sont sélectionnés de manière à répartir la charge d'enquête le mieux possible dans la population. Il arrive toutefois que certaines personnes soient sélectionnées pour cette enquête lors de deux ou, dans de très rares cas, plus de deux années consécutives. Les contrôles effectués n'ont jusqu'à présent jamais mis en évidence un défaut ou dysfonctionnement des programmes utilisés. Les raisons qui permettent d'expliquer individuellement les sélections répétées sont diverses et exposées dans la suite de ce document. L'existence de ces cas indésirables est certes surprenante, mais la complexité du problème ne doit pas être sous-estimée : il est aisé d'éviter ces cas en produisant des enquêtes *fausses*, c'est à dire en admettant des biais et erreurs d'estimation, mais il est extrêmement difficile voire impossible de les éviter totalement tout en assurant l'exactitude des procédures d'enquête.

7.1 Objectifs de la coordination des échantillons

Le système de coordination d'échantillons a été développé en 2009 pour répondre aux besoins en matière d'enquêtes auprès des entreprises. Il devait permettre de

- sélectionner des échantillons d'entreprises avec un libre choix des probabilités de sélection, celles-ci étant exactement respectées (les enquêtes auprès des entreprises utilisent usuellement des probabilités très inégales, égales à 1 pour les grandes unités, très petites pour les petites entreprises et très différentes selon les secteurs d'activité),

- permettre une coordination négative entre les enquêtes, c'est à dire obtenir le plus petit recouvrement possible entre les échantillons des différentes enquêtes,
- permettre de mettre à jour un échantillon de panel ou d'en réaliser une rotation partielle,
- être adapté à une population qui évolue dans le temps. La fréquence des mises à jour du cadre de sondage, qui n'était pas encore décidée lors du développement du système, a été fixée en 2014 à deux fois par an.

Répondre à ces exigences n'est pas trivial et bon nombre d'instituts nationaux de statistique utilisent des outils qui ne respectent pas exactement les probabilités de sélection spécifiées, ne sont pas adaptés à une population dynamique ou bien imposent de définir des strates communes pour toutes les enquêtes et d'utiliser un plan de sondage stratifié [voir à ce sujet [Nedyalkova et col., 2008](#)].

La solution développée pour l'OFS [voir [Qualité, 2009](#)] satisfait les critères énumérés mais ne répond pas à certains autres souhaits qui pourraient être formulés :

- elle ne permet pas de sélectionner des échantillons à un autre niveau (établissements, groupes, etc. ou dans le cas des enquêtes en population, personnes, ménages, bâtiment etc.) avec coordination,
- la sélection d'une unité ne peut pas dépendre de la sélection ou non d'autres unités (par exemple il est impossible d'imposer des règles qui empêcheraient la sélection simultanée de deux ménages dans un même bâtiment), ou plus exactement ces règles ne peuvent pas être respectées par le système de tirage coordonné,
- les tailles d'échantillons, totales ou dans des strates, des densifications, ne peuvent être imposées qu'approximativement (c'est en fait une conséquence du point précédent).

Ce système a été repris en 2010 pour constituer la base des procédures de sélection d'échantillons de personnes et de ménages dans le SRPH. Il a dû être adapté dès le départ et ensuite plusieurs fois pour répondre à des besoins nouvellement exprimés et assurer la pérennité des traitements informatiques. Ces adaptations ont eu des répercussions sur les performances du système en termes de coordination et de nombres de sélections répétées des différentes unités. L'évolution du système de coordination au fil du temps est décrite dans la section 7.2 et des éléments sur les sélections répétées de personnes et de ménages sont donnés respectivement dans les sections 7.3 et 7.4.

7.2 Historique du système de coordination

7.2.1 En 2010 - prototype

Sélection coordonnée de personnes. Les durées entre deux sélections d'une même personne dépendent uniquement des taux de sondage et des priorités de coordination choisies. Le système permet en théorie de quasiment garantir à chaque personne environ 15 ans entre deux sélections pour le relevé structurel dans un canton qui densifie son échantillon et 30 ans dans les autres cantons. Ce prototype n'a jamais été utilisé car il ne permet pas de contrôler le nombre de sélections au sein d'un même ménage.

Possibilités de sélection répétée indésirable :

- modification du numéro AVS dans le registre (concerne en particulier les numéros fictifs),
- numéros AVS qui disparaissent temporairement du SRPH sur un ou plusieurs trimestres, par exemple en cas de déménagement à l'étranger puis de retour ultérieur en Suisse (un suivi de ces cas aurait été envisageable, mais ne faisait pas partie de la conception du système de coordination d'enquête).

7.2.2 De 2010 à 2012

Sélection en deux étapes, tirage coordonné de personnes puis élimination des sélections multiples dans un ménage ou des sélections répétées dans une même enquête thématique. Pour obtenir une seule personne cible par ménage, les personnes vivant dans des grands ménages sont dans un premier temps sur-échantillonnées, puis dans un deuxième temps une seule personne par ménage est retenue (les probabilités de tirage sont calculées à chaque étape pour donner le résultat souhaité). Cette procédure conserve la meilleure coordination possible pour les personnes vivant seules mais dégrade la coordination pour les grands ménages. Dans le cas extrême, un ménage de 15 ou 16 personnes cibles dans un canton de densification est sélectionné chaque année. En 2012, il fallait de plus éliminer de l'échantillon les personnes dont le ménage avait déjà été sélectionné en 2011. Cette procédure serait devenue inefficace à partir de 2013 car elle aurait alors impliqué de sélectionner quasiment toutes les personnes vivant dans des grands ménages (à partir de 5 ou 6 personnes) avant de les éliminer de l'échantillon. Pour ces personnes, tout le bénéfice de la coordination aurait de ce fait été perdu.

Possibilités de sélection répétée indésirable (en plus des précédentes) :

- Les personnes vivant dans des très grands ménages (15 et 16 personnes cibles) dans les cantons de densification peuvent être sélectionnées deux années successives. La coordination est également fortement dégradée dans les grands ménages (8 à 14 personnes cibles, le temps minimum entre deux sélections passe de 15 à 10 à 4 ans selon la taille du ménage). Dans les cantons qui ne densifient pas, la coordination est légèrement dégradée pour les personnes dans des grands ménages (le temps minimum entre deux sélections passe de 30 à 20-25 ans selon la taille du ménage).

7.2.3 De 2013 à 2018

Passage à une sélection coordonnée de ménages et à des tirages en deux phases pour les échantillons de personnes. Les ménages sont sélectionnés avec coordination par rapport aux enquêtes passées, puis une personne cible est sélectionnée au hasard dans le ménage. Cette adaptation était nécessaire pour pouvoir continuer à assurer une coordination entre les ménages du relevé structurel et pour intégrer l'enquête sur le budget des ménages (EBM) et l'enquête sur les revenus et conditions de vie (SILC), ces deux dernières étant des enquêtes auprès des ménages. Il a fallu définir comment sont traitées les évolutions de composition des ménages. Il est toujours possible de choisir les règles de coordination pour les ménages. Par contre, il n'y a pas de procédure spécifique pour organiser une coordination au sein du ménage. Dans le cas où toutes les personnes du ménage ont la même probabilité de sélection, une solution basée sur la permutation des nombres aléatoires permanents pourrait être considérée. Cependant, ce n'est pas le cas de toutes les enquêtes. Ce n'est par exemple pas le cas de l'enquête suisse sur la population active (ESPA) dans laquelle la population étrangère est surreprésentée. Dans un ménage qui contient à la fois des personnes suisses et des personnes d'autres nationalités, les probabilités de sélection ne sont pas nécessairement égales. De plus, certains calculs de probabilité d'inclusion jointes, pertinents pour les «pooling» et enquêtes par panel seraient compliqués. L'implémentation ne serait pas forcément non plus viable et certaines questions restent à traiter (Le tirage est-il exact ? Comment traiter les ménages dont la composition change ? etc).

Possibilités de sélection répétée indésirable (en plus des précédentes) :

- les personnes qui changent de numéro de ménage perdent leur «protection» fournie par le système et donc peuvent être re-sélectionnées (cela concerne des personnes dont le

ménage change de composition. Dans ce cas, l'identifiant de ménage suit à chaque trimestre le plus grand groupe de personnes cibles),

- les ménages de 8 personnes et plus dans les cantons de densification et 15 personnes et plus dans les autres cantons peuvent être sélectionnés deux années consécutives. Il est possible que la même personne cible y soit retenue pour l'enquête. La durée minimum entre deux sélections du même ménage dépend de la taille du ménage et est bien inférieure à ce que permet une coordination au niveau des personnes.

7.2.4 Depuis fin 2018

Essentiellement la même procédure que depuis 2013 mais en coordonnant uniquement par ordre chronologique inverse : de l'enquête la plus récente à la plus ancienne. Le système initial permettait de choisir l'ordre des coordinations et au besoin de coordonner positivement. Tout en étant très simple, il nécessite un nombre d'opérations et génère une masse de données qui augmentent comme le carré du nombre d'enquêtes. Or le nombre d'enquêtes par trimestre a aussi tendance à augmenter. Il a fallu alléger le système pour garantir son fonctionnement à moyen et long terme. Cela implique qu'il n'est plus possible de coordonner prioritairement les échantillons du relevé structurel entre eux.

Possibilités de sélection répétée indésirable (en plus des précédentes) :

- Avec le nombre d'enquêtes, l'intensité de sélection a augmenté dans le temps. Dans des cantons de densification il y a maintenant des ménages de 5 ou 6 personnes qui sont nécessairement sélectionnés au moins une fois chaque année. Pour peu que leur probabilité de sélection cumulée sur un an soit proche de 1, il est alors possible qu'ils soient sélectionnés plusieurs années de suite au relevé structurel puis plusieurs années de suite à d'autres enquêtes (par opposition à alterner chaque année entre relevé structurel et une autre enquête si l'on pouvait choisir de coordonner en priorité les échantillons des relevés structurels entre eux). Comme il n'y a pas de coordination supplémentaire pour les personnes, la sélection peut alors concerner la même personne. Le nombre de ces ménages étant non négligeable il n'est pas anormal d'observer quelques cas de personnes sélectionnées 4 ou 5 fois à la suite.

7.3 Sélections répétées de personnes

Le nombre total de sélections de personnes lors de quatre trimestres consécutifs dans le SRPH figure dans la table 22 (colonne «Toutes enq.»). La ligne «2010» donne les sélections à partir du 4e trimestre 2010 - tirages en novembre 2010 sur des données SRPH au 30 septembre 2010 - et jusqu'au 3e trimestre 2011 inclus, ce qui correspond aux enquêtes dont le terrain se déroule en 2011, parmi lesquelles la SE2010. Cet effectif contient l'ensemble des sélections dans le SRPH, y compris les réserves et les sélections multiples dans les ménages qui sont filtrées avant enquête.

La colonne «SE (act.)» indique le nombre de sélections dans l'échantillon principal du relevé structurel, sans les réserves. Il s'agit de l'échantillon qui est effectivement activé. Dans la colonne «SRPH» figure le nombre de sélections effectivement effectuées pour le relevé structurel. Cela comprend donc les échantillons de réserves et, pour 2010, 2011 et 2012, les sélections filtrées par la suite de personnes dans un même ménage ou dans un ménage enquêté l'année précédente. Les six colonnes suivantes détaillent les sélections activées en fonction du nombre de fois que la personne cible a été sélectionnée dans un relevé structurel (échantillon principal, sans les réserves).

Les trois périodes de fonctionnement du système de coordination décrites ci-avant se distinguent nettement dans la table 22. De 2010 à 2012 la coordination était réalisée au niveau des personnes cibles et permettait très bien de les protéger de sélections répétées.

Entre 2013 et 2017, la perte de coordination pour les personnes qui changent de ménage et les possibilités de sélection répétées dans les grands ménages expliquent l'augmentation du nombre de personnes sélectionnées 2 fois ou plus. Le nombre de sélections répétées augmente essentiellement par paliers lors des années à partir desquelles les ménages d'une taille donnée peuvent être à nouveau sélectionnés dans un canton de densification ou bien dans un canton sans densification.

Après 2018, le système ne permet plus autant de limiter les sélections répétées dans le relevé structurel. Cela se combine avec un nombre annuel de sélections, devenu conséquent aussi dans les autres enquêtes, pour expliquer en partie l'augmentation des sélections répétées. De plus, une fraction importante des ménages a déjà été sollicitée au moins une fois, ce qui rend inévitable des sélections multiples hormis dans les ménages d'une ou de deux personnes. Dans les cantons qui densifient leurs échantillons, ces derniers vont, de manière naturelle, pouvoir être re-sélectionnés dans les années qui viennent. En effet, un ménage stable d'une personne dans un canton de densification devrait être sélectionné environ tous les 15 ans, or le système de coordination est utilisé depuis 2010. À moyen terme, une fraction relativement faible de l'échantillon sera sollicitée pour la première fois, essentiellement alimentée par les arrivées dans la population sélectionnable.

TABLE 22 Nombre de sélections totales de personnes dans l'année, pour le relevé structurel et détail en fonction du nombre de sélections dans le relevé structurel

	Toutes enq.	SE (act.)	SRPH	1ère sél.	2e sél.	3e sél.	4e sél.	5e	6e
2010	551 550	365 883	433 116	365 883	0	0	0	0	0
2011	510 719	326 308	377 748	326 230	78	0	0	0	0
2012	590 033	329 185	435 693	328 948	237	0	0	0	0
2013	688 111	325 915	358 754	322 184	3 730	1	0	0	0
2014	817 199	333 054	366 528	313 337	19 700	17	0	0	0
2015	712 263	314 109	345 731	269 056	44 887	166	0	0	0
2016	683 191	322 431	322 431	269 822	51 974	632	3	0	0
2017	818 518	314 464	314 464	260 666	52 481	1 309	8	0	0
2018	870 868	314 155	314 155	236 733	72 985	4 379	57	1	0
2019	865 631	315 444	315 444	219 396	88 018	7 875	153	2	0
2020	994 933	353 208	353 208	238 035	103 171	11 548	447	7	0
2021	923 521	338 289	338 289	216 275	106 534	14 673	783	22	2
2022	933 523	343 666	343 666	211 801	112 460	18 114	1 247	43	1
Total	9 960 060	4 296 111	4 619 227						

7.4 Sélections répétées de ménages

Les sélections de ménages dans l'échantillon du relevé structurel sont détaillées dans la table 23. Le nombre total de ménages distincts sélectionnés dans l'échantillon actif figure dans la colonne «Total». La définition du ménage avant 2013 dans cette table ne correspond pas exac-

tement à celle qui avait été utilisée pour les tirages². En effet, un identifiant de ménages suivi dans le temps a été créé en 2013 seulement. Avant cela la variable utilisée pour les tirages était essentiellement basée sur les identifiants de bâtiments et de logements et couplée à une indicatrice de «validité du ménage».

La colonne «Nv. sél.» donne la proportion de ménages qui sont sélectionnés pour la première fois dans le relevé structurel. Elle est calculée en observant la composition du ménage lors de sa sélection et en examinant si l'une au moins des personnes de ce ménage a déjà été sélectionnée dans un relevé structurel antérieur. Il faut noter que le système de coordination, quant à lui vise, à étaler les sélections des identifiants de ménage qui peuvent désigner un groupe de personnes qui varie dans le temps. Il est donc possible qu'un ménage apparaisse comme déjà sélectionné dans cette table alors qu'il n'est pas considéré comme tel dans le système de coordination, par exemple s'il s'agit d'un nouveau ménage constitué de personnes qui appartenaient auparavant à d'autres ménages.

La colonne «Sél. an préc.» indique le nombre de ménages (groupes de personnes - numéros AVS) dont au moins un membre a été sélectionné pour le relevé structurel l'année précédente. La colonne «Selon SRPH» donne le nombre d'identifiants de ménages qui avaient été sélectionnés au précédent relevé structurel. C'est la seule colonne de cette table qui concerne le concept de ménage effectivement utilisé par le système de coordination d'échantillons depuis 2013. Ces deux colonnes font clairement apparaître :

- le tirage de la SE2011, sans protection particulière pour les ménages touchés en 2010,
- la SE2012 avec un soin particulier pour éviter de sélectionner dans les mêmes «ménages-groupes de personnes» qu'en 2011,
- la période 2013-2017 avec une coordination prioritaire entre les relevés structurels, où quelques grands «ménages-identifiants» sont re-sélectionnés et environ 1500 «ménages-groupes de personnes» dont l'identifiant a changé le sont également,
- le fait qu'après 2018, la coordination entre relevés structurels ne vient plus en première priorité mais passe après les autres enquêtes de l'année, ce qui, selon les taux de sondage des différentes enquêtes intercalées, rajoute entre 1900 et 5600 sélections répétées.

Enfin, dans la colonne «Anc.» figure l'ancienneté moyenne en années depuis la dernière sélection dans le relevé structurel d'un membre du ménage-groupe de personnes lorsque le ménage a déjà été touché par l'enquête. Par exemple, parmi les 226 869 ménages sélectionnés dans l'échantillon de la SE2022 qui avaient déjà été sélectionnés pour le relevé structurel par le passé (cf. table 24), 5678 l'avaient été pour la dernière fois un an auparavant, 25 913 deux ans auparavant, etc. La durée moyenne entre la précédente sélection et la SE2022 est de 6 ans. Cette ancienneté moyenne cache une répartition assez étalée, comme le montre la table 24 qui détaille les «ménages-groupes de personnes» sélectionnés pour la SE2022 en fonction de la dernière fois qu'un de ses membres a été sélectionné pour le relevé structurel, et les sélections répétées de personnes en fonction de leur dernière sélection.

Sur les 1613 personnes sélectionnées à la fois dans la SE2021 et la SE2022, 904 étaient dans le même ménage lors de ces deux sélections. Dans tous ces cas, la probabilité de sélection cumulée du ménage sur douze mois approche ou dépasse 1. Il s'agit essentiellement (716/904) de ménages de 4, 5 ou 6 personnes.

La répartition donnée dans la table 24 dépend non seulement des performances du système de coordination, mais aussi du mouvement naturel de la population, de l'évolution de la composition des ménages, des taux de sélection et plans de sondage des relevés structurels et des

2. Ce qui explique l'écart d'une unité dans le relevé structurel 2011 entre la table 22 et la table 23.

TABLE 23 Sélections de ménages dans le relevé structurel : nombre de ménages activés, proportion de ménages sélectionnés pour la première fois dans le relevé structurel, nombre de ménages sélectionnés l'année passée et ancienneté moyenne de sélection pour les ménages sélectionnés à nouveau

	Total	Nv. sél. (en %)	Sél. an préc.	selon SRPH	Anc.
2010	365 883	100.0	-	-	-
2011	326 307	92.0	26 006	-	1.0
2012	329 185	91.4	146	-	2.0
2013	325 915	95.7	1 727	-	2.2
2014	333 054	82.4	1 871	139	2.9
2015	314 109	65.8	1 645	125	3.9
2016	322 431	63.2	1 732	64	4.5
2017	314 464	62.7	1 676	47	5.1
2018	314 155	49.9	3 236	1 904	4.6
2019	315 444	41.8	5 261	4 015	5.1
2020	353 208	39.6	4 066	2 665	5.3
2021	338 289	35.9	6 946	5 641	5.6
2022	343 666	34.0	5 678	4 445	6.0

TABLE 24 Sélections de ménages dans la SE2022 en fonction de la SE de dernière sélection (d'un de ses membres), et sélections de personnes dans la SE2022 en fonction de leur dernière sélection passée

	Ménages	Personnes
Jamais	116 797	211 801
SE2010	12 333	11 405
SE2011	12 274	10 462
SE2012	13 035	10 525
SE2013	14 804	10 906
SE2014	21 733	14 953
SE2015	15 686	10 219
SE2016	21 756	12 447
SE2017	29 304	15 660
SE2018	28 584	14 465
SE2019	25 769	10 681
SE2020	25 913	8 529
SE2021	5 678	1 613
Total	343 666	343 666

autres enquêtes sélectionnées dans le SRPH. Il faut noter qu'il y a eu environ 4 300 000 sélections de ménages dans les relevés structurels depuis leur introduction, soit plus que le nombre total de ménages dans la population.

La proportion de ménages qui ont été déjà sélectionnés entre 2010 et 2022 par l'OFS pour ses enquêtes ou dans les échantillons livrés à des tiers est représenté dans la figure 13. Les proportions y sont détaillées par taille de ménage. Tout comme la table 24, les courbes de la figure 13 résultent non seulement de la taille et du nombre des échantillons sélectionnés, de la composition des ménages, de la performance du système de coordination, mais aussi de l'évolution

naturelle de la population des ménages. L'aplatissement des courbes observé dans la période récente ne traduit pas une baisse des tailles d'échantillon mais résulte plutôt de l'évolution de la structure des ménages combinée à la sélection répétée de certains ménages dans les échantillons.

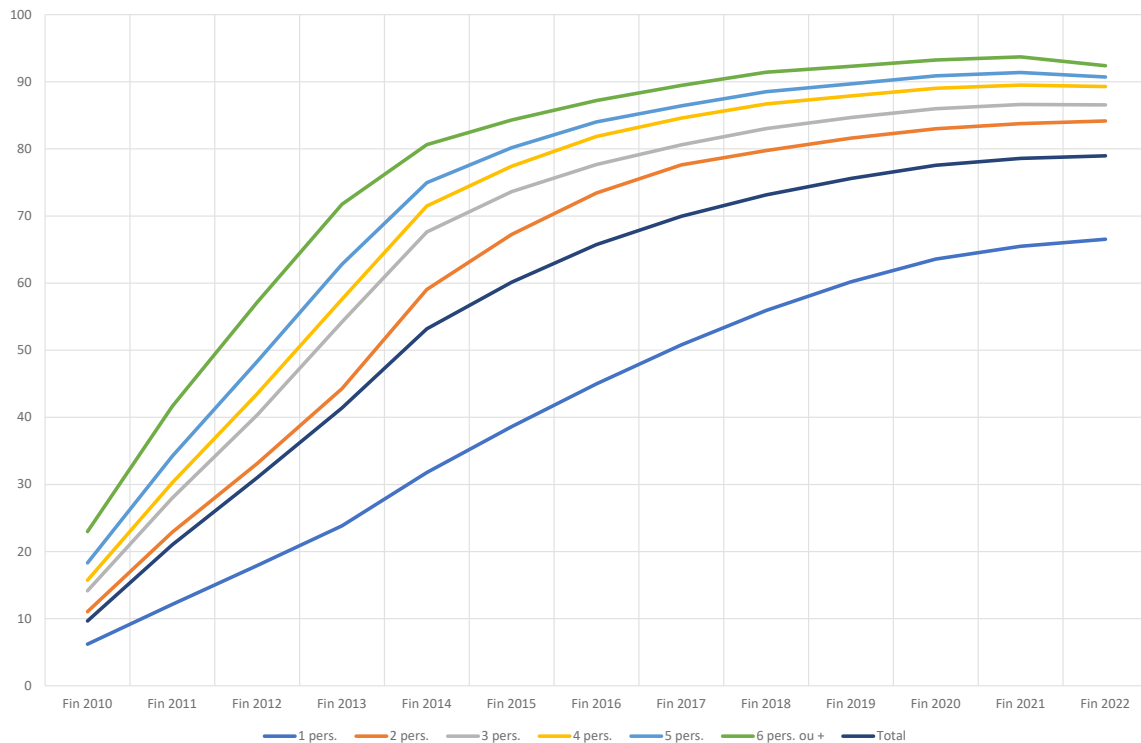


FIGURE 13 Proportion des ménages sélectionnés au moins une fois dans une enquête du système de coordination, en fonction de la taille du ménage.

La répartition de la population selon le nombre de sélections individuelles dans l'ensemble des échantillons tirés par l'OFS dans le cadre de sondage SRPH est illustrée dans la figure 14. Il faut noter qu'un peu plus de 20% de la population n'est pas concernée par ces enquêtes, essentiellement car ce sont des personnes de moins de 15 ans ou bien des personnes qui vivent dans des ménages collectifs.

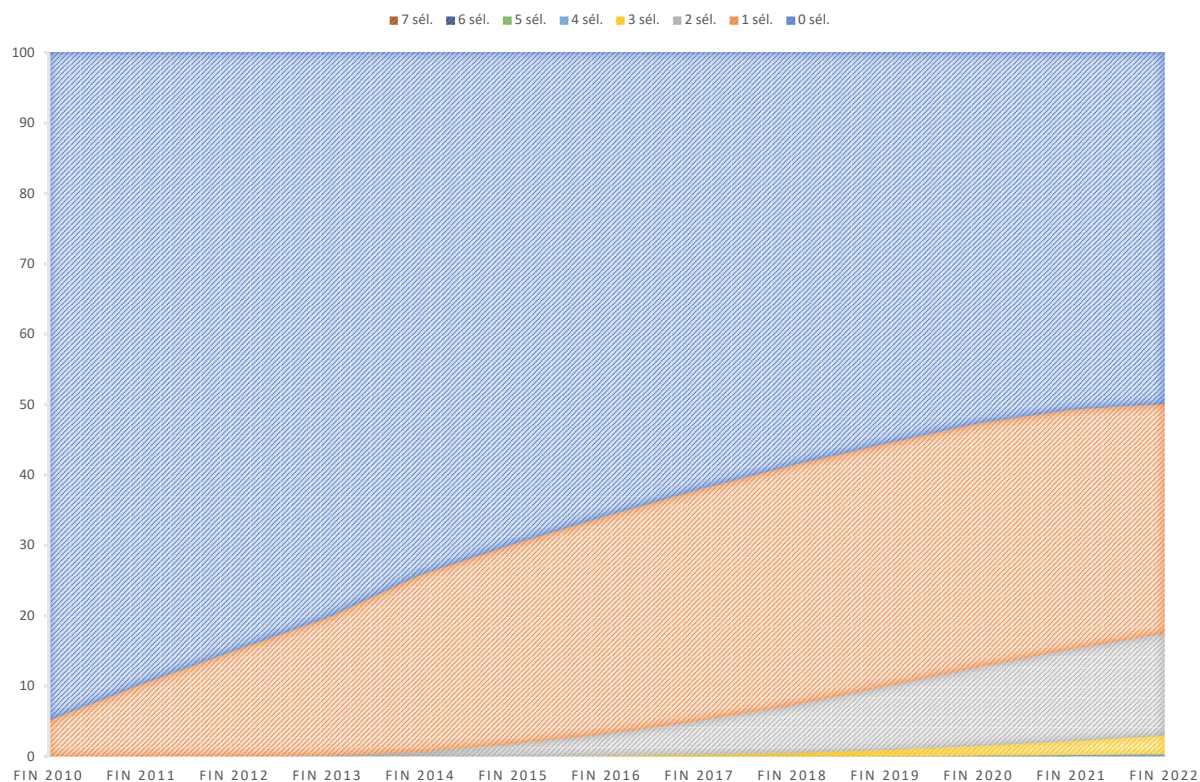


FIGURE 14 Répartition de la population en fonction du nombre de sélections dans des enquêtes du système de coordination.

8 Développements en cours et conclusions

L'OFS évalue en permanence les méthodes et pratiques appliquées pour la sélection des échantillons et leur utilisation. Ces méthodes et pratiques sont révisées périodiquement. La mise à disposition, la préparation et l'utilisation de données nouvelles, par exemple des sources administratives, permettent de réduire la charge d'enquête et d'améliorer la précision de leurs résultats. L'intégration des données sur les revenus détenues par les caisses de compensation constitue une avancée notable pour les statistiques portant sur la population et les ménages. Le niveau de revenu est en effet lié à une grande part des sujets d'intérêt couverts par les enquêtes de l'OFS. Cette variable vient utilement compléter les informations démographiques traditionnellement utilisées, telles que le sexe, l'âge, la nationalité ou le lieu de résidence, qui n'ont pas toujours un très grand pouvoir explicatif sur les thèmes d'enquête.

La pondération de l'échantillon du relevé structurel est désormais adaptée pour reproduire fidèlement la distribution des revenus connus. Cela permet d'obtenir des estimations qui sont plus fidèles à la réalité de la population et des ménages vivant en Suisse. Pour autant, cette nouvelle variable est imparfaite : elle ne couvre pas les rentes des retraités ou les revenus de la fortune. Le patrimoine des personnes et des ménages ne s'y reflète pas non plus. D'autres sources de données, comme les données fiscales par exemple, seront à l'avenir envisagées pour compléter les données des caisses de compensation.

Outre les grandes avancées permises par l'utilisation de données nouvelles, des optimisations dans le procédé de collecte, le traitement des données et les calculs de pondération sont effectuées ou envisagées.

Lors de la SE2018, une modification majeure de la procédure de contact des personnes sélectionnées a été implémentée par l'OFS. Les personnes âgées de 15 à 64 ans sont désormais invitées à répondre prioritairement par voie électronique. Un formulaire papier ne leur est envoyé que sur demande ou lors du deuxième rappel³. Il en a résulté une forte évolution de la proportion de questionnaires remplis électroniquement. Celle-ci est en effet passée de 31.1% en 2017 (35.7% parmi les 15-64 ans) à 72.1% en 2018 (87.7% parmi les 15-64 ans). Dans le même temps, le taux de réponse restait relativement stable : 85.4% à la SE2017 contre 84.8% à la SE2018. Une étude complète de l'éventuel impact de ce changement sur les résultats reste à effectuer, pour autant qu'elle soit possible. Le processus de préparation statistique des données connaît lui aussi des évolutions ponctuelles qui ne sont pas documentées ici.

Une autre évolution qui peut être envisagée concerne le calage des poids et la définition de la population de référence. Jusqu'à présent, celle-ci est constituée d'une part des résidents permanents âgés de 15 ans ou plus en ce qui concerne les résultats sur les «personnes cibles» qui sont directement enquêtées et d'autre part de l'ensemble des résidents permanents en ce qui concerne les résultats produits à partir des informations recueillies sur la partie ménage du questionnaire. Il y a chaque année quelques centaines de ménages (précisément 735 ménages en 2018) qui ne contiennent aucun résident permanent âgé de 15 ans ou plus, mais qui contiennent des mineurs de moins de 15 ans ayant la qualité de résident permanent. Ces ménages font partie de la population de référence puisqu'ils contiennent des résidents permanents mais ne peuvent pas être observés à l'enquête. Les retirer de la population de référence permettrait de rapprocher l'échantillon de la population qu'il est censé représenter et en même temps d'appliquer une procédure de calage intégré [voir à ce sujet [Le Guennec et Sautory, 2002](#)], qui

3. En 2018, les personnes dont la langue de contact est le Romanche recevaient également un questionnaire papier car la traduction de l'interface de réponse électronique dans cette langue n'était pas encore disponible.

garantirait une forme de cohérence entre les poids ménages et les poids des personnes cibles. Le point le plus visible est que les poids des personnes vivant seules seraient alors égaux au poids de leur ménage, ce qui n'est pas le cas actuellement.

Un calcul de la précision des estimations permettrait de réellement évaluer l'impact du changement de pondération sur les résultats présentés dans la section 6. Il serait également utile pour caractériser l'amélioration de la qualité permise par l'utilisation de nouvelles données dans la pondération. Cela pourra faire l'objet de travaux ultérieurs.

Le système de coordination d'échantillons a été adapté plusieurs fois aux besoins des enquêtes auprès des personnes et des ménages et pour faire face à l'augmentation du nombre d'échantillons à sélectionner. Avec l'intensité de sélection actuelle, un ménage qui comporte 3 ou 4 personnes âgées de 15 ans ou plus peut s'attendre à être sélectionné dans un échantillon de l'OFS environ tous les deux ans. La fréquence de sélection pour les ménages plus grands est plus élevée encore. Naturellement, une bonne part des sélections concernent le relevé structurel qui est, de loin, la plus grande enquête réalisée par l'OFS (voir table 22). Le système de coordination permet de répartir au mieux les sélections dans le temps, mais il ne permet pas d'éviter les sélections répétées car cela entraînerait des biais de couverture. Seules des optimisations de son implémentation informatique sont actuellement envisagées.

Annexes

A Approximation des probabilités d'inclusion des ménages

Les π_m définis par l'équation 1 dans la section 2 sont une approximation des probabilités de sélection réelles des ménages. En effet, ils ne sont strictement égaux aux probabilités de sélection que pour les ménages de STATPOP qui ne contiennent aucun nouveau membre qui aurait par ailleurs été sélectionnable dans un autre ménage selon le cadre de sondage SRPH utilisé pour sélectionner l'échantillon. Ce dernier, pour pouvoir réaliser la collecte au tout début de l'année, est construit sur la base des données des registres de populations datés du 30 septembre précédent l'enquête. Il y a donc naturellement une différence entre les ménages du cadre de sondage et ceux de STATPOP.

L'écart entre les π_m calculés et les probabilités de sélection exactes reste toutefois minime et ne concerne qu'une très faible part de l'échantillon. Dans le cas d'un ménage m de STATPOP composé de deux personnes k et ℓ qui étaient sélectionnables et dans deux ménages différents du SRPH, la valeur obtenue est

$$\pi_m = \pi_k + \pi_\ell,$$

alors que la vraie probabilité de sélection du ménage vaut $\pi_k + \pi_\ell - \pi_k\pi_\ell$. Une application numérique avec une probabilité de sélection individuelle de 3.5%, ce qui est proche du taux de sondage du relevé structurel dans un canton sans densification, donne pour le premier calcul $\pi_m=7\%$ et, pour le second, 6.88%. Les équations 2 et 3 reposent sur des approximations similaires.

En 2018, ces approximations de calcul touchent 1602 ménages parmi les 231 145 ménages de l'échantillon net, soit 0.7% de l'échantillon net. Pour ces 1602 ménages, l'écart relatif moyen entre les poids avant calage calculés selon la méthode simplifiée et ceux qui seraient calculés avec la méthode exacte est de 2%, avec un écart maximum de 7%. L'erreur d'approximation a donc un impact négligeable sur les résultats.

B Taux de réponse des personnes et ménages selon les caractéristiques STATPOP

TABLE 25 Échantillon selon le sexe

Sexe	Éch. brut	Éch. net	Tx. réponse (%)
Homme	155 314	129 712	83.5
Femme	158 841	136 825	86.1
Total	314 155	266 537	84.8

TABLE 26 Échantillon selon l'état-civil

État-civil	Éch. brut	Éch. net	Tx. réponse (%)
Célibataire	107 900	86 259	79.9
Marié(e)	157 842	139 284	88.2
Autre	48 413	40 994	84.7

TABLE 27 Échantillon de personnes cibles selon le canton de résidence

Canton	Éch. brut	Éch. net	Tx. réponse (%)
Argovie	19 175	16 321	85.1
Appenzell Rh.-Int.	438	360	82.2
Appenzell Rh.-Ext.	1 608	1 336	83.1
Berne	33 716	29 053	86.2
Bâle campagne	8 304	7 130	85.9
Bâle ville	5 557	4 368	78.6
Fribourg	8 950	7 871	87.9
Genève	26 252	21 146	80.6
Glaris	1 127	906	80.4
Grisons	5 950	5 072	85.2
Jura	2 171	1 888	87.0
Lucerne	22 426	19 167	85.5
Neuchâtel	10 339	8 841	85.5
Nidwald	1 319	1 169	88.6
Obwald	1 137	1 003	88.2
Saint-Gall	14 760	12 354	83.7
Schaffouse	2 430	2 034	83.7
Soleure	7 987	6 723	84.2
Schwytz	4 553	3 945	86.6
Thurgovie	7 898	6 713	85.0
Tessin	21 065	18 060	85.7
Uri	1 026	913	89.0
Vaud	45 685	39 056	85.5
Valais	9 945	8 578	86.3
Zoug	7 358	6 403	87.0
Zürich	42 979	36 127	84.1

TABLE 28 Échantillon selon la taille du ménage

Taille du ménage	Éch. brut	Éch. net	Tx. réponse (%)
1 pers.	59 671	49 361	82.7
2 pers.	105 465	93 107	88.3
3 pers.	55 604	46 486	83.6
4 pers.	58 763	50 229	85.5
5 pers.	22 977	18 699	81.4
6 pers. et +	11 675	8 655	74.1
Total	314 155	266 537	84.8

Bibliographie

- DEVILLE, J.-C. & SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- HAZIZA, D., CHEN, S. & GAO, Y. (2022). Targeting key survey variables at the nonresponse treatment stage. *Journal of Survey Statistics and Methodology*, 10, 25-49.
- LE GUENNEC, J. & SAUTORY, O. (2002). CALMAR2 : une nouvelle version de la macro CALMAR de redressement d'échantillon par calage. *Actes des Journées de Méthodologie*, pp.33-38, Insee, Paris.
- NEDYALKOVA, D., PEA, J. & TILLÉ, Y. (2008). Sampling Procedures for Coordinating Stratified Samples : Methods Based on Microstrata. *International Statistical Review*, 76, 368-386.
- OFS (2008). *NOGA 2008 Nomenclature générale des activités économiques. Introduction*. Office fédéral de la statistique, Neuchâtel. ISBN 978-3-303-00358-9.
- POTTERAT, J., QUALITÉ, L. & ASSOULIN, D. (2019). *Strukturerhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018*. Methodenberichte der Sektion Statistische Methoden, Bundesamt für Statistik, Neuchâtel.
- QUALITÉ, L. (2009). *Unequal probability sampling and repeated surveys*. Thèse de Doctorat, Université de Neuchâtel, Suisse.

Rapports de méthodes de la section méthodes statistiques de l'OFS
Methodenberichte der Sektion Statistische Methoden des BFS
Methodology reports published by the FSO's Statistical Methods Section

Qualité, L. (2023). Relevé structurel de la population suisse 2018. Révision de la pondération et sélections répétées. Numéro OFS : 338-0087

Potterat, J., Assoulin, D. (2023). Forschung und Entwicklung 2021. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0086

Chevalier, C. (2023). Méthodes de correction des effets calendrier en usage à l'OFS en 2023. Construction, choix et application des régresseurs mis à disposition. Numéro OFS : 338-0085

Mariétan, R. (2022). Comportement en matière de voyages 2020. Révision de la pondération. Numéro OFS : 338-0084

Massiani, A. (2021). Enquête suisse sur la population active 2015-2020 : Révision de la pondération et de l'estimation de la variance. Numéro OFS : 338-0083

Chevalier, C., Qualité, L. (2021). Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Numéro OFS : 338-0082

Qualité, L., Potterat, J. (2020). Enquête suisse sur la structure des salaires 2018. Révision du plan de sondage. Numéro OFS : 338-0081

Potterat, J., Qualité, L., Assoulin, D. (2019). Strukturhebung der eidgenössischen Volkszählung. Stichprobenplan und -ziehung, Gewichtung, Schätzverfahren und Pooling 2010-2018. BFS-Nummer : 338-0080

Potterat, J., Assoulin, D., Nicoletti, J.-M. (2019). Beschäftigungsstatistik BESTA. Revision 2015 : Stichprobenrahmen, Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0079

Massiani, A. (2017). Estimation de la couverture du nouveau recensement en Suisse en 2012. Numéro OFS : 338-0078

Nedyalkova, D., Assoulin, D. (2017). Construction of full-time equivalents for the Swiss structural business statistics. FSO number : 338-0077

Ferster, M. (2016). Energieverbrauchsstatistik EVS2014 - Stichprobe, Hochrechnung und Vergleichbarkeit mit der EVS2013. BFS-Nummer : 338-0076

Panchard, C. (2014). Enquête sur la participation aux activités culturelles 2008. Plan de sondage et estimations. BFS-Nummer : 338-0073

Assoulin, D. (2014). Wertschöpfungsstatistik. Revision 2009 : Statistische Datenaufbereitung und Hochrechnung. BFS-Nummer : 338-0072

Wilhelm, M. (2014). Echantillonnage boule de neige. La méthode de sondage déterminé par les répondants. Numéro OFS : 338-0071

Assoulin, D. (2013). Wertschöpfungsstatistik. Revision 2009 : Stichprobenrahmen und Stichprobenplan. BFS-Nummer : 338-0070

Ferster, M. (2013). EVS I - Energieverbrauchsstatistik 2002 bis 2007 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0069

Assoulin, D. (2013). Zusatzerhebung für die landwirtschaftliche Betriebszählung 2010 : Stichprobenplan und Hochrechnung. BFS-Nummer : 338-0068

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. FSO number : 338-0067-05

Andrade, B., Salamin P.-A. (2012). Enquête sur la situation sociale et économique des étudiant-e-s des hautes écoles suisses 2009. Cadre de sondage, plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0066

Potterat, J., Panchard, C., Kilchmann, D. (2012). Umweltschutzausgaben der Unternehmen 2009 (UWSA2009). Stichprobenplan, Einsetzungen, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0065

Potterat, J. (2012). Benutzung der Umsteigeschlüssel NOGA 2002-2008. BFS-Nummer : 338-0064

Potterat, J. (2011). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2009 (KNBB09). Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0063

Kilchmann, D., Potterat, J., Genoud, S. (2011). Gütertransporterhebung 2008. Stichprobenplan, Datenaufbereitung, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0062

Graf, E. (2010). Enquête suisse sur la santé 2007. Plan d'échantillonnage, pondérations et analyses pondérées des données. Numéro OFS : 338-0061

Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. FSO number : 338-0060

Graf, E. (2010). Étude empirique de l'attrition du Panel Suisse de Ménages : vers une caractérisation du profil du non-répondant. Numéro OFS : 338-0059

Graf, E. (2009). Weightings of the Swiss Household Panel : SHP_I wave 9, SHP_II wave 4, SHP_I et SHP_II combined. FSO number : 338-0058

Graf, E. (2009). Pondérations du Panel Suisse de Ménages : PSM_I vague 9, PSM_II vague 4, PSM_I et PSM_II combinés. Numéro OFS : 338-0057-05

Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro OFS : 338-0056

Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro OFS : 338-0055

Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro OFS : 338-0054

Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium". Numéro OFS : 338-0053

Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro OFS : 338-0052

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

Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. BFS-Nummer : 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 OFS : 338-0049

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

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

Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer : 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 OFS : 338-0045

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

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

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

- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 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 OFS : 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS : 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. FSO number : 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer : 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer : 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer : 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer : 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS : 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS : 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer : 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer : 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number : 338-0027
- Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindizes und deren Varianzschätzer. BFS-Nummer : 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS : 338-0025
- Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS : 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 OFS : 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS : 338-0022
- Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. BFS-Nummer : 338-0021
- Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. BFS-Nummer : 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 OFS : 338-0019

Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. BFS-Nummer : 338-0018

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

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

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

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

Renaud, A. (2002). Programme international pour le suivi des acquis des élèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro OFS : 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 OFS : 338-0012

Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. BFS-Nummer : 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 OFS : 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 OFS : 338-0009

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

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

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

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

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

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

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

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

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

Programme des publications de l'OFS

En tant que service statistique central de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public. Il utilise plusieurs moyens et canaux pour diffuser ses informations statistiques par thème.

Les domaines statistiques

- 00 Bases statistiques et généralités
- 01 Population
- 02 Espace et environnement
- 03 Travail et rémunération
- 04 Économie nationale
- 05 Prix
- 06 Industrie et services
- 07 Agriculture et sylviculture
- 08 Énergie
- 09 Construction et logement
- 10 Tourisme
- 11 Mobilité et transports
- 12 Monnaie, banques, assurances
- 13 Sécurité sociale
- 14 Santé
- 15 Éducation et science
- 16 Culture, médias, 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, disparités régionales et internationales

Les principales publications générales

L'Annuaire statistique de la Suisse



L'Annuaire statistique de la Suisse de l'OFS constitue depuis 1891 l'ouvrage de référence de la statistique suisse. Il englobe les principaux résultats statistiques concernant la population, la société, l'État, l'économie et l'environnement de la Suisse.

Le Mémento statistique de la Suisse



Le mémento statistique résume de manière concise et attrayante les principaux chiffres de l'année. Cette publication gratuite de 52 pages au format A6/5 est disponible en cinq langues (français, allemand, italien, romanche et anglais).

Le site Internet de l'OFS: www.statistique.ch

Le portail «Statistique suisse» est un outil moderne et attrayant vous permettant d'accéder aux informations statistiques actuelles. Nous attirons ci-après votre attention sur les offres les plus prisées.

La banque de données des publications pour des informations détaillées

Presque tous les documents publiés par l'OFS sont disponibles gratuitement sous forme électronique sur le portail Statistique suisse (www.statistique.ch). Pour obtenir des publications imprimées, vous pouvez passer commande par téléphone (+41 58 463 60 60) ou par e-mail (order@bfs.admin.ch).
www.statistique.ch → Trouver des statistiques → Catalogues et banques de données → Publications

Vous souhaitez être parmi les premiers informés?



Abonnez-vous à un Newsmail et vous recevrez par e-mail des informations sur les résultats les plus récents et les activités actuelles concernant le thème de votre choix.
www.news-stat.admin.ch

STAT-TAB: la banque de données statistiques interactive



La banque de données statistiques interactive vous permet d'accéder simplement aux résultats statistiques dont vous avez besoin et de les télécharger dans différents formats.
www.stattab.bfs.admin.ch

Statatlas Suisse: la banque de données régionale avec ses cartes interactives



L'atlas statistique de la Suisse, qui compte plus de 4500 cartes, est un outil moderne donnant une vue d'ensemble des thématiques régionales traitées en Suisse dans les différents domaines de la statistique publique.
www.statatlas-suisse.admin.ch

Pour plus d'informations

Centre d'information statistique

+41 58 463 60 11, info@bfs.admin.ch

En 2018, l'Office Fédéral de la Statistique (OFS) a amélioré la pondération de l'échantillon du relevé structurel grâce à l'utilisation de nouvelles données administratives dans son calcul. Les caisses de compensation de l'assurance-vieillesse et survivants (AVS) livrent à l'OFS des informations sur les personnes qui leur sont affiliées. Ces données sont désormais utilisées à différentes étapes du calcul des poids d'extrapolation du relevé structurel. En outre des données provenant du registre des bâtiments et logements (RegBL) et du registre des entreprises et établissements (REE) sont également utilisées. Jusqu'en 2017, la pondération du relevé structurel utilisait uniquement des données issues de la statistique de la population et des ménages STATPOP.

Ce rapport présente les traitements introduits pour le relevé structurel 2018, les analyses qui ont conduit à cette adaptation et une étude de l'impact de la modification de la pondération sur les résultats. À la fin de ce document, la question des sélections répétées de personnes et de ménages dans l'échantillon du relevé structurel est également abordée.

En ligne

www.statistique.ch

Imprimés

www.statistique.ch
Office fédéral de la statistique
CH-2010 Neuchâtel
order@bfs.admin.ch
tél. +41 58 463 60 60

Numéro OFS

338-0087

ISBN

978-3-303-00728-0

Les informations publiées ici contribuent à mesurer la réalisation des objectifs de développement durable (ODD).



Système d'indicateurs MONET 2030

www.statistique.ch → Trouver des statistiques → Développement durable → Système d'indicateurs MONET 2030

**La statistique
compte pour vous.**

www.la-statistique-compte.ch