

2018

00

Bases
statistiques
et généralités

Neuchâtel 2021

Enquête suisse sur la structure des salaires 2018

Révision de la pondération

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

Enquête suisse sur la structure des salaires 2018

Révision de la pondération

Rédaction	Clément Chevalier et Lionel Qualité, OFS
Contenu	Section Méthodes statistiques, OFS
Éditeur	Office fédéral de la statistique (OFS)

Neuchâtel 2021

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Clément Chevalier, tél. 058 481 00 19
clement.chevalier@bfs.admin.ch

Rédaction: Clément Chevalier et 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: section DIAM, Prepress/Print
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. 058 463 60 60
Impression réalisée en Suisse

Copyright: OFS, Neuchâtel 2021
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.

Numéro OFS: 338-0082

ISBN: 978-3-303-00682-5

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Variables, plan de sondage et unités	7
2.1 Variables importantes	7
2.1.1 Données auxiliaires connues sur tout le cadre de sondage	7
2.1.2 Salaire brut mensuel standardisé et temps de travail standardisé	8
2.1.3 Nombre d'employés et nombre de salaires	8
2.2 Plan de sondage, poids de sondage	9
2.3 Processus de préparation statistique des données	9
2.4 Unités inéligibles, unités inactives	10
3 Non-réponse et calage	11
3.1 Vue d'ensemble de la pondération - un plan à trois degrés	11
3.2 Modèle de réponse	12
3.2.1 Généralités	12
3.2.2 Modèles de réponse	13
3.2.3 Modèle de non-élimination des données	14
3.3 Construction des variables de calage	15
3.3.1 Situation en l'absence de calage et critères de choix des variables de calage	15
3.3.2 Choix et construction des variables de calage	16
3.4 Méthode de calage	18
4 Calcul de coefficients de variations et hypothèses	20
4.1 Hypothèses effectuées pour construire le plan de sondage	20
4.2 Calcul des coefficients de variation	21
5 Résultats	22
5.1 Nombre d'unités et nombre d'individus	22
5.1.1 Secteur privé	23
5.1.2 Administration publique	23
5.2 Coefficients de variation d'estimateurs de la médiane et du total	23
6 Conclusions	32
Bibliographie	33
Annexes	33
A Unités et nombre d'employés dans l'échantillon net du secteur privé par croisement NOGA x GR	33
B Coefficient de variation de l'estimateur du salaire médian dans le secteur privé par croisement NOGA x GR	36
Rapports de méthodes de la section méthodes statistiques de l'OFS	37

Préambule

L'enquête suisse sur la structure des salaires (ESS) est réalisée tous les deux ans par l'Office fédéral de la statistique (OFS) auprès d'un échantillon d'entreprises et d'administrations publiques. Elle relève des informations sur les salaires versés, la nature des emplois correspondants et les caractéristiques des employés qui perçoivent ces salaires. La section des méthodes statistiques (METH) de l'OFS a récemment révisé le plan de sondage de l'ESS2018. La suite naturelle de ce travail est une révision de la procédure de pondération. En particulier, tout comme lors de la révision du plan de sondage, les améliorations proviennent de l'utilisation de données administratives disponibles sur les revenus des personnes salariées en Suisse. Cette nouvelle source de données permet d'améliorer la qualité des résultats produits. Les auteurs remercient la section Salaires et conditions de travail (LOHN) de l'OFS, qui est responsable de l'ESS et en particulier F. Santi, D. Droz, D. Froidevaux, J. Ignaczewska, A. Rouvinez pour le soutien qu'ils ont apporté à ce travail. Les auteurs remercient également D. Assoulin, D. Kilchmann et J.-P. Renfer de la section METH pour leur relecture rigoureuse de ce rapport.

Résumé

La méthode de pondération de l'ESS a fait l'objet d'une révision approfondie pour l'enquête de l'année 2018. Cette révision s'inscrit dans la continuité des travaux effectués sur le plan de sondage présentés dans [Qualité & Potterat \(2020\)](#). L'objectif est d'utiliser les données administratives disponibles sur les revenus des personnes salariées pour améliorer la qualité des résultats. Les poids des unités primaires (entreprises privées ou publiques, administrations publiques) répondant à l'enquête sont obtenus à l'aide du plan de sondage, mais aussi via une modélisation du mécanisme de réponse à l'enquête et un calage des poids. Pour ces deux dernières opérations, les données administratives sur les revenus des personnes ont permis, comme pour le plan de sondage, d'améliorer la précision de l'enquête. Les poids des unités primaires et les poids des salaires des personnes employées dans ces unités sont liés par une relation simple. Au final, un peu plus de 2 millions de salaires relevés permettent, une fois pondérés, de produire les résultats de l'enquête. Les analyses montrent qu'un calage des poids utilisant les données administratives sur les revenus des personnes améliore le coefficient de variation de l'estimateur du salaire médian d'environ 10% à 20% selon le domaine considéré (NOGA, Grande Région).

1 Introduction

L'enquête Suisse sur la structure des salaires (ESS) est effectuée tous les deux ans par l'Office Fédéral de la Statistique (OFS), sous la coordination de la section LOHN. Elle vise entre autres à estimer des salaires médians sur l'ensemble de la Suisse ou bien dans des domaines tels que des cantons, grandes régions ou catégories d'activité. Les estimations se basent sur un échantillon aléatoire de salaires relevés qui, une fois standardisés pour tenir compte de la durée du travail, servent de variable principale à l'enquête. Le présent rapport se focalise sur la méthodologie de pondération des données de l'échantillon net de l'ESS2018, c'est-à-dire des salaires relevés dans les entités qui sont interrogées et qui ont répondu à l'enquête. Les étapes principales de ce processus sont notamment le traitement de la non-réponse à l'enquête et le calage sur des variables auxiliaires importantes. Tout comme pour le plan de sondage, une procédure séparée est utilisée pour les administrations publiques (environ 250 unités dans l'échantillon net) et pour le secteur privé et entreprises publiques (environ 37'100 unités dans l'échantillon net). La construction du plan de sondage des administrations publiques (communes, districts, cantons et Confédération) et du secteur privé et entreprises publiques est décrite dans [Qualité & Potterat \(2020\)](#).

2 Variables, plan de sondage et unités

Certaines opérations sont effectuées en amont du processus de pondération mais ont une influence sur celui-ci et doivent être brièvement présentées pour mieux comprendre les décisions prises pour la pondération.

2.1 Variables importantes

2.1.1 Données auxiliaires connues sur tout le cadre de sondage

Au cours de l'enquête, des unités primaires sont interrogées. Il s'agit d'entreprises privées, publiques ou d'administrations publiques, sélectionnées aléatoirement dans un cadre de sondage décrit dans [Qualité & Potterat \(2020\)](#). Ces entités seront simplement nommées "unités" ou "unités primaires" dans ce rapport. Le nombre total d'employés d'une unité donnée est une variable connue sur tout le cadre de sondage. Cette variable est basée sur un extrait du registre des entreprises et des établissements (REE) et est ici appelée `betot_cadre`.

La catégorie d'activité d'une unité primaire est une variable importante connue sur tout le cadre de sondage. Celle-ci est obtenue à partir de la nomenclature générale des activités (NOGA). Des regroupements de NOGA en 39 catégories d'activité économique (OFS39) sont effectués et sont documentés dans le rapport de méthodes sur le plan de sondage ([Qualité & Potterat, 2020](#)). Par la suite, le terme NOGA, même s'il est utilisé seul, fait toujours référence à ce *regroupement de NOGA*, OFS39, en 39 classes différentes.

Le présent rapport de méthodes présente deux pondérations bien distinctes pour les unités des administrations publiques (communes, districts, cantons et Confédération) et du secteur privé et entreprises publiques. Par abus de langage, cette seconde pondération est référencée comme la pondération "du secteur privé", même si celle-ci inclut des entreprises publiques. Concrètement, toutes les administrations publiques font partie de la NOGA 84, et sont traitées dans la pondération du secteur public. La NOGA 84 inclut également des entreprises publiques qui sont traitées dans la pondération du secteur privé.

La grande région, GR, d'une unité primaire est utilisée aussi bien pour l'élaboration du plan de sondage que pour la pondération. Elle intervient aussi dans la définition d'objectifs de précision. Par exemple, les estimations de salaires médians à l'intérieur de chaque grande région doivent atteindre des précisions données, voir le rapport de [Qualité & Potterat \(2020\)](#) pour plus de détails. Les grandes régions en Suisse sont des regroupements d'un ou plusieurs cantons. La liste des grandes régions, avec entre parenthèse les cantons qui la composent, est la suivante : Région lémanique (GE, VS, VD), Espace Mittelland (BE, FR, JU, NE, SO), Suisse du Nord-Ouest (AG, BL, BS), Zurich (ZH), Suisse orientale (AI, AR, GL, GR, SG, SH, TG), Suisse centrale (LU, NW, OW, SZ, UR, ZG), et Tessin (TI).

En plus de variables relatives à une unité primaire telles que `betot_cadre` ou GR, l'OFS dispose de données sur les individus, mises à disposition par les caisses de compensation AVS. En particulier, un revenu annuel est disponible pour les personnes ayant cotisé à l'AVS au cours de l'année 2018. Les revenus AVS ne remplacent pas les salaires bruts relevés au cours de l'enquête. Cependant leurs corrélation avec ces derniers en font une information précieuse qui sera notamment utilisée pour la modélisation des probabilités de réponse et le calage, voir sections [3.2](#) et [3.3](#).

Les unités interrogées dans l'ESS livrent à l'OFS plusieurs variables portant sur l'unité primaire elle-même ainsi que plusieurs variables portant sur les individus qui y travaillent. Certaines variables importantes relevées lors de l'enquête sont décrites dans les sections suivantes.

2.1.2 Salaire brut mensuel standardisé et temps de travail standardisé

À partir d'informations relevées dans l'enquête telles que le salaire brut et le nombre d'heures travaillées des employés des unités répondantes, l'OFS calcule un salaire standardisé, MBLS (monatlicher Bruttolohn standardisiert), qui correspond à un salaire brut mensuel pour 40 heures de travail hebdomadaire. Le salaire standardisé est associé à une variable qui décrit le temps de travail hebdomadaire de la personne qui touche un salaire : IBGRS (individueller Beschäftigungsgrad standardisiert). Cette variable représente le taux d'activité ramené à 40 heures hebdomadaires. La variable IBGRS est importante dans la pondération. En effet, c'est au final, d'une certaine façon, chaque heure de travail effectuée en Suisse qui recevra un poids. Ce point est détaillé dans la section [3.1](#).

2.1.3 Nombre d'employés et nombre de salaires

Lorsqu'une unité livre des données, elle déclare le nombre d'employés qui travaillent dans l'unité. Ce nombre peut être légèrement différent de la variable `betot_cadre` décrite précédemment. Le nombre total d'employés que l'entreprise déclare avoir est stocké à l'OFS dans la variable FB01. L'unité livre des données sur la totalité ou une partie des employés. Le nombre d'emplois sur lesquels des données sont livrées et validées est stocké dans la variable `anzlohn`. Pour chaque unité, `anzlohn` est nécessairement inférieur ou égal à FB01. Selon les directives fournies aux entreprises, l'OFS fait l'hypothèse que les `anzlohn` employés dont les données sont livrées parmi les FB01 de l'unité primaire ont été tirés au hasard via un plan de sondage simple de taille fixe. Les conséquences sur la pondération sont discutées dans la section [3.1](#). Pour beaucoup d'unités, `anzlohn` est strictement inférieur à FB01. Cela est dû aux deux facteurs suivants :

1. Les unités ne sont pas toujours tenues de livrer l'intégralité de leurs données. La règle est basée sur une variable de classe de taille, TA3, qui dépend de `betot_cadre`. La variable TA3 prend les valeurs 1 si `betot_cadre` est compris entre 3 et 19, 2 si `betot_cadre` est

compris entre 20 et 49 et 3 si `betot_cadre` est supérieur ou égal à 50. Lors de l'enquête, il est demandé aux unités de la classe de taille $TA3 = 1$ de livrer les données sur la totalité de leurs employés. Pour les classes de taille $TA3 = 2$ et 3 il est demandé de livrer les données sur au moins la moitié ($TA3 = 2$) et le tiers ($TA3 = 3$) des employés. La variable $TA3$ sera citée à de nombreuses reprises dans ce rapport.

2. Les données livrées ne sont pas toujours intégralement validées. Certaines données sont susceptibles d'être éliminées des calculs, en partie ou en totalité. Cela peut se produire si certaines variables importantes sont manquantes ou si certaines données sont jugées comme étant peu plausibles. Cette étape de filtrage et de recherche de valeurs aberrantes ne fait pas partie du processus de pondération à proprement parler, mais elle a une influence sur la valeur de `anzlohn` (le nombre de personnes travaillant dans l'unité primaire dont les données livrées sont validées) et donc sur les résultats de la pondération. Cette étape de préparation des données fait partie du processus de préparation statistique des données (PPSD, "SDAP" en allemand) et est brièvement résumée dans la section 2.3.

2.2 Plan de sondage, poids de sondage

Le plan de sondage est présenté en détail dans le rapport de méthodes [Qualité & Potterat \(2020\)](#). Il s'agit d'un plan de Poisson où les unités sont sélectionnées indépendamment avec une probabilité de tirage qui dépend de la strate de l'unité. Dans [Qualité & Potterat \(2020\)](#), on trouve la construction des strates ainsi que la procédure pour choisir les taux de sondage dans chaque strate. Une fois le plan construit, une unité i dans le cadre de sondage a une probabilité π_i d'être tirée dans l'échantillon brut. Cela confère naturellement à l'unité i un poids de sondage $1/\pi_i$ déjà calculé au moment d'effectuer la pondération. Le poids des unités primaires obtenu lors de la pondération ne sera en général pas égal à ce poids de sondage. Ces poids seront adaptés pour prendre en compte deux principes essentiels en théorie des sondage. Premièrement, il conviendra de calculer un poids avant calage pour les unités, qui sera calculé comme l'inverse de leur probabilité d'inclusion dans l'échantillon net. Une unité fait partie de l'échantillon net si elle est tirée dans l'échantillon brut, si elle répond et si les données livrées sont validées lors du PPSD pour au moins un employé. Deuxièmement, les poids avant calage seront adaptés par un calage sur des variables auxiliaires bien choisies. Le calage sur des totaux de variables auxiliaires vise à améliorer la précisions des estimations et assurer la cohérence des résultats avec des totaux connus par ailleurs.

2.3 Processus de préparation statistique des données

Lorsque les unités primaires répondent à l'enquête en livrant des données à l'OFS, la section LOHN leur applique un processus de préparation statistique des données (PPSD). Il s'agit d'une étape importante pour la pondération car elle est susceptible dans certains cas d'éliminer tout ou partie des données livrées par une unité primaire. Ainsi une unité qui répond à l'enquête peut se voir exclue de l'échantillon net avec une probabilité non nulle qu'il convient de prendre en compte pour la pondération. Ce point est détaillé davantage dans la section 3.2. Le PPSD consiste essentiellement en les étapes suivantes.

- Un calcul de certaines variables importantes telles que MBLs (voir : section 2.1.2) à partir des données livrées telles que le salaire brut ou le nombre d'heures de travail.
- Des filtres qui provoquent une alerte si les données ne satisfont pas un certain nombre de tests de plausibilité. Les alertes sont vérifiées manuellement et une décision (ne pas conserver les données, ou bien continuer le processus) est prise par la section. Par exemple, des

comparaisons sont faites entre le nombre d'employés que l'unité déclare avoir (FB01) et les valeurs dans les registres.

- Une succession de filtres nommés filtres "Behalten", décrits succinctement ci-dessous.
 - Certains filtres sont très simples et consistent seulement en la suppression d'observations avec des valeurs manquantes sur des variables cruciales (par exemple : si le salaire est manquant).
 - Un filtre calcule plusieurs indicateurs basés sur la différence entre le salaire communiqué par l'entreprise et celui dont l'OFS dispose via les données des caisses de compensation AVS (voir section 2.1.1). En cas d'écart trop important, les indicateurs dépassent un seuil et la donnée concernée n'est pas retenue.
 - Le mécanisme Behalten_Cook consiste à effectuer une régression linéaire des salaires standardisés relevés à l'aide de prédicteurs bien choisis et à analyser les résidus de cette régression, ou plus précisément la distance de Cook de chaque observation. Celle-ci est une fonction du résidu et du levier de l'observation. Les observations avec une distance de Cook supérieure à un seuil choisi sont exclues. Ce mécanisme vise à détecter des données aberrantes et est notamment utilisé pour la pondération spécifique à l'outil de calcul de salaire [Salarium](#)¹.
 - Finalement, un dernier filtre élimine totalement les données d'une unité primaire si la proportion de salaires retenus après l'application des filtres précédents est inférieure à un seuil qui dépend de la taille de l'unité considérée. Une unité répondante de classe de taille TA3 = 1 (respectivement 2, 3) est entièrement éliminée si le ratio $\text{anzlohn}/\text{FB01}$ (c'est à dire la proportion de salaires livrés et validés pour tous les filtres précédents) est inférieur à 1/12 (respectivement 1/30, 1/60).

L'idée de filtrer les données livrées est appliquée depuis longtemps à l'OFS. Le processus tel que décrit ci-dessus est en production depuis l'enquête ESS2014. Pour l'enquête en question dans ce rapport (ESS2018), l'étape Behalten_Cook a fait l'objet de révisions concernant les poids utilisés dans la régression linéaire. Il a été décidé d'appliquer une régression équipondérée pour limiter la dépendance de la distance de Cook au poids de l'observation. Une alternative qui utilise tout de même des poids (obtenus lors d'une première pondération sans mécanisme Behalten_Cook) mais calcule un indicateur légèrement différent de la distance de Cook est à l'étude pour l'ESS2020. Pour conclure cette section, il apparaît que deux mécanismes peuvent justifier qu'une unité sélectionnée dans l'échantillon brut soit absente de l'échantillon net. C'est le cas quand elle n'a pas répondu à l'enquête ou bien quand elle a répondu mais a été entièrement éliminée par le PPSD. Ces deux mécanismes vont être modélisés dans la section 3. Dans le secteur privé, le PPSD élimine un peu moins de 60'000 salaires, c'est à dire environ 3% des salaires livrés, voir table 5. De ce fait, un peu moins de 300 unités primaires du secteur privé sont exclues de l'échantillon net en raison du PPSD (voir : table 3) car tous les salaires qu'elles ont livré sont filtrés. Cela représente un peu moins d'1% d'unités primaires écartées par le PPSD.

2.4 Unités inéligibles, unités inactives

Les unités primaires qui répondent à l'enquête et dont les données passent les filtres du PPSD vont recevoir des poids selon une procédure décrite en section 3. Il existe cependant des unités supplémentaires présentes dans l'échantillon brut qui font l'objet d'un traitement spécifique et qui sont parfois rajoutées à l'échantillon net même si elles ne livrent aucun salaire.

Premièrement, certaines unités de l'échantillon brut sont dites "inéligibles". Il s'agit d'unités qui auraient du être écartées de l'échantillon brut (pas sélectionnables) voire du cadre de sondage.

1. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/travail-remuneration/salaires-revenus-cout-travail/niveau-salaires-suisse/salarium.html>

C'est la section LOHN qui déduit l'éligibilité d'une unité de l'échantillon brut sur la base de renseignements actualisés. Le traitement des unités inéligibles dans les différentes étapes de la pondération (notamment : modèle de réponse et calage) est décrit en section 3. Pour l'ESS2018, l'échantillon brut comporte environ 45'600 unités, dont environ 400 sont inéligibles.

Certaines unités de l'échantillon brut sont considérées comme "inactives". Il peut s'agir par exemple d'entreprises qui n'ont plus d'activité en Suisse, plus de personnel, ou qui ont fusionné avec une autre entité. C'est aussi la section LOHN qui déduit, à partir d'informations actualisées, le statut d'activation des unités de l'échantillon brut. Le traitement de ces unités est aussi décrit en section 3. Pour l'ESS2018, environ 2'850 unités sont inactives.

3 Non-réponse et calage

3.1 Vue d'ensemble de la pondération - un plan à trois degrés

L'objectif de la pondération est d'attribuer des poids à des individus qui touchent un salaire standardisé (MBLS). D'après la théorie des sondages, pour éviter les biais dans les estimations, il suffit que le poids d'un individu soit égal à l'inverse de sa probabilité d'inclusion dans l'échantillon net (voir Tillé, 2019, chapitre 2). Dans le cas de l'ESS, le tirage des individus se fait selon un plan de sondage à deux degrés. Le premier degré est le tirage des unités primaires. C'est dans cette étape que la plupart du travail de pondération est effectué. En effet, les unités primaires vont recevoir un poids calé, noté w_i selon la procédure suivante.

- Premièrement, des poids d'unités primaires avant calage, notés d_i , sont calculés. Ces poids sont égaux à l'inverse de la probabilité d'inclusion estimée (dans l'échantillon net) de l'entreprise. Mathématiquement

$$d_i = \frac{1}{\pi_i \hat{p}_i^{\text{rep}} \hat{p}_i^{\text{PPSD}}} \quad (1)$$

où π_i est la probabilité de tirage dans l'échantillon brut ; $\hat{p}_i^{\text{rep}}$ est la probabilité conditionnelle de réponse d'une unité primaire i sélectionnée, estimée à l'aide d'un modèle de réponse (voir section 3.2) et $\hat{p}_i^{\text{PPSD}}$ est la probabilité conditionnelle que le PPSD ne filtre pas l'intégralité des données d'une unité répondante i , c'est à dire qu'il y ait au minimum un employé de l'unité i dont les données passent les filtres. Cette probabilité est aussi estimée en utilisant un modèle décrit en section 3.2. La séparation entre réponse et PPSD est possible car l'échantillon net est disponible pour la pondération avant et après le PPSD². Les quantités $\hat{p}_i^{\text{rep}}$ et $\hat{p}_i^{\text{PPSD}}$, obtenues à l'aide de modèles, sont des estimations des probabilités de réponse et de non-élimination des unités, respectivement. Ainsi, d_i est une estimation de l'inverse de la probabilité d'inclusion de l'unité i . Il est important de noter que les unités inéligibles et inactives reçoivent aussi un poids avant calage, parce que ces unités ne sont connues que pour l'échantillon brut et pas pour tout le cadre de sondage. Elles doivent donc être intégrées dans le calage car on ne connaît pas les marges qui concernent uniquement les unités éligibles et actives. Les choix effectués pour ces unités pour les calcul de $\hat{p}_i^{\text{rep}}$ et $\hat{p}_i^{\text{PPSD}}$ sont décrits en section 3.2.2 et 3.2.3.

- Ensuite, un calage des poids d_i sur des totaux de variables auxiliaires bien choisies est effectué. Le résultat du calage est un poids final, w_i de l'unité primaire i . Cette procédure est décrite dans les sections 3.3 et 3.4.

2. Ce n'est pas le cas dans toutes les enquêtes. En l'absence de séparation, la probabilité dite "de réponse" englobe généralement le mécanisme de réponse et le mécanisme de retenir une unité lors du PPSD.

La seconde phase du plan de sondage consiste à tirer `anzlohn` individus parmi les `FB01` employés de l'unité primaire (voir section 2.1.3 pour un rappel sur la définition de ces variables). Ce tirage est effectué par l'unité primaire qui répond à l'enquête et l'OFS n'a donc pas de contrôle direct sur celui-ci. Cependant, compte tenu des directives données par l'OFS aux unités primaires et de la pondération qui sera appliquée aux unités secondaires, on peut considérer que cette seconde phase se déroule selon un plan simple de taille fixe et que ces tirages sont indépendants d'une unité primaire à l'autre. Ainsi, pour une unité i avec un poids final après calage de w_i , un individu employé dans cette unité recevra un poids de $w_i \times \text{FB01} / \text{anzlohn}$. Cette démarche repose aussi sur l'hypothèse que l'exclusion des données d'un individu par le PPSD à l'intérieur d'une unité primaire se fait à probabilité égale. Cela signifie qu'à l'intérieur d'une unité primaire un employé n'a pas plus de chance qu'un autre de voir ses données exclues.

Les poids individu obtenus à partir des poids des unités primaires ne sont pas les poids finaux utilisés pour des estimations telles que celle du salaire médian en Suisse ou dans un domaine. En effet, l'OFS s'intéresse au salaire médian *d'heures de travail standardisées effectuées en Suisse*. Pour comprendre la différence, il convient d'imaginer que chaque individu effectue un certain nombre d'heures de travail correspondant parfois à un poste à plein temps et parfois non. D'une certaine façon, l'OFS devrait donner un poids à chaque heure de travail effectuée en Suisse. Il y a donc au final un troisième degré dans le plan de sondage, où chaque individu "livre" une certaine quantité d'heures de travail qu'il a effectuées. Chacune de ces heures reçoit un poids identique au poids de l'individu qui les a livrées. Une manière efficace de réaliser cette opération sans explicitement donner des poids à chaque heure de travail effectuée en Suisse est de conserver les poids individus, $w_i \times \text{FB01} / \text{anzlohn}$, et de les multiplier par la variable `IBGRS` de l'individu (voir section 2.1.2 pour une définition). La variable `IBGRS` est en effet directement proportionnelle au nombre d'heures de travail effectuées. Les poids finaux sont ainsi livrés par l'OFS sous forme d'une table contenant les salaires standardisés des individus (ainsi que d'autres informations) et des poids finaux, égaux à $w_i \times \text{IBGRS} \times \text{FB01} / \text{anzlohn}$. Les variables `FB01` et `anzlohn` sont spécifiques à l'unité primaire i qui emploie l'individu. La variable `IBGRS` est spécifique à l'individu. La construction du poids w_i de l'unité primaire qui emploie l'individu est détaillée dans les sections suivantes : voir sections 3.2, 3.3, et 3.4.

3.2 Modèle de réponse

3.2.1 Généralités

La construction d'un poids avant calage pour les unités primaires de l'échantillon net nécessite une modélisation des probabilités $\hat{p}_i^{\text{rep}}$ et $\hat{p}_i^{\text{PPSD}}$, voir équation (1).

La probabilité de réponse d'une unité i , $\hat{p}_i^{\text{rep}}$, est estimée à l'aide d'une régression logistique avec modèle logit. La quantité $\text{logit}(\hat{p}_i^{\text{rep}}) := \log(\hat{p}_i^{\text{rep}} / (1 - \hat{p}_i^{\text{rep}}))$ est modélisée comme dépendant linéairement d'un certain nombre de prédicteurs, sous la forme classique :

$$\text{logit}(\hat{p}_i^{\text{rep}}) = \beta_0 + \sum_{j=1}^{n_{\text{pred}}} \beta_j x_{i,j}$$

où n_{pred} est le nombre de prédicteurs (sans la constante), $\beta_0, \dots, \beta_{n_{\text{pred}}}$ sont les coefficients de la régression logistique à estimer, et $x_{i,j}$ est la valeur du j -ème prédicteur pour l'unité i . Les coefficients $\beta_0, \dots, \beta_{n_{\text{pred}}}$ ont été estimés par maximum de vraisemblance à l'aide de la procédure "logistic" du logiciel SAS (version 7.15).

3.2.2 Modèles de réponse

Pour l'enquête 2018, deux pondérations distinctes sont effectuées, comme expliqué en section 2.1.1. Ces deux pondérations vont comporter chacune leurs modèles de réponse et leur calage.

Dans le secteur privé, il y a environ 45'600 unités dans l'échantillon brut. Un modèle de régression logistique est ajusté en excluant les entreprises inéligibles. Les entreprises inactives sont à l'inverse bien prises en compte dans l'ajustement du modèle, et sont comptées comme des unités répondantes. Au final, le modèle logistique est ajusté sur environ 45'200 unités, avec environ 36'700 répondantes, ce qui permet d'estimer $\hat{p}_i^{\text{rep}}$ pour ces unités. Pour les 400 unités inéligibles exclues du modèle, on utilise $\hat{p}_i^{\text{rep}} = 1$.

Les prédicteurs du modèle de régression logistique sont sélectionnés à l'aide de l'option "step-wise" par défaut de la procédure "logistic" dans SAS. Cette option permet de sélectionner les prédicteurs les plus influents sur la probabilité de réponse. Seuls les prédicteurs ayant un impact statistiquement significatif sur la probabilité de réponse sont retenus dans le modèle final. Dans le modèle de réponse final du secteur privé, les prédicteurs retenus sont:

- la NOGA de l'unité (variable catégorielle à 39 modalités),
- le pourcentage de salariés de l'unité primaire payés, selon les données AVS, au-dessus du revenu AVS médian de la NOGA de l'unité. Cette variable est calculée à l'aide d'une table de revenus AVS (voir section 2.1.1). Concrètement, pour chaque NOGA, un revenu AVS médian est calculé, ce qui permet d'obtenir pour chaque unité primaire la proportion de revenus AVS au-dessus de la médiane de sa NOGA. Ce pourcentage dépend de l'unité primaire et est transformé en une variable catégorielle à 11 modalités. Les modalités sont obtenues en arrondissant le pourcentage au 10% le plus proche, de sorte que les valeurs possibles sont 0%, 10%, 20%, ..., 100%. Le fait de rendre cette variable catégorielle permet de modéliser avec plus de flexibilité des effets non monotones et non linéaires de ce pourcentage sur le logit de la probabilité de réponse de l'unité primaire. Dans le secteur privé, il y a globalement de meilleurs taux de réponse chez les unités dont une plus grande part des employés ont un revenu AVS supérieur à la médiane.
- le croisement entre la variable GR (7 modalités) et TA3 (classe de taille à 3 modalités, décrite dans la section 2.1.3).
- une variable de taille, Taille8 en 8 modalités, basée sur betot_cadre. La variable prend respectivement les valeurs 1, 2, 3, 4, 5, 6, 7 si betot_cadre est compris respectivement dans les intervalles [3, 9], [10, 49], [50, 249], [250, 499], [500, 999], [1'000, 1'999], [2'000, 4'999]. Enfin, si betot_cadre est supérieur ou égal à 5'000 alors Taille8 = 8. L'introduction de cette variable vise notamment à mieux prédire les taux de réponse des très grandes unités.
- le pourcentage de salariés de l'unité qui sont de sexe masculin. Cette variable continue est rendue catégorielle en 11 modalités selon le même procédé que pour le pourcentage de salariés payés au-dessus du revenu AVS médian. La variable a été introduite dans le modèle à cause du taux de réponse légèrement meilleur (et statistiquement significatif) des unités qui emploient un plus grand pourcentage de femmes.
- la variable betot_cadre croisée avec la NOGA. L'introduction de cette variable permet de modéliser un effet linéaire de betot_cadre sur le logit de la probabilité de réponse, et autorise cet effet à dépendre de la NOGA.
- la variable betot_cadre croisée avec GR.

Le modèle de régression logistique retenu nécessite d'estimer au total 129 paramètres, ce qui est à mettre en relation avec la taille de l'échantillon brut du secteur privé (environ 45'200 unités dans le modèle, et un peu plus de 80% de taux de réponse).

Dans l'administration publique, un modèle plus simple est choisi. Le modèle concerne les 283 communes présentes dans l'échantillon brut. La probabilité de réponse des cantons et de la Confédération n'est pas modélisée car ces derniers répondent tous et reçoivent directement un poids final de 1. Pour modéliser des probabilité de réponse sur un échantillon aussi petit, avec de plus seulement 23 unités qui ne répondent pas, il convient d'être parcimonieux sur le nombre de prédicteurs utilisés dans la régression logistique. Les prédicteurs retenus pour modéliser la réponse dans l'administration publique sont :

- une variable catégorielle à 4 modalités. Les trois premières modalités coïncident avec les classes de taille 1, 2 et 3 de la variable TA3 décrite en section 2.1.3. La quatrième modalité regroupe les communes dites "profiling" qui font l'objet d'un suivi plus important dans le processus de collecte des données de l'OFS. Une commune profiling est ainsi classée dans la quatrième modalité tandis qu'une commune non profiling est classé dans l'une des trois première modalités selon son `betot_cadre`.
- la variable GR (7 modalités).
- la variable `betot_cadre` croisée avec GR. L'introduction de cette variable permet de prendre en compte un effet linéaire de `betot_cadre` sur le logit de la probabilité de réponse. Cet effet peut varier avec la grande région.

Le modèle ainsi obtenu contient 17 paramètres à estimer, ce qui est déjà important compte tenu de la taille de l'échantillon.

3.2.3 Modèle de non-élimination des données

Le second modèle vise à estimer, pour les unités de l'échantillon brut qui répondent à l'enquête, la probabilité $\hat{p}_i^{\text{PPSD}}$ que les données d'au moins un employé de l'unité primaire passent les différents filtres du PPSD. Le PPSD est un mécanisme interne à l'OFS. L'élimination par le PPSD ne peut pas être considérée comme une non-réponse, et est causée par des mécanismes différents.

Pour le secteur privé, $\hat{p}_i^{\text{PPSD}}$ est à nouveau estimé à l'aide d'un modèle de régression logistique. Les unités prises en compte pour ajuster le modèle sont les unités répondantes en excluant les unités inactives et inéligibles ce qui donne environ 33'850 unités. Parmi elles, environ 300 sont éliminées par le PPSD. L'élimination par le PPSD est donc un phénomène relativement rare qu'il convient de modéliser avec un nombre limité de prédicteurs pour éviter la sur-paramétrisation. Le modèle permet d'estimer la quantité $\hat{p}_i^{\text{PPSD}}$ pour toutes les unités utilisées lors de l'ajustement du modèle. Les unités inactives et inéligibles ne peuvent par définition pas être éliminées par le PPSD et on pose ainsi $\hat{p}_i^{\text{PPSD}} = 1$ pour ces unités.

Il faut rappeler qu'une unité est considérée comme éliminée lorsque tous les salaires livrés sont filtrés. Intuitivement, la probabilité qu'une unité ne soit pas éliminée est liée (possiblement de manière non linéaire) au nombre d'employés de l'unité. Les prédicteurs utilisés dans le modèle ont été choisis selon cette idée. Les prédicteurs sont:

- la variable `Taille8` décrite dans la section 3.2.2,
- la grande région, GR,
- la variable `betot_cadre` croisée avec la variable TA3 décrite dans la section 2.1.3.

Ce modèle pour le secteur privé possède 17 paramètres.

Pour l'administration publique, seule une unité parmi les 260 répondantes voit ses données entièrement filtrées par le PPSD. Comme il n'est pas raisonnable de mettre en place un modèle

logistique pour modéliser l'élimination d'une seule unité, il a été décidé (pour l'administration publique uniquement) de fusionner les modèles de réponse et d'élimination par le PPSD dans l'administration publique. Ainsi, au lieu de considérer que 23 unités ne répondent pas et qu'une unité est éliminée, un unique modèle qui modélise la non-réponse de 24 unités est utilisé. Ce modèle a déjà été décrit dans la section 3.2.2.

3.3 Construction des variables de calage

3.3.1 Situation en l'absence de calage et critères de choix des variables de calage

Les poids avant calage d_i sont à présent construits, voir équation (1). Ces poids d'unités primaires pourraient en théorie être utilisés pour la pondération définitive pour effectuer des estimations approximativement sans biais, en supposant que la modélisation de la probabilité de réponse est adéquate. L'OFS pourrait donc livrer un fichier avec des poids individus égaux à $d_i \times \text{IBGRS} \times \text{FB01} / \text{anzlohn}_i$, voir section 3.1.

Dans le rapport de méthodes sur le plan de sondage, [Qualité & Potterat \(2020\)](#) montrent qu'avec le plan de sondage utilisé et en l'absence de calage, un estimateur à priori conservateur (voir aussi [Kim & Kim \(2007\)](#) pour le détail) de la variance du total d'une variable y relevée sur des individus est:

$$\widehat{\text{var}}(\hat{T}) = \sum_{i \in S} d_i (d_i - 1) \hat{T}_i^2 + \sum_{i \in S} d_i \text{FB01}_i^2 \left(1 - \frac{\text{anzlohn}_i}{\text{FB01}_i} \right) \frac{s_i^2}{\text{anzlohn}_i} \quad (2)$$

où:

- S désigne l'échantillon net d'unités primaires,
- d_i est le poids avant calage de l'unité i , donné par l'équation (1),
- FB01_i et anzlohn_i sont les valeurs des variables FB01 et anzlohn pour l'unité i ,
- $\hat{T}_i$ est l'estimateur du total de la variable y (mesurée sur les salariés) pour l'unité primaire i . Si S_i désigne l'ensemble des individus de l'unité i dont les données sont collectées par l'enquête, alors :

$$\hat{T}_i = \frac{\text{FB01}_i}{\text{anzlohn}_i} \sum_{k \in S_i} y_k$$

- s_i^2 désigne la variance empirique de la variable y au sein de l'unité i . Pour être plus explicite, dans le cas où l'unité i livre des données sur $\text{anzlohn}_i > 1$ individus, alors:

$$s_i^2 = \frac{1}{\text{anzlohn}_i - 1} \sum_{k \in S_i} (y_k - \bar{y}_i)^2, \text{ avec } \bar{y}_i = \frac{1}{\text{anzlohn}_i} \sum_{k \in S_i} y_k$$

Lorsque $\text{anzlohn}_i = 1$, l'équation ci-dessus ne peut pas être utilisée et un autre estimateur doit être choisi pour obtenir s_i^2 . Ce point est discuté en section 4.2.

Lorsqu'un calage est effectué avec des variables auxiliaires bien choisies, [Qualité & Potterat \(2020\)](#) montrent que l'estimateur de variance de l'équation (2) s'écrit quasiment de la même façon. Seul le terme $\hat{T}_i^2$ doit être remplacé par $\hat{\varepsilon}_i^2$, où $\hat{\varepsilon}_i$ est le résidu de l'unité i dans la régression linéaire où l'on cherche à prédire (pour les unités de l'échantillon net) la variable $\hat{T}$ à l'aide des variables de calage. Ainsi, il faut trouver des variables de calage disponibles sur l'échantillon net, dont le total sur le cadre est connu, et capables de bien prédire $\hat{T}$ par un modèle linéaire. Avec de telles variables, les carrés des résidus $\hat{\varepsilon}_i^2$ sont petits par rapport à $\hat{T}_i^2$, ce qui signifie une baisse de la variance de l'équation (2); ou en d'autres termes une meilleure précision de l'estimation.

Les tables 9, 10 et 11 donnent les gains de précision obtenus par l'ajout de certaines variables de calage liées aux revenus AVS et détaillées en section 3.3.2.

Les arguments avancés ici concernent l'amélioration de la précision de l'enquête pour l'estimation du *total* d'une variable, notée y , mesurée sur des individus. Le calcul est généralisé par [Qualité & Potterat \(2020\)](#) dans le cas où l'on veut une variance pour l'estimateur de la *médiane* de y et non de son total. À ce moment là, ce n'est plus la variable y qui est agrégée au niveau de chaque unité primaire, mais la "linéarisée de l'estimateur de la médiane de y ":

$$\hat{Z}_i = \frac{FB01_i}{anzlohn_i} \sum_{k \in S_i} z_k \quad (3)$$

où z_k est une linéarisée pour l'unité k de l'estimateur de la médiane. La linéarisée dépend du paramètre d'intérêt estimé et plusieurs définitions sont proposées dans la littérature. Pour l'ESS2018, c'est la méthode de [Deville \(1999\)](#) et [Osier \(2009\)](#) qui est retenue. La linéarisée de la médiane vaut ainsi:

$$z_k = C(2 I(y_k > q_{0.5}) - 1),$$

où C est une constante (voir [Qualité & Potterat \(2020\)](#), section 3.1.5, pour une formule explicite), $q_{0.5}$ est un estimateur de la médiane des y_k produit en utilisant la procédure FREQ du logiciel SAS et I est la fonction indicatrice. Il apparaît ainsi que la linéarisée de la médiane est égale à une constante multipliée par 1 ou -1 selon que le salaire de l'unité secondaire en question est au-dessus ou en dessous du salaire médian dans le domaine. Ainsi, l'agrégation $\hat{Z}_i$ des linéarisées au niveau entreprise donnée dans l'équation (3) est proportionnelle à la différence entre le nombre de salariés (équivalent plein temps) payés au-dessus d'un salaire médian et le nombre de salariés payés en dessous. Cette intuition est très utile pour le choix des variables de calage puisque celles-ci sont choisies de sorte à bien prédire $\hat{Z}_i$.

Dans ce rapport, la variable d'intérêt y , mesurée sur les salariés, est le salaire standardisé, dont on veut estimer une médiane et des totaux dans des domaines. Les domaines sont ici en adéquation avec les statistiques publiées par l'OFS. Ainsi, pour le salaire médian, les domaines sont la NOGA et les grandes régions. Pour les totaux, les domaines sont la Section NOGA et la taille de l'unité, voir section 5.

3.3.2 Choix et construction des variables de calage

Secteur privé :

Lors des ESS de 2012, 2014 et 2016, des variables de calage relativement classiques ont été choisies pour la pondération du secteur privé. Ces variables de calage ont toutes été reprises pour l'ESS2018, et sont listées ci-dessous:

- calage sur le nombre d'employés, `betot_cadre`, par NOGA. Concrètement, 39 variables de calage différentes sont créées, correspondant aux 39 NOGA. Pour une unité primaire appartenant à la i -ème NOGA, les 39 variables de calage prennent toutes la valeur 0, à l'exception de la i -ème variable de calage qui prend la valeur du `betot_cadre` de l'unité considérée. Pour résumer cette opération de multiplication d'une variable quantitative (`betot_cadre`) par la variable indicatrice de chaque NOGA, on parlera des variables "betot_cadre par NOGA".
- calage sur le nombre d'employés par croisement $TA3 \times GR$. Le croisement $TA3 \times GR$ est ici une variable catégorielle avec $3 \times 7 = 21$ modalités. Il y a donc ici 21 variables `betot_cadre` par croisement $TA3 \times GR$.
- calage sur le nombre d'unités primaires par NOGA. Si "UN" est la variable constante égale à 1 pour toutes les unités primaires, il y a concrètement 39 variables de calage UN par NOGA.

- calage sur le nombre d'unités primaires par croisement TA3 × GR. Il y a 21 variables de calage UN par croisement TA3 × GR.

En plus des 120 variables mentionnées ci-dessus, qui sont toutes conservées pour l'ESS2018, 85 nouvelles variables de calage ont été ajoutées en recherchant de bons prédicteurs dans des régressions linéaires selon les idées présentées dans la section 3.3.1. Parmi elles, 12 variables sont relativement simples et sont présentées en premier :

- calage sur le nombre d'employés (*betot_cadre*) par *Taille6* : la variable *Taille6* est une simple variable de taille de l'unité primaire, en 6 modalités, basée sur *betot_cadre*. *Taille6* prend respectivement les valeurs 1, 2, 3, 4, 5 si *betot_cadre* est strictement inférieur à respectivement 10, 50, 250, 500, 1000. Si *betot_cadre* est supérieur ou égal à 1000, alors *Taille6* = 6. Ces classes de taille sont utilisées dans les publications.
- calage sur le nombre d'unités primaires par *Taille6*.

Parmi les 85 nouvelles variables de calage, 73 variables sont basées sur la table de revenus AVS décrite en section 2.1.1. En plus de leur utilisation pour le calage, ces données ont servi notamment à l'élaboration du plan de sondage, à certains filtres du PPSD, et pour le modèle de réponse. Les 73 variables de calage supplémentaires sont les suivantes :

- *nb_sup40*, *nb_sup50*, *nb_sup60* : à partir de la table de revenus AVS, des percentiles à 40%, 50%, 60% des revenus AVS sur toute la Suisse sont calculés. Ensuite, pour chaque unité primaire, toujours à partir des revenus AVS, le nombre de revenus supérieurs aux percentiles à 40%, 50%, 60% est compté et stocké dans ces 3 variables. L'utilisation de ces 3 variables de calage dans la pondération du secteur privé permet d'assurer que ces trois percentiles sont exactement respectés dans l'échantillon pondéré.
- La variable *nb_sup50NOGA* est construite selon un schéma similaire. À partir de la table de revenus AVS, un revenu médian est calculé dans chaque NOGA. Pour une unité donnée, le nombre de revenus supérieurs à ce revenu médian de la NOGA à laquelle l'unité appartient est alors compté. La variable *nb_sup50NOGA* est multipliée par les 39 variables indicatrices des NOGA. L'utilisation de ces 39 variables de calage permet d'améliorer l'estimation de la médiane par NOGA.
- Dans le même esprit, la variable *nb_sup50GR* multipliée par les variables indicatrices des grandes régions est utilisée. Il s'agit de 7 variables de calages, associées aux 7 grandes régions, qui permettent de caler le nombre d'individus dont le revenu (selon l'AVS) dépasse le revenu AVS médian calculé au niveau de leur grande région.
- Enfin, la variable *incometotalAVS*, obtenue en agrégeant au niveau des unités primaires l'ensemble des revenus AVS annuels, est utilisée. Pour la construction des variables de calage, *incometotalAVS* est multiplié par les 18 variables indicatrices des Sections NOGA ; ainsi que par les 6 variables indicatrices de chaque *Taille6*. L'utilisation de ces variables permet d'améliorer l'estimation de totaux par Section NOGA et *Taille6*.

L'introduction de ces 85 nouvelles variables de calage porte le nombre total de variables de calage à 205, ce qui reste raisonnable devant le nombre d'unités dans l'échantillon net qui est de l'ordre de 37'100. Les variables de calage qui comptent le nombre de revenus AVS supérieurs à un revenu AVS médian dans un domaine visent à améliorer la précision de l'estimateur du salaire médian dans ce domaine. Les variables de calage basées sur *incometotalAVS* visent plutôt à améliorer la précision d'estimateurs du coût total de la main d'œuvre dans des domaines. Des variables de calage supplémentaires, dépendant notamment du genre, ont été testées mais n'ont pas été retenues car elles n'amélioreraient pas significativement les précisions par rapport aux autres variables déjà introduites.

Administration publique :

Comme expliqué précédemment, une pondération est effectuée séparément pour les unités de l'administration publique telles que les communes. Dans le cas présent cela concerne environ 250 unités présentes dans l'échantillon net sur lesquelles un poids avant calage est calculé. Compte-tenu de la taille de cet échantillon, il est exclu d'utiliser les mêmes ordres de grandeur que dans le secteur privé quant au nombre de variables de calage. Les enquêtes ESS de 2012, 2014 et 2016 utilisaient un calage sur le nombre d'unités par croisement $GR \times TA4$, où $TA4$ est une variable égale à $TA3$ (voir section 2.1.3 pour une définition) si la commune n'est pas une commune profiling et prend la valeur 4 si la commune est une commune profiling. Cela représente donc 28 variables de calage, ce qui est déjà conséquent. Pour l'ESS2018, le choix est fait de ne pas reprendre ces variables telles quelles. Au lieu de cela, les 16 variables de calage suivantes sont proposées:

- calage sur le nombre d'unités par $TA4$, en regroupant les classes $TA4= 4$ et $TA4= 3$. Notons que toutes les communes profiling ont aussi un `betot_cadre` supérieur ou égal à 50, ce qui rend le regroupement proposé cohérent (3 variables de calage).
- calage sur le nombre d'employés par $TA4$, en regroupant les classes $TA4= 4$ et $TA4= 3$ (3 variables de calage).
- calage sur le nombre d'employés par GR (7 variables de calage).
- utilisation des variables de calage `nb_sup60`, `nb_sup60GR`, `incometotalAVS`, construites exactement selon la même procédure que dans la pondération du secteur privé (3 variables de calage). Il est à noter que ces variables ne sont multipliées par aucune variable indicatrice pour conserver un nombre total de variables de calage limité.

Le choix du calage du nombre d'employés ou d'unités par classe de taille et par grande région vise à maintenir une précision comparable aux enquêtes précédentes en limitant le nombre total de variables de calage. Le choix de ne pas caler sur le croisement entre $TA4$ et GR permet de plus d'ajouter des variables de calage basées sur les revenus AVS, ce qui vise à améliorer la précision de l'estimateur du salaire médian.

3.4 Méthode de calage

La construction des poids avant calage d_i est décrite en section 3.1 et en particulier par l'équation (1). Les variables de calage sont elles décrites en section 3.3. Pour l'ESS2018, dans le secteur privé, environ 37'100 unités ont ainsi un poids avant calage. Cela inclut les unités inéligibles et inactives. Notons $x_{i,j}^{cal}$ la valeur de la j -ème variable de calage de l'unité primaire i . Si S désigne l'échantillon net des unités primaires, le calage consiste à transformer les poids avant calage d_i en poids calés w_i de sorte à satisfaire l'équation de calage

$$\sum_{i \in S} w_i x_{i,j}^{cal} = \sum_{\text{cadre de sondage}} x_{i,j}^{cal} \quad (4)$$

pour chacune des variables de calage, et ce en modifiant le moins possible les poids avant calage (voir Deville & Särndal, 1992). Le terme de droite de l'équation (4) est un total sur le cadre de sondage de la j -ème variable de calage qui est supposé connu et qui est noté t_j^{cadre} . Le calage a été réalisé en utilisant la distance "logit" (cas 6 de Deville & Särndal, 1992) au moyen du programme Calmar de l'Insee (voir Sautory & Le Guennec, 2003). Les facteurs q_i qui interviennent dans le calage ont été choisis comme suit :

$$q_i = \begin{cases} 1/\text{nb_income} & \text{si nb_income} > 0, \\ 1/\text{betot_cadre} & \text{sinon,} \end{cases}$$

où nb_income est le nombre de revenus pour l'unité i présents dans la table de revenus AVS. De plus, l'équation (4) a été légèrement modifiée de sorte à garantir des poids après calages $w_i \geq 1$. Le fait de garantir des poids après calage supérieurs ou égaux à 1 permet d'avoir au final des poids individus, $w_i \times \text{FB01}/\text{anzlohn}$, supérieurs ou égaux à 1. L'équation (4) est modifiée de la façon suivante :

$$\sum_{i \in S} w_i x_{i,j}^{\text{cal}} - \sum_{i \in S} x_{i,j}^{\text{cal}} = t_j^{\text{cadre}} - \sum_{i \in S} x_{i,j}^{\text{cal}}$$

ce qui donne

$$\sum_{i \in S} (w_i - 1) x_{i,j}^{\text{cal}} = t_j^{\text{cadre}} \quad (5)$$

où $t_j^{\text{cadre}} = t_j^{\text{cadre}} - \sum_{i \in S} x_{i,j}^{\text{cal}}$. En notant $w'_i := w_i - 1$, l'équation (5) est une nouvelle équation de calage sur les "poids moins 1", similaire à l'équation (4). Pour appliquer cette équation de calage, il est possible d'utiliser l'algorithme de calage tel que décrit précédemment, mais avec des poids avant calage de $d_i - 1$ (au lieu de d_i) qui sont calés sur le total t_j^{cadre} (au lieu de t_j^{cadre}). L'algorithme renvoie alors les poids w'_i , ce qui permet d'avoir des poids après calage $w_i = w'_i + 1$ supérieurs ou égaux à 1. Cette opération fonctionne en pratique car les poids avant calage d_i sont supérieurs ou égaux à 1 et parce que l'algorithme de calage fournit nécessairement des poids positifs $w'_i \geq 0$ (avec la distance "logit"), ce qui garantit $w_i \geq 1$.

Pour analyser les transformations sur les poids obtenus via le calage, il est courant d'analyser les g-weights, définis comme étant les ratios w_i/d_i . Le g-weight maximal est d'environ 2.5 et 18 unités ont un g-weight supérieur à 2. Le g-weight minimal est d'environ 0.2 et 11 unités ont un g-weight inférieur à 0.25. La moyenne des g-weights est proche de 1, ce qui est le signe de poids avant calage convenablement construits. De plus, 95% des g-weights sont situés dans l'intervalle [0.89, 1.20]. Pour la pondération de l'administration publique, les poids de 259 unités sont calés à l'aide des 16 variables de calage décrites dans ce rapport. Les g-weights sont tous compris entre 0.25 et 2.14 et valent en moyenne environ 1.07. Ainsi, le calage a un impact limité sur les poids avant calage comme souhaité.

Le prix à payer pour assurer des poids calés supérieurs ou égaux à 1 est un calage parfois plus difficile à atteindre. L'équation de calage (5) offre une flexibilité plus faible pour modifier les poids des unités qui sont initialement proches de 1. Lorsque le calage n'est pas atteint il est possible de modifier une nouvelle fois l'équation de calage de sorte à caler "les poids moins 1" pour certaines unités choisies et "les poids" pour d'autres. Ceci est réalisé de la façon suivante : on divise l'échantillon net S en 2 parties disjointes $S^{(1)}$ et $S^{(0)}$. Les unités dans $S^{(1)}$ sont celles pour lesquelles on souhaite garantir des poids supérieurs ou égaux à 1. Les unités dans $S^{(0)}$ sont celles où l'on a besoin de plus de flexibilité pour atteindre le calage. On définit ainsi $w''_i = w_i - 1$ si $i \in S^{(1)}$ et $w''_i = w_i$ sinon. L'équation (4) est modifiée de la façon suivante :

$$\sum_{i \in S} w_i x_{i,j}^{\text{cal}} - \sum_{i \in S^{(1)}} x_{i,j}^{\text{cal}} = t_j^{\text{cadre}} - \sum_{i \in S^{(1)}} x_{i,j}^{\text{cal}}$$

ce qui donne la nouvelle équation de calage

$$\sum_{i \in S} w''_i x_{i,j}^{\text{cal}} = t_j^{\text{cadre}} \quad (6)$$

où $t_j^{\text{cadre}} = t_j^{\text{cadre}} - \sum_{i \in S^{(1)}} x_{i,j}^{\text{cal}}$. Pour les unités dans $S^{(1)}$ on a bien $w_i = w''_i + 1 \geq 1$. En revanche, pour les unités dans $S^{(0)}$, il n'y a plus de garantie d'avoir des poids supérieurs ou égaux à 1. Dans le secteur privé, le calage avec l'équation (5) fonctionne sans avoir besoin de faire

appel à l'équation (6). Pour l'administration publique, il a été nécessaire d'utiliser l'équation (6). Les 259 poids après calage obtenus sont cependant tous supérieurs ou égaux à 1, même si cela n'était pas garanti.

La table 1 donne des statistiques sur les poids obtenus pour les unités primaires du secteur privé. Les poids de sondage sont ceux obtenus lors de l'élaboration du plan de sondage. La colonne "poids après NR" indique les poids après prise en compte du modèle de réponse. Les poids avant calage sont donnés dans la colonne "Poids après élim. PSD". Enfin, les poids après calage sont indiqués sur la dernière colonne. Pour l'administration publique, la table 2 comporte une colonne de moins car les modèles de réponse et d'élimination par le PSD ont été réunis. La table présente ainsi les poids de sondage et les poids avant et après calage.

TABLE 1 Statistiques (moyenne, minimum, percentiles à 10%, 25%, 50%, 75%, 90%, maximum) sur les poids obtenus dans le secteur privé.

	Poids sondage	Poids après NR	Poids après élim. PSD	Poids après calage
moyenne	4.11	5.07	5.12	5.07
minimum	1	1	1	1
p10	1	1.14	1.14	1.13
p25	1	1.24	1.26	1.26
p50	1.56	1.87	1.92	1.89
p75	3.13	3.80	3.82	3.80
p90	8.18	10.21	10.21	10.56
maximum	95	135	136	131

TABLE 2 Statistiques (moyenne, maximum, minimum, percentiles à 10%, 25%, 50%, 75%, 90%) sur les poids obtenus dans l'administration publique.

	Poids sondage	Poids avant calage	Poids après calage
moyenne	6.69	7.21	7.75
minimum	1	1	1
p10	1	1	1
p25	2.52	1.14	1.61
p50	4.87	5.20	4.93
p75	11.56	12.89	11.12
p90	14.54	16.02	18.93
maximum	25	27	41

4 Calcul de coefficients de variations et hypothèses

4.1 Hypothèses effectuées pour construire le plan de sondage

Le rapport de méthodes [Qualité & Potterat \(2020\)](#) sur le plan de sondage de l'ESS fournit plusieurs tableaux sur le nombre d'unités primaires (entreprises) et secondaires (individus) attendus dans l'échantillon net ainsi que sur les précisions à obtenir et les précisions attendues. Plus exactement, les précisions sont mesurées par les coefficients de variation (CV) des estimateurs du salaire médian dans un domaine ou d'un total dans un domaine. Le rapport de [Qualité & Potterat \(2020\)](#) fournit aussi des objectifs de précision à respecter (CV visés, voir table 9), lesquels

servent à construire et optimiser le plan de sondage. Des CV attendus, sous certaines hypothèses, sont aussi donnés dans le rapport et reproduits en table 9.

Pour calculer les CV attendus, des hypothèses sont faites sur les probabilités de réponse des unités, ainsi que sur la proportion de salaires livrés par celles-ci. Dans l'administration publique, il est supposé que la Confédération et les cantons répondent toujours et que les districts et communes ont une probabilité de réponse de 80%. Dans le secteur privé, les hypothèses sont des probabilités de réponse de 80% pour les grandes entreprises (TA3= 3) et les entreprises du programme profiling, de 70% pour les entreprises moyennes (TA3= 2) et de 60% pour les petites entreprises (TA3= 1). De plus, il est supposé que les grandes entreprises livrent un tiers de leurs salaires, que les entreprises moyennes en livrent la moitié et que les petites entreprises livrent la totalité, comme demandé dans les instructions du questionnaire. Il est aussi supposé que les salaires livrés sont tous valides, c'est-à-dire qu'aucune donnée n'est éliminée par le PPSD. Enfin, les CV attendus sont calculés sans tenir compte du choix des nouvelles variables de calage basées sur les revenus AVS. Les différentes hypothèses pour calculer la taille attendue de l'échantillon net sont détaillées dans le rapport de [Qualité & Potterat \(2020\)](#), section 3.2.

Comme le montreront les résultats, la taille de l'échantillon net de salaires est largement plus élevée qu'attendu. Cela est dû à deux facteurs. D'une part, le taux de réponse des grandes unités est supérieur au taux utilisé pour construire le plan et calculer une taille attendue d'échantillon net. D'autre part, beaucoup de grandes unités livrent 100% de leurs données, au lieu du tiers présumé dans l'élaboration du plan de sondage. L'échantillon net plus grand qu'attendu mène globalement à des CV meilleurs que prévus, et ce d'autant plus que l'apport des revenus AVS dans le calage est parfois non négligeable.

4.2 Calcul des coefficients de variation

Les CV réellement obtenus sont calculés à l'aide de l'équation (2) adaptée pour tenir compte du calage effectué:

$$\widehat{\text{var}}(\hat{T}) = \sum_{i \in S} w_i (d_i - 1) \hat{\varepsilon}_i^2 + \sum_{i \in S} d_i \text{FB01}_i^2 \left(1 - \frac{\text{anzlohn}_i}{\text{FB01}_i} \right) \frac{s_i^2}{\text{anzlohn}_i} \quad (7)$$

où :

- S , d_i , FB01_i et anzlohn_i sont déjà décrits après l'équation (2),
- w_i est le poids après calage de l'unité i ,
- ε_i est le résidu dans une régression linéaire où l'on cherche à prédire $\hat{T}_i$ à l'aide des variables de calage. Pour calculer le CV d'un estimateur du total, $\hat{T}_i$ est un estimateur de la masse salariale au sein de l'unité i :

$$\hat{T}_i = \frac{\text{FB01}_i}{\text{anzlohn}_i} \sum_{k \in S_i} y_k$$

Pour calculer le CV d'un estimateur du salaire médian, l'équation (7) est utilisée en remplaçant $\hat{T}_i$ par $\hat{Z}_i$ (voir : équation (3)) pour calculer les résidus ε_i . Cela revient à remplacer les salaires y_k par les z_k , voir section 3.3.1.

- s_i^2 est un estimateur de la dispersion des salaires y_k (pour l'estimation du CV d'un total) ou des z_k (pour l'estimation du CV d'une médiane) au sein de l'unité primaire i .

Une formule pour le calcul de s_i^2 est donnée en section 3.3.1 et réécrite ci-dessous. Pour estimer le CV d'un total, la formule est :

$$s_i^2 = \frac{1}{\text{anzlohn}_i - 1} \sum_{k \in S_i} (y_k - \bar{y}_i)^2, \text{ avec } \bar{y}_i = \frac{1}{\text{anzlohn}_i} \sum_{k \in S_i} y_k$$

et pour estimer le CV d'une médiane, la même équation est utilisée en remplaçant les salaires y_k par les z_k . La formule ci-dessus pose des soucis pour les entreprises qui ne livrent des données que pour $\text{anzlohn}_i = 1$ salarié. Dans ce cas là, s_i ne peut pas être calculé facilement. Si l'entreprise livre un seul salaire ($\text{anzlohn}_i = 1$) et déclare n'avoir qu'un seul salarié ($\text{FB01}_i = 1$) alors l'estimation de s_i n'est pas nécessaire à cause du facteur $1 - \frac{\text{anzlohn}_i}{\text{FB01}_i}$ présent dans l'équation (7). En revanche, une estimation de s_i est nécessaire dans le cas $\text{FB01}_i > 1$ et $\text{anzlohn}_i = 1$.

Dans notre programme, l'imputation des valeurs de s_i pour les unités i concernées est faite à l'aide de la valeur moyenne prise par s_i sur un ensemble d'entreprises donneuses. Le choix des entreprises donneuses peut être différent pour chaque unité i . Une option simple consiste à choisir comme donneuses à l'unité i toutes les entreprises qui ont à la fois la même classe de taille (TA3), NOGA et grande région que l'unité i . C'est cette option qui est utilisée pour produire les résultats présentés dans ce rapport. Notons que le second terme de l'équation (7), qui fait intervenir s_i , est généralement faible en comparaison du premier terme. De plus, l'imputation de s_i a lieu pour moins de 1% des unités et toujours pour des unités petites. Ainsi, le choix de la technique pour obtenir les entreprises donneuses a une influence très faible sur les CV finaux. Pour conclure cette section, il faut mentionner que les CV produits via l'équation (7) ne sont pas ceux officiellement publiés par l'OFS. En l'état actuel, la publication de CV identiques à ceux calculés dans ce rapport par l'OFS (mais aussi par les autres utilisateurs de l'enquête) impliquerait de pouvoir livrer des données sensibles telles que les revenus AVS. Les CV publiés actuellement par l'OFS ne prennent pas en compte le bénéfice lié à l'utilisation du calage. Ces CV sont donc généralement plus conservateurs que ceux calculés dans ce rapport.

5 Résultats

5.1 Nombre d'unités et nombre d'individus

Les tables 3, 4 et 7 donnent le nombre d'unités primaires présentes dans l'échantillon brut et l'échantillon net, par domaine, pour le secteur privé et l'administration publique. Des tableaux similaires sont produits pour les unités secondaires (i.e. les employés), voir tables 5, 6 et 8. Voici une description des variables présentes dans ces tableaux :

- Nb. unités brut : nombre d'unités primaires présentes dans l'échantillon brut.
- attendu net : nombre d'unités primaires attendues dans l'échantillon net compte tenu des hypothèses effectuées lors de l'élaboration du plan de sondage sur le taux de réponse et sur la proportion des données livrées (voir section 2.1.3).
- betot brut : nombre d'unités secondaires dans l'échantillon brut, en utilisant la variable `be-tot_cadre`.
- betot attendu net : nombre d'unités secondaires attendues dans l'échantillon net. Le calcul se base sur la valeur de betot brut et sur les hypothèses faites lors de l'élaboration du plan de sondage.
- net avant elim. : nombre d'unités primaires (ou secondaires, selon le tableau) réellement présentes dans l'échantillon net avant d'appliquer le PPSD. Pour les unités primaires, la variable "net avant elim." inclut les entreprises inactives.
- actif avant elim. : nombre d'unités primaires réellement présentes dans l'échantillon net avant d'appliquer le PPSD et sans prendre en compte les entreprises inactives.
- net après elim. : nombre d'unités primaires (ou secondaires, selon le tableau) actives et présentes dans l'échantillon net après le PPSD.

5.1.1 Secteur privé

Nombre d'unités primaires attendu et réalisé par domaine.

Les tables 3 et 4 montrent que, dans le secteur privé, le nombre d'unités primaires dans l'échantillon net est légèrement supérieur à ce qui est attendu. Cela est dû à des taux de réponses un peu meilleurs que ceux retenus dans les hypothèses. De plus, le PPSD élimine peu d'unités primaires.

Nombre d'unités secondaires attendu et réalisé par domaine.

Dans le secteur privé, le nombre d'unités secondaires dans l'échantillon net est souvent très supérieur à ce qui est attendu, comme le montrent les tables 5 et 6. Cela est surtout dû au fait que certaines grandes unités livrent tous leurs salaires au lieu d'un tiers, comme expliqué en section 4.1.

5.1.2 Administration publique

Pour l'administration publique, la taille d'échantillon net au niveau des unités primaires est légèrement meilleure que prévu en raison de taux de réponses plus élevés qu'attendu, voir table 7. Pour les unités secondaires, la taille d'échantillon net est largement plus grande que prévu (voir table 8). Cela s'explique du fait que la Confédération et la plupart des cantons ont livré l'intégralité de leurs salaires, au lieu d'un tiers attendu.

TABLE 3 Secteur privé : nombre d'unités primaires dans l'échantillon brut et attendues/présentes dans l'échantillon net, par NOGA et grande région.

Noga	Nb. brut	attendu net	net avant élim.	actif avant élim.	net après élim.
05-09	110	75	78	73	73
10-11	798	554	553	520	516
12	10	7	10	10	10
13-15	309	201	228	214	210
16-18	584	398	436	419	412
19-20	171	127	141	136	136
21	150	109	127	123	120
22-23	374	266	287	277	277
24-25	665	478	523	515	512
26	418	320	345	340	340
27	195	140	157	153	153
28	422	324	341	341	341
29-30	80	57	61	59	59
31-33	690	453	551	527	521
35	179	136	166	154	154
36-39	389	259	304	270	270
41-43	1'344	1'002	1'033	1'007	1'004
45-46	2'624	1'798	2'159	2'074	2'062
47	2'000	1'321	1'575	1'463	1'456
49-52	901	651	679	651	641
53	109	72	76	72	72
55-56	4'095	2'627	3'116	2'760	2'736
58-60	603	394	469	430	427
61	123	83	106	95	95
62-63	1'467	969	1'264	1'178	1'173
64, 66	1'262	859	1'125	1'012	996
65	198	146	186	176	176
68	3'397	2'090	2'831	2'271	2'246
69-71	3'301	2'148	2'814	2'652	2'629
72	320	211	264	249	248
73-75	2'790	1'715	2'341	2'180	2'159
77, 79-82	2'426	1'587	1'861	1'759	1'743
78	651	500	429	402	400
84	196	128	182	144	144
85	1'516	1'059	1'269	1'148	1'136
86-88	2'722	2'049	2'446	2'370	2'362
90-93	2'670	1'683	2'033	1'682	1'667
94-95	3'580	2'292	3'103	2'617	2'597
96	1'728	1'056	1'428	1'317	1'289
grande région					
Région lémanique	7'149	4'794	5'631	5'212	5'158
Espace Mittelland	9'950	6'585	8'072	7'411	7'348
Suisse du Nord-ouest	5'145	3'478	4'128	3'685	3'661
Zurich	6'932	4'749	5'758	5'328	5'299
Suisse orientale	6'592	4'372	5'415	4'848	4'819
Suisse centrale	6'880	4'481	5'674	5'103	5'052
Tessin	2'919	1'884	2'419	2'253	2'225
Suisse	45'567	30'341	37'097	33'840	33'562

TABLE 4 Secteur privé : nombre d'unités primaires dans l'échantillon brut et attendues/présentes dans l'échantillon net, par Section NOGA et classe de taille.

Section NOGA	Nb. brut	attendu net	net avant élim.	actif avant élim.	net après élim.
B	110	75	78	73	73
C	4'866	3'432	3'760	3'634	3'607
D	179	136	166	154	154
E	389	259	304	270	270
F	1'344	1'002	1'033	1'007	1'004
G	4'624	3'118	3'734	3'537	3'518
H	1'010	724	755	723	713
I	4'095	2'627	3'116	2'760	2'736
J	2'193	1'446	1'839	1'703	1'695
K	1'460	1'006	1'311	1'188	1'172
L	3'397	2'090	2'831	2'271	2'246
M	6'411	4'074	5'419	5'081	5'036
N	3'077	2'086	2'290	2'161	2'143
O	196	128	182	144	144
P	1'516	1'059	1'269	1'148	1'136
Q	2'722	2'049	2'446	2'370	2'362
R	2'670	1'683	2'033	1'682	1'667
S	5'308	3'348	4'531	3'934	3'886
Taille					
3 à 9	22'374	13'462	18'216	15'606	15'385
10 à 49	11'950	7'885	9'628	9'133	9'091
50 à 249	9'460	7'568	7'679	7'548	7'536
250 à 499	1'044	835	903	891	890
500 à 999	415	332	369	365	364
1'000 et +	324	259	302	297	296

TABLE 5 Secteur privé : nombre d'unités secondaires (individus) dans l'échantillon brut et attendues/présentes dans l'échantillon net, par NOGA et grande région

Noga	betot brut	betot attendu net	net avant élim.	net après élim.
05-09	3'226	1'137	1'910	1'851
10-11	69'287	19'833	41'109	40'377
12	2'114	579	1'433	1'344
13-15	9'857	3'195	6'178	5'941
16-18	26'639	8'064	11'409	11'130
19-20	26'328	7'187	18'264	17'923
21	45'767	12'390	35'273	34'267
22-23	30'207	8'608	16'296	16'056
24-25	51'529	14'536	30'538	29'775
26	93'082	25'047	60'879	57'906
27	26'876	7'446	17'864	17'454
28	61'196	16'601	36'873	36'268
29-30	14'887	4'068	8'824	8'679
31-33	29'944	9'243	16'776	16'226
35	27'445	7'547	19'573	19'218
36-39	11'944	3'987	7'081	6'939
41-43	121'506	33'594	70'287	69'064
45-46	163'572	47'693	94'087	90'788
47	220'141	62'209	166'767	162'142
49-52	139'345	38'246	94'199	91'164
53	50'336	13'649	32'553	31'835
55-56	120'302	40'499	63'788	61'091
58-60	26'543	8'279	16'174	15'763
61	26'109	7'191	22'510	22'273
62-63	56'719	17'790	33'168	32'035
64, 66	130'324	36'994	88'805	86'195
65	50'622	13'873	35'526	35'067
68	41'496	17'388	19'836	19'066
69-71	124'381	39'209	70'343	68'053
72	17'162	5'176	11'491	11'058
73-75	35'521	15'081	22'572	21'528
77, 79-82	131'992	39'792	68'343	65'996
78	121'077	32'961	52'968	48'242
84	5'609	1'915	3'465	3'361
85	140'321	39'750	82'315	79'293
86-88	499'889	135'013	348'170	337'189
90-93	52'761	19'628	23'698	22'679
94-95	73'472	27'539	33'669	32'664
96	18'673	7'876	11'159	10'550
grande région				
Région lémanique	497'848	144'948	278'982	262'225
Espace Mittelland	594'421	176'030	399'564	389'114
Suisse du Nord-ouest	379'531	109'428	237'963	230'066
Zurich	748'428	210'269	491'859	480'970
Suisse orientale	318'449	96'359	188'571	183'824
Suisse centrale	257'929	81'473	145'004	141'021
Tessin	101'595	32'307	54'230	51'230
Suisse	2'898'201	850'813	1'796'173	1'738'450

TABLE 6 Secteur privé : nombre d'unités secondaires (individus) dans l'échantillon brut et attendues/présentes dans l'échantillon net, par Section NOGA et classe de taille

Sect. NOGA	betot brut	betot attendu net	net avant élim.	net après élim.
B	3'226	1'137	1'910	1'851
C	487'713	136'796	301'716	293'346
D	27'445	7'547	19'573	19'218
E	11'944	3'987	7'081	6'939
F	121'506	33'594	70'287	69'064
G	383'713	109'903	260'854	252'930
H	189'681	51'895	126'752	122'999
I	120'302	40'499	63'788	61'091
J	109'371	33'261	71'852	70'071
K	180'946	50'868	124'331	121'262
L	41'496	17'388	19'836	19'066
M	177'064	59'466	104'406	100'639
N	253'069	72'754	121'311	114'238
O	5'609	1'915	3'465	3'361
P	140'321	39'750	82'315	79'293
Q	499'889	135'013	348'170	337'189
R	52'761	19'628	23'698	22'679
S	92'145	35'415	44'828	43'214
Taille				
3 à 9	108'372	65'233	71'781	67'889
10 à 49	266'655	112'734	136'493	130'320
50 à 249	940'964	250'924	459'777	445'459
250 à 499	356'097	94'959	203'593	196'690
500 à 999	287'628	76'701	194'586	187'708
1'000 et +	938'485	250'263	729'943	710'384

TABLE 7 Administration publique : nombre d'unités primaires dans l'échantillon brut et attendues/présentes dans l'échantillon net, par grande région et classe de taille

grande région	Nb. brut	attendu net	net avant élim.	actif avant élim.	net après élim.
Région lémanique	54	44	52	50	50
Espace Mittelland	64	52	61	61	60
Suisse du Nord-ouest	33	27	31	31	31
Zurich	35	28	31	31	31
Suisse orientale	42	35	39	38	38
Suisse centrale	51	42	48	48	48
Tessin	31	25	25	25	25
Suisse	310	253	287	284	283
Taille					
3 à 19	97	78	87	85	84
20 à 49	75	60	70	70	70
50 ou plus	64	51	57	57	57
Profiling	47	38	46	45	45
Cantons et Conf.	27	27	27	27	27

TABLE 8 Administration publique : nombre d'unités secondaires (individus) dans l'échantillon brut et attendues/présentes dans l'échantillon net, par grande région et classe de taille

grande région	betot brut	betot attendu net	net	net
			avant élim.	après élim.
Région lémanique	65'433	21'076	67'404	64'232
Espace Mittelland	115'451	37'987	125'220	123'405
Suisse du Nord-ouest	40'283	13'355	33'472	33'040
Zurich	67'000	20'377	62'985	62'435
Suisse orientale	30'092	9'536	24'973	24'697
Suisse centrale	31'612	10'069	23'529	23'376
Tessin	15'320	4'906	4'725	4'603
Suisse	365'191	117'304	342'308	335'788
Taille				
3 à 19	847	678	886	848
20 à 49	2'417	967	1'526	1'468
50 ou plus	9'423	2'513	4'533	4'373
Profiling	65'315	17'417	50'353	49'275
Cantons et Conf.	287'189	95'730	285'010	279'824

5.2 Coefficients de variation d'estimateurs de la médiane et du total

Les tables 9, 10 et 11 reproduisent les CV visés et attendus publiés dans le rapport de [Qualité & Potterat \(2020\)](#). Le CV visé correspond à un objectif de précision minimale à atteindre et le plan de sondage est optimisé en fonction de celui-ci. Le CV attendu est calculé une fois le plan de sondage élaboré à l'aide des hypothèses décrites précédemment sur le taux de réponse et la proportion de salaires livrés (voir section 4.1). Les CV attendus du rapport de [Qualité & Potterat \(2020\)](#) tiennent compte des éventuelles densifications d'échantillons qui sont financées par les cantons qui souhaitent avoir un échantillon plus grand. Les CV attendus doivent obligatoirement être inférieurs ou égaux aux CV visés lors de la construction du plan de sondage.

Dans les tables 9, 10 et 11, les CV obtenus sont donnés dans les colonnes "CV noAVS" et "CV". La colonne "CV noAVS" donne le CV qui serait obtenu sans intégrer les variables de calage basées sur les revenus AVS, c'est à dire en effectuant un calage semblable à celui des enquêtes précédentes. La colonne "CV" donne le CV réellement obtenu. Il est ainsi remarquable que l'ajout des variables de calage basées sur les revenus AVS améliore sensiblement certains CV. De plus les CV obtenus sont parfois très inférieurs aux CV attendus, même si la comparaison est faite uniquement par rapport au "CV noAVS". Les deux dernières colonnes (% betot attendu et % `anzlohn` obtenu) visent à expliquer partiellement ce dernier point en illustrant la correction de population finie attendue dans le plan et celle observée dans l'échantillon:

- La colonne "% betot attendu" donne, en pour-cent, la proportion de salaires attendue dans l'échantillon net, par rapport au cadre tout entier : $\text{betot attendu net} / \text{betot(cadre entier)}$.
- La colonne "% `anzlohn` obtenu" est aussi un ratio exprimé en pour-cent. Au numérateur il y a le total du nombre d'employés livrés, validés, par les entreprises de l'échantillon net obtenu : net après élimination. Au dénominateur, au lieu de reprendre les betot sur le cadre tout entier, une estimation du total de la variable FB01 est faite à l'aide de l'échantillon net et des poids d'entreprises calés (construits dans la pondération) : $\text{FB01 estimé (cadre entier)}$.

Pour le secteur privé et l'administration publique, les CV sont globalement meilleurs qu'attendu. Cela est lié à la taille nette d'échantillon plus grande que prévue, surtout au niveau du nombre d'unités secondaires (salaires).

TABLE 9 Secteur privé : CV visés, attendus et obtenus pour l'estimateur du salaire médian par NOGA et grande région

Noga	CV visé	CV attendu	CV noAVS	CV	% betot attendu	% anzlohn obtenu
05-09	3.0	1.2	0.7	0.6	26%	46%
10-11	3.0	1.3	0.6	0.5	21%	46%
12	10.3	8.9	2.7	2.7	27%	68%
13-15	4.0	3.3	1.3	1.0	26%	53%
16-18	3.0	1.1	0.8	0.7	13%	20%
19-20	3.0	2.0	1.3	0.7	24%	62%
21	3.0	1.8	0.5	0.3	27%	76%
22-23	3.0	1.0	0.6	0.4	21%	43%
24-25	3.0	0.9	0.6	0.5	16%	34%
26	3.0	1.5	0.5	0.3	24%	61%
27	3.0	2.8	0.7	0.4	24%	60%
28	3.0	1.1	0.6	0.4	22%	48%
29-30	3.0	1.8	0.6	0.5	24%	57%
31-33	3.0	1.1	1.2	0.9	20%	36%
35	3.0	2.0	0.6	0.5	24%	67%
36-39	3.0	1.2	1.1	0.9	24%	46%
41-43	3.0	0.8	0.5	0.5	11%	23%
45-46	3.0	0.9	0.7	0.5	17%	34%
47	3.0	1.0	0.3	0.3	20%	56%
49-52	3.0	1.9	0.7	0.5	21%	55%
53	4.0	3.8	0.2	0.2	27%	72%
55-56	3.0	1.1	0.2	0.2	18%	33%
58-60	3.0	3.0	1.3	1.2	26%	55%
61	3.0	2.1	1.6	0.9	27%	82%
62-63	3.0	1.2	0.7	0.4	21%	38%
64, 66	3.0	1.8	1.0	0.6	22%	56%
65	3.0	2.4	0.4	0.3	26%	71%
68	4.0	2.6	0.4	0.4	35%	51%
69-71	3.0	1.0	0.6	0.5	16%	30%
72	3.0	2.2	0.8	0.6	26%	57%
73-75	4.0	3.2	0.7	0.5	36%	55%
77, 79-82	5.0	3.8	0.6	0.5	23%	42%
78	3.0	1.2	0.8	0.7	26%	48%
84	4.0	3.0	0.9	0.8	31%	61%
85	3.0	2.6	0.4	0.4	24%	56%
86-88	3.0	1.2	0.2	0.2	21%	56%
90-93	5.0	3.3	1.3	1.0	31%	41%
94-95	4.0	2.7	0.4	0.4	30%	46%
96	3.0	1.4	0.5	0.5	25%	41%
grande région						
Région lémanique	3.0	1.0	0.5	0.4	20%	41%
Espace Mittelland	3.0	0.8	0.2	0.2	21%	50%
Suisse du Nord-ouest	3.0	1.1	0.4	0.4	21%	47%
Zurich	3.0	0.9	0.3	0.3	21%	54%
Suisse orientale	3.0	1.0	0.4	0.4	19%	41%
Suisse centrale	3.0	1.0	0.4	0.4	22%	42%
Tessin	3.0	1.4	0.9	0.8	20%	36%
Suisse	-	0.27	0.14	0.11	21%	47%

TABLE 10 Secteur privé : CV visés, attendus et obtenus pour l'estimateur du total des salaires par Section NOGA et classe de taille

Section NOGA	CV visé	CV attendu	CV noAVS	CV CV	% betot attendu	% anzlohn obtenu
B	3.0	2.8	2.3	1.8	26%	46%
C	3.0	0.7	0.3	0.3	21%	48%
D	3.0	2.0	1.2	0.9	24%	67%
E	3.0	2.1	1.8	1.5	24%	46%
F	3.0	1.4	1.3	1.2	11%	23%
G	3.0	1.2	0.8	0.5	18%	45%
H	3.0	2.4	1.2	0.5	23%	59%
I	3.0	1.7	1.0	0.9	18%	33%
J	3.0	1.9	2.1	1.8	23%	50%
K	3.0	1.8	0.9	0.6	23%	59%
L	3.0	1.6	1.0	0.8	35%	51%
M	3.0	1.5	0.9	0.6	19%	36%
N	3.0	1.4	1.3	1.0	24%	44%
O	4.0	3.7	1.4	1.2	31%	61%
P	3.0	2.3	1.6	1.3	24%	56%
Q	3.0	1.5	0.8	0.8	21%	56%
R	3.0	2.4	3.0	1.8	31%	41%
S	3.0	1.2	0.9	0.6	29%	45%
Taille						
3 à 9	3.0	1.6	1.7	1.2	11%	13%
10 à 49	3.0	1.3	1.0	0.7	12%	15%
50 à 249	3.0	0.7	0.4	0.2	27%	50%
250 à 499	3.0	1.9	1.4	0.4	27%	61%
500 à 999	3.0	3.0	1.9	0.6	27%	69%
1'000 et +	3.0	1.5	0.8	0.2	27%	81%

TABLE 11 Administration publique : CV visés, attendus et obtenus pour l'estimateur du salaire médian par grande région

grande région	CV visé	CV attendu	CV noAVS	CV CV	% betot attendu	% anzlohn obtenu
Région lémanique	3.0	2.4	0.2	0.2	27%	81%
Espace Mittelland	3.0	2.4	0.5	0.5	29%	86%
Suisse du Nord-ouest	3.0	2.8	0.7	0.7	27%	67%
Zurich	3.0	2.8	0.5	0.5	27%	86%
Suisse orientale	3.0	3.0	0.9	0.9	22%	64%
Suisse centrale	3.0	3.0	0.8	0.8	27%	69%
Tessin	3.0	3.0	1.0	1.0	28%	50%

6 Conclusions

La révision de la pondération de l'enquête sur la structure des salaires (ESS) permet d'améliorer la précision des résultats publiés par l'OFS. Cela est dû notamment à une modélisation soigneuse de la réponse basée sur l'utilisation de nouvelles variables auxiliaires. Ceci permet d'obtenir de meilleurs poids avant calage, ce qui réduit le risque de biais et facilite le calage. De plus, l'utilisation d'information auxiliaire, notamment les revenus AVS, apporte à travers le calage des gains de précision pour estimer des salaires médians ou des totaux dans de nombreux domaines. Par exemple, le calage utilisant cette information auxiliaire améliore en moyenne les CV de l'estimateur du salaire médian d'environ 20% pour les salaires médians par NOGA et 10% pour les salaires médians par Grande Région. Ces améliorations augmentent la qualité de la procédure de pondération. Elles peuvent également servir à diminuer la charge, notamment sur les petites entreprises, tout en maintenant la précision des résultats. Pour les enquêtes futures, le secteur privé et l'administration publique seront intégrés dans une unique pondération, avec néanmoins toujours des variables de calage spécifiques au public et au privé. Un autre travail en cours est l'utilisation des formules de ce rapport pour les calculs de CV (notamment, l'estimation de variance à l'aide d'une linéarisée) pour les publications de l'OFS. Outre ces projets déjà à l'étude, les améliorations futures peuvent porter sur une modélisation, plutôt au moment d'effectuer le plan de sondage, du fait que certaines grandes unités livrent la totalité de leurs salaires et non un tiers. Ce point conduit parfois en effet parfois à des précisions bien meilleures qu'anticipé. Les travaux pourraient aussi porter sur l'ajout de davantage d'informations auxiliaires, ainsi que, à plus long terme, sur un calage qui s'effectuerait directement sur les poids des unités secondaires.

Bibliographie

- DEVILLE, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology* **25**, 193–204.
- DEVILLE, J.-C. & SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.
- KIM, J. K. & KIM, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics* **35**, 501–514.
- OSIER, G. (2009). Variance estimation for complex indicators of poverty and inequality using linearization techniques. *Survey Research Methods* **3**, 167–195.
- QUALITÉ, L. & POTTERAT, J. (2020). Rapport de méthodes : Révision du plan de sondage pour l'enquête suisse sur la structure des salaires 2018. Rapport de méthodes, Office Fédéral de la statistique OFS.
- SAUTORY, O. & LE GUENNEC, J. (2003). La macro CALMAR2 : redressement d'un échantillon par calage sur marges – documentation de l'utilisateur. Tech. rep., Insee, Paris.
- TILLÉ, Y. (2019). *Théorie des sondages: Échantillonnage et estimation en populations finies*. 2e Édition. Dunod.

Annexes

A Unités et nombre d'employés dans l'échantillon net du secteur privé par croisement NOGA x GR

TABLE 12 Nombre d'unités primaires attendues (att.) et présentes (réa.) dans l'échantillon net, par croisement NOGA × GR

unités Noga	GR1		GR2		GR3		GR4		GR5		GR6		GR7	
	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.
05-09	11	12	20	23	11	7	4	3	17	13	13	14	1	1
10-11	88	82	134	121	47	50	66	59	110	106	68	52	43	46
12	1	1	2	2	2	3	1	1	-	-	2	2	1	1
13-15	22	24	39	34	28	26	31	34	39	47	36	39	6	6
16-18	58	60	102	114	39	34	55	57	48	53	70	70	27	24
19-20	17	17	23	23	16	16	23	23	24	30	20	22	5	5
21	18	20	18	22	27	28	8	10	15	17	10	11	14	12
22-23	26	25	58	62	44	45	28	30	56	58	29	27	26	30
24-25	62	65	142	154	58	60	50	60	78	81	56	63	32	29
26	50	45	127	134	30	32	40	43	37	41	21	24	16	21
27	20	21	30	31	23	27	25	26	19	22	16	18	6	8
28	17	16	86	90	44	43	47	55	70	65	49	58	11	14
29-30	6	6	13	13	6	5	8	10	10	9	10	12	5	4
31-33	61	71	92	106	61	64	83	101	64	73	78	90	14	16
35	27	31	22	26	15	16	10	13	38	39	17	22	8	7
36-39	35	39	61	66	31	35	26	26	50	53	39	32	17	19
41-43	155	134	215	226	130	124	167	174	143	154	140	141	51	51
45-46	448	508	301	354	160	173	269	292	208	231	327	396	84	108
47	197	207	424	469	79	78	154	157	207	244	163	180	97	121
49-52	110	111	120	126	92	87	91	82	106	102	82	78	52	55
53	15	16	10	8	6	8	13	13	10	8	15	13	4	6
55-56	306	306	672	709	530	540	164	163	428	441	377	400	151	177
58-60	96	116	92	90	41	41	52	56	48	47	41	47	24	30
61	16	18	14	17	8	12	14	16	7	6	15	15	8	11
62-63	162	195	139	159	80	101	240	281	94	124	205	253	50	60
64, 66	145	169	116	133	50	60	139	162	80	92	267	298	63	82
65	20	25	27	30	16	20	43	51	18	23	18	20	5	7
68	223	255	328	352	315	299	412	461	378	411	313	321	120	147
69-71	319	371	388	475	155	189	239	284	280	351	548	684	220	275
72	60	66	30	38	17	18	45	52	16	16	26	32	19	26
73-75	249	302	280	349	158	198	615	787	171	213	167	206	76	104
77, 79-82	254	248	312	344	178	185	225	237	150	173	260	300	208	256
78	89	71	73	54	84	62	108	96	56	44	73	58	19	15
84	25	22	26	34	13	15	14	13	30	35	15	19	4	6
85	152	150	245	285	86	85	201	204	191	214	138	141	47	57
86-88	374	407	496	579	236	275	310	352	286	335	253	306	95	108
90-93	400	382	383	379	228	226	178	174	245	245	151	159	99	102
94-95	333	381	691	812	259	283	375	429	372	391	191	218	71	83
96	131	163	237	305	77	91	181	212	175	212	166	211	89	95

TABLE 13 Nombre (en milliers) d'unités secondaires (employés) attendues (att.) et présentes (réa.) dans l'échantillon net, par croisement NOGA × GR

empl. Noga	GR1		GR2		GR3		GR4		GR5		GR6		GR7	
	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.
05-09	0.2	0.2	0.3	0.6	0.2	0.2	0.1	0.3	0.2	0.3	0.2	0.2	0.0	0.1
10-11	3.1	3.8	4.1	9.3	2.8	6.8	3.4	9.0	3.2	5.9	2.6	4.7	0.6	0.9
12	0.0	0.0	0.4	1.1	0.0	0.0	0.0	0.0	-	-	0.1	0.2	0.0	0.0
13-15	0.2	0.2	0.6	0.9	0.3	0.4	0.4	0.6	0.9	2.4	0.3	0.3	0.5	1.3
16-18	0.7	0.8	1.8	2.7	1.3	2.0	1.0	1.2	1.5	2.3	1.4	1.9	0.3	0.3
19-20	2.6	6.7	0.9	1.8	1.4	4.3	0.6	0.8	1.1	2.9	0.5	1.3	0.1	0.2
21	1.1	2.1	1.2	1.9	7.8	25.0	0.1	0.2	1.0	2.7	0.7	1.5	0.6	0.9
22-23	0.6	1.0	1.8	3.1	1.6	3.9	0.6	0.9	2.8	5.5	0.8	1.0	0.4	0.6
24-25	1.2	1.8	4.2	8.5	1.8	3.5	1.6	3.6	3.3	7.9	1.6	3.0	0.9	1.5
26	4.0	9.6	10.6	21.9	2.6	5.1	4.0	11.4	2.1	5.4	0.9	2.3	0.9	2.2
27	0.5	1.2	1.0	1.8	2.2	6.7	0.8	1.8	1.3	3.4	1.3	1.6	0.4	0.9
28	1.1	1.7	4.5	10.4	2.0	4.5	2.6	6.1	3.7	7.9	2.4	4.6	0.5	0.9
29-30	0.3	0.6	1.3	3.5	0.1	0.1	0.5	1.1	1.1	2.4	0.6	0.9	0.1	0.1
31-33	0.9	1.5	2.0	4.9	1.4	1.7	1.9	2.4	1.2	2.3	1.4	2.5	0.5	1.0
35	1.5	4.5	1.9	5.7	1.3	1.9	1.0	3.3	0.9	1.8	0.6	1.4	0.4	0.7
36-39	0.7	1.2	0.8	1.0	0.6	1.2	0.7	2.0	0.6	0.6	0.5	0.7	0.2	0.2
41-43	4.8	8.0	6.8	14.6	4.2	7.3	8.6	22.6	3.9	6.5	4.1	8.0	1.3	2.0
45-46	8.7	11.9	7.9	18.9	6.2	11.6	12.2	24.8	4.2	7.3	6.6	13.7	1.9	2.6
47	7.0	17.2	9.4	22.6	7.5	12.9	25.8	79.8	6.4	18.0	4.4	8.5	1.7	3.1
49-52	5.1	9.6	12.5	40.5	4.5	8.5	8.2	18.5	4.6	9.7	2.5	3.2	0.8	1.2
53	0.3	0.4	8.5	27.5	0.9	2.1	2.7	1.4	1.0	0.3	0.1	0.1	0.0	0.1
55-56	6.2	9.0	7.6	10.2	6.1	7.4	8.9	20.5	5.3	5.3	4.8	6.8	1.6	2.0
58-60	1.1	1.3	1.1	1.4	0.7	0.9	3.9	10.6	0.7	0.6	0.5	0.4	0.3	0.4
61	0.5	1.5	5.1	18.1	0.1	0.2	1.3	2.3	0.1	0.0	0.1	0.1	0.1	0.1
62-63	2.8	4.6	2.4	4.6	1.4	2.5	6.7	13.3	1.4	2.8	2.3	3.3	0.8	0.9
64, 66	7.3	14.1	4.1	11.4	2.6	6.7	15.1	38.0	2.8	7.7	3.2	5.7	1.9	2.6
65	1.2	3.4	1.2	3.1	2.3	7.4	6.4	14.7	0.3	0.9	2.4	5.6	0.1	0.1
68	3.1	2.9	2.7	3.1	2.1	2.1	4.4	6.2	2.3	2.3	2.0	1.8	0.8	0.7
69-71	6.7	9.5	5.3	8.8	3.6	7.8	13.2	27.1	3.2	5.0	5.2	7.0	2.0	2.9
72	0.8	1.4	0.5	1.2	1.9	4.8	1.0	2.1	0.3	0.4	0.5	0.9	0.2	0.4
73-75	2.0	2.5	2.2	2.8	1.3	1.4	6.0	10.3	1.4	1.9	1.8	2.2	0.5	0.6
77, 79-82	8.7	9.5	6.4	10.5	3.4	4.5	13.5	30.9	2.6	3.6	3.4	4.9	1.8	2.1
78	11.5	19.5	5.7	10.6	3.9	4.6	4.7	6.2	2.7	3.2	2.6	2.4	1.8	1.7
84	0.6	1.0	0.5	1.1	0.2	0.3	0.1	0.2	0.3	0.4	0.2	0.3	0.0	0.0
85	9.1	23.7	5.1	8.6	4.4	10.3	12.5	26.4	5.1	5.9	2.4	2.8	1.2	1.7
86-88	29.2	62.1	29.7	73.3	18.2	51.3	25.4	67.4	16.0	41.7	11.6	29.4	4.9	11.9
90-93	4.1	5.0	4.2	4.7	2.7	3.1	3.7	5.0	2.2	1.8	1.7	1.9	1.0	1.1
94-95	4.3	5.6	8.2	10.4	3.1	3.5	5.1	6.1	3.6	3.4	2.4	3.0	0.8	0.8
96	1.2	1.5	1.5	2.3	0.9	1.8	1.7	2.1	1.1	1.3	0.9	1.0	0.5	0.5

B Coefficient de variation de l'estimateur du salaire médian dans le secteur privé par croisement NOGA x GR

TABLE 14 CV attendu (att.) et réalisé (réa.) de l'estimateur du salaire médian par croisement NOGA × GR

CV Noga	GR1		GR2		GR3		GR4		GR5		GR6		GR7	
	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.	att.	réa.
05-09	3.2	2.8	1.7	0.8	3.5	1.7	3.0	0.7	2.2	2.2	2.5	1.5	3.3	0.1
10-11	6.9	2.1	3.3	1.0	3.5	1.8	3.3	0.9	2.4	1.2	4.0	1.4	2.9	1.6
12	18.5	1.0	10.0	1.2	3.0	49.6	0.0	0.0	-	-	5.1	2.0	12.9	5.3
13-15	3.4	2.9	3.6	2.7	3.9	3.4	3.3	2.0	3.7	1.7	6.0	2.4	3.3	1.6
16-18	2.4	2.0	2.7	1.1	2.8	2.2	2.8	1.5	2.6	2.6	2.3	1.8	3.1	2.1
19-20	1.9	1.8	2.8	1.4	2.8	1.7	3.0	4.0	2.3	2.1	2.7	1.7	7.0	6.2
21	1.8	1.5	2.2	1.3	1.6	0.7	5.0	4.5	5.0	2.1	9.7	5.2	2.3	2.2
22-23	2.7	2.3	1.7	1.3	2.8	1.3	2.5	1.3	2.0	0.7	2.2	1.0	3.1	3.8
24-25	2.7	1.6	1.5	0.9	2.4	2.1	2.8	1.6	1.9	0.8	2.5	2.1	3.0	2.3
26	2.7	0.9	1.6	0.7	4.5	2.2	2.8	0.8	3.6	1.2	3.6	1.6	1.9	1.7
27	3.8	2.9	2.6	1.6	4.0	0.9	2.7	2.1	3.7	1.1	3.4	1.5	4.4	3.1
28	4.3	2.4	1.3	1.0	2.8	0.9	2.4	0.9	2.7	0.7	3.1	1.2	4.9	2.4
29-30	3.9	3.2	3.9	1.0	3.8	5.9	4.7	2.0	2.4	0.6	2.7	2.0	5.0	2.8
31-33	2.1	1.5	2.7	1.3	2.8	2.7	2.8	1.9	2.5	1.2	2.6	2.1	4.9	3.3
35	5.1	1.5	3.0	0.6	5.1	1.2	4.9	1.1	2.9	1.4	4.6	1.2	2.9	1.6
36-39	2.3	2.5	2.8	2.3	3.4	2.9	3.1	2.4	2.8	2.2	3.8	5.1	4.8	5.2
41-43	1.9	1.5	1.5	1.0	2.2	1.4	1.9	0.8	2.2	1.2	2.2	1.5	3.2	2.6
45-46	2.4	2.4	2.0	1.2	2.6	1.8	2.0	1.3	2.7	1.7	2.6	1.6	2.8	1.8
47	2.5	0.9	2.6	0.7	3.2	0.8	1.5	0.4	3.7	0.9	3.9	1.3	2.8	1.2
49-52	3.0	2.3	3.3	0.9	3.2	2.3	2.8	1.4	3.8	1.7	3.3	2.1	4.1	5.0
53	14.1	3.2	1.0	0.0	24.4	4.3	3.4	1.7	8.4	4.2	18.5	2.5	21.7	3.5
55-56	1.7	0.7	2.6	0.5	6.0	0.6	2.4	0.5	1.7	0.7	2.2	0.5	1.8	0.9
58-60	8.6	3.6	6.5	2.3	8.8	2.4	3.1	1.3	9.5	2.1	8.0	3.5	8.1	3.4
61	5.8	6.9	2.0	1.9	5.6	3.7	4.6	6.3	7.2	7.7	7.1	5.4	7.8	5.2
62-63	2.1	1.3	2.6	1.8	2.8	1.7	2.4	1.2	2.9	1.5	2.2	1.4	10.2	6.9
64, 66	3.6	2.7	3.1	1.1	3.8	2.5	1.8	0.9	2.7	1.9	2.5	1.3	5.0	2.4
65	5.6	1.4	8.0	1.2	2.6	0.5	4.1	0.5	8.2	1.5	3.4	0.7	11.6	2.7
68	3.9	1.0	8.3	0.8	7.7	1.3	4.6	0.6	6.4	1.2	6.4	1.2	7.5	1.4
69-71	2.3	1.5	2.5	1.1	2.8	1.4	2.2	1.3	2.6	1.2	2.0	1.0	2.8	1.5
72	2.9	2.0	3.1	2.4	4.0	1.5	3.3	1.7	4.9	1.9	7.9	3.5	5.8	3.3
73-75	4.9	2.0	3.3	1.6	5.4	1.6	10.6	1.0	4.3	1.4	11.4	2.0	4.4	3.2
77, 79-82	4.4	1.0	12.9	1.2	6.2	2.1	7.5	1.0	4.6	1.7	9.6	2.2	7.5	1.5
78	2.5	1.6	1.3	0.9	2.4	1.6	4.4	1.4	2.0	1.9	2.7	2.2	5.6	5.1
84	4.3	0.9	7.3	2.2	11.1	2.4	13.6	2.1	10.6	1.2	9.8	1.4	17.8	2.9
85	4.7	0.7	6.5	1.2	7.8	1.7	4.4	0.5	3.6	1.4	6.5	1.4	7.0	2.9
86-88	2.8	0.5	2.2	0.5	2.8	0.7	2.3	0.6	2.7	0.7	2.8	0.5	3.1	1.0
90-93	6.3	2.5	5.6	1.5	7.7	1.5	9.2	1.9	7.0	1.6	6.5	1.8	10.8	3.3
94-95	4.6	1.2	6.8	0.6	6.1	1.0	5.1	1.0	4.6	1.1	6.8	1.2	8.9	2.1
96	2.2	1.3	3.6	1.0	4.6	1.6	2.9	1.3	4.0	1.0	3.7	1.1	3.9	1.6

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Potterat, J. (2012). Use of conversion keys for NOGA2002-2008. Order 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 de commande: 338-0066

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

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

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

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

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

- Eichenberger P., Hulliger B., Potterat J. (2010). Describing the Anticipated Accuracy of the Swiss Population Survey. Order 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 de commande: 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. Order 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 de commande: 338-0057-05
- Qualité, L., Tillé, Y. (2009). Estimation de la précision d'évolutions dans l'enquête sur la valeur ajoutée. Numéro de commande: 338-0056
- Renaud, A., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro de commande: 338-0055
- Graf, E. (2008). Pondérations du PSM. PSM_I vague 8, PSM_II vague 3, PSM_I et PSM_II combinés. Numéro de commande: 338-0054
- Andrade, B., Graf, M. (2008). Enquête suisse sur la structure des salaires 2006. Aspects méthodologiques du modèle des salaires "Salarium". Numéro de commande: 338-0053
- Renaud, A. (2008). Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage. Numéro de commande: 338-0052
- Graf, E. (2008). Pondérations du SILC pilote. SILC_I vague 2, SILC_II vague 1, SILC_I et SILC_II combinés. Numéro de commande: 338-0051
- Kilchmann, D. (2008). Statistik der sozialmedizinischen Institutionen 1999-2004 und Krankenhausstatistik 1999-2002. Einsetzungen für fehlende Daten. Bestellnummer: 338-0050
- Renaud, A. (2008). Technologies de l'information et de la communication. Estimations sur la base de la statistique de la valeur ajoutée. Numéro de commande: 338-0049
- Assoulin, D. (2007). Wertschöpfungsstatistik. Einsetzungsversuche für fehlende Antworten grosser Unternehmen. Bestellnummer: 338-0048
- Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. Bestellnummer: 338-0047
- Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. Bestellnummer: 338-0046
- Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro de commande: 338-0045
- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro de commande: 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. Bestellnummer: 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro de commande: 338-0041
- Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro de commande: 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro de commande: 338-0039
- Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample: Analysis and precision assessment of wage components. Order number: 338-0038

- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. Bestellnummer: 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. Bestellnummer: 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. Bestellnummer: 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Bestellnummer: 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. Bestellnummer: 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. Bestellnummer: 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro de commande: 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro de commande: 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. Bestellnummer: 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. Bestellnummer: 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. Order number: 338-0027
- Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. Bestellnummer: 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro de commande: 338-0025
- Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro de commande 338-0024
- Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro de commande: 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro de commande: 338-0022
- Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. Bestellnummer: 338-0021
- Potterat, J. (2003). Landwirtschaftliche Betriebszählung 2003. Stichprobenplan der Zusatzerhebung. Bestellnummer: 338-0020.
- Renaud, A. (2003). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sur-couverture (E-sample). Numéro de commande: 338-0019
- Hulliger, B. (2003). Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. Erhebung über F+E in der schweizerischen Privatwirtschaft 2000. Bestellnummer: 338-0018
- Renfer, J.-P. (2003). Enquête 2000 sur la recherche et le développement dans l'économie privée en Suisse. Plan d'échantillonnage. Numéro de commande: 338-0017
- Potterat, J. (2003). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe. Schätzverfahren. Bestellnummer: 338-0016
- Graf, M., Matei, A. (2003). Stratégie de choix des modèles de désaisonnalisation. Application aux séries de l'emploi total. Numéro de commande: 338-0015

- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Methoden für die Datenbereinigung. Bestellnummer: 338-0014
- Renaud, A. (2002). Programme international pour le suivi des acquis des é lèves (PISA). Plans d'échantillonnage pour PISA 2000 en Suisse. Numéro de commande: 338-0013
- Renfer, J.-P. (2002). Enquête 2001 sur les coûts et l'utilité de la formation des apprentis du point de vue des établissements. Plan d' échantillonnage. Numéro de commande: 338-0012
- Potterat, J., Salamin, P.A. (2002). Betriebszählung 2001. Stichprobenplan und Schätzverfahren für die provisorischen Ergebnisse. Bestellnummer: 338-0011
- Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. Numéro de commande: 338-0010
- Renaud, A., Eichenberger P. (2002). Estimation de la couverture du recensement de la population de l'an 2000. Procédure d'enquête et plan d'échantillonnage de l'enquête de couverture. Numéro de commande: 338-0009
- Kilchmann, D., Hulliger, B. (2002). Stichprobenplan für die Obstbaumzählung 2001. Bestellnummer: 338-0008
- Graf, M. (2002). Passage du concept établissement au concept entreprise. Numéro de commande: 338-0007
- Salamin, P.A. (2001). La technique de la double enquête pour la statistique du transport routier de marchandise. Numéro de commande: 338-0006
- Peters, R., Renfer, J.-P. et Hulliger, B. (2001). Statistique de la valeur ajoutée 1997-1998. Procédure d'extrapolation des données. Numéro de commande: 338-0005
- Potterat, J., Hulliger, B. (2001). Schätzung der Sägereiproduktion mit der Sägerei-Erhebung PAUL. Bestellnummer: 338-0004
- Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande: 338-0003
- Hüsler, J., Müller, S. (2001). Schlussbericht Betriebszählung 1995 (BZ 95), Mehrfach imputierte Umsatzzahlen. Bestellnummer: 338-0002
- Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande: 338-0001
- Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen: Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer: 338-0000

Programme des publications de l'OFS

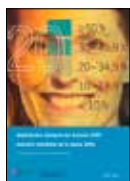
En tant que service statistique central de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public. Il utilise plusieurs moyens et canaux pour diffuser ses informations statistiques par thème.

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 (058 463 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

058 463 60 11, info@bfs.admin.ch

La méthode de pondération de l'ESS a fait l'objet d'une révision approfondie pour l'enquête de l'année 2018. Cette révision s'inscrit dans la continuité des travaux effectués sur le plan de sondage présentés dans *Qualité & Potterat* (2020). L'objectif est d'utiliser les données administratives disponibles sur les revenus des personnes salariées pour améliorer la qualité des résultats. Les poids des unités primaires (entreprises privées ou publiques, administrations publiques) répondant à l'enquête sont obtenus à l'aide du plan de sondage, mais aussi via une modélisation du mécanisme de réponse à l'enquête et un calage des poids. Pour ces deux dernières opérations, les données administratives sur les revenus des personnes ont permis, comme pour le plan de sondage, d'améliorer la précision de l'enquête. Les poids des unités primaires et les poids des salaires des personnes employées dans ces unités sont liés par une relation simple. Au final, un peu plus de 2 millions de salaires relevés permettent, une fois pondérés, de produire les résultats de l'enquête. Les analyses montrent qu'un calage des poids utilisant les données administratives sur les revenus des personnes améliore le coefficient de variation de l'estimateur du salaire médian d'environ 10% à 20% selon le domaine considéré (NOGA, Grande Région).

En ligne

www.statistique.ch

Imprimés

www.statistique.ch

Office fédéral de la statistique

CH-2010 Neuchâtel

order@bfs.admin.ch

tél. 058 463 60 60

Numéro OFS

338-0082

ISBN

978-3-303-00682-5

La statistique
compte pour vous.

www.la-statistique-compte.ch