

2021

00

Bases
statistiques
et généralités

Neuchâtel 2025

Statistique de la valeur ajoutée 2021

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.

Statistique de la valeur ajoutée 2021

Révision de la pondération

Rédaction Clément Chevalier, OFS

Éditeur Office fédéral de la statistique (OFS)

Neuchâtel 2025

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Clément Chevalier, tél. +41 58 481 00 19
clement.chevalier@bfs.admin.ch

Rédaction: Clément Chevalier, OFS

Contenu: section méthodes statistiques, OFS

Série: Statistique de la Suisse

Domaine: 00 Bases statistiques et généralités

Langue du texte original: français

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

En ligne: www.statistique.ch

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

Copyright: OFS, Neuchâtel 2025
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.

Numéro OFS: 338-0092

ISBN: 978-3-303-00769-3

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Variables importantes et rappels sur le plan de sondage	7
2.1 NOGA, grandes régions, nombre d'employés et classes de taille	7
2.2 Chiffre d'affaires TVA, productivité TVA et données AVS	8
2.3 Plan de sondage et VIU	9
3 Non-réponse, robustification et calage	10
3.1 Vue d'ensemble de la pondération	10
3.2 Modèle de réponse	11
3.3 Robustification	14
3.3.1 Motivations et calculs de variance	14
3.3.2 Algorithme de robustification	15
3.3.3 Variables de l'enquête considérées par l'algorithme	16
3.3.4 Robustification des unités qui changent de NOGA	18
3.3.5 Impact de la robustification sur les estimations	18
3.4 Calage	19
3.4.1 Motivations du calage	19
3.4.2 Variables de calage	20
3.4.3 Méthode de calage	21
4 Résultats	22
4.1 Taille de l'échantillon net	22
4.2 Interprétations sur le modèle de réponse	25
4.2.1 Statistiques descriptives et odds ratios	25
4.2.2 Interprétations relatives à la classe de taille	27
4.2.3 Interprétations relatives aux revenus AVS	28
4.2.4 Interprétations relatives à la productivité TVA	30
4.2.5 Interprétations relatives aux grandes régions	31
4.3 Coefficients de variation d'estimateur du total	32
5 Conclusions et perspectives	38
Bibliographie	39
Annexes	40
A Coefficients de variation pour la valeur ajoutée t029	40
B Impact de la robustification sur l'estimation des totaux	42
Rapports de méthodes de la section méthodes statistiques de l'OFS	43

Préambule

La statistique de la valeur ajoutée (en allemand, Wertschöpfungsstatistik, abrégé "WS") est une enquête par échantillonnage aléatoire réalisée chaque année par l'Office Fédéral de la Statistique (OFS). L'enquête relève des informations sur les entreprises telles que le chiffre d'affaires (t010), la production brute (t017), les consommations intermédiaires (t022), la masse salariale (t053) ou la valeur ajoutée (t029). Ce dernier indicateur est particulièrement important dans la mesure où son total est lié au calcul du produit intérieur brut. La section des méthodes statistiques (METH) de l'OFS a été sollicitée pour réviser le plan de sondage de la WS 2021. La suite naturelle de ce travail est une révision de la procédure de pondération, c'est-à-dire la procédure qui attribue un poids à chaque unité (entreprise) répondante. Les poids d'extrapolation sont utilisés pour l'estimation de totaux sur la Suisse ou sur d'autres domaines, comme par exemple le total de la valeur ajoutée par domaines d'activité économique. La nouvelle méthodologie optimise l'utilisation de données administratives issues de décomptes TVA, mais aussi de l'AVS. Ces informations auxiliaires sont utilisées à plusieurs étapes importantes de la pondération, telles que la modélisation de la non-réponse, la robustification ou le calage. Il en résulte une amélioration de la qualité des résultats produits. L'auteur remercie les membres de la section METH pour la relecture attentive de ce rapport, Daniel Assoulin, Blandine Bianchi, Daniel Kilchmann, Lionel Qualité et Jean-Pierre Renfer. L'auteur remercie également les collaborateurs de la section Statistique monétaire des entreprises (MON) qui ont relu le rapport et travaillé avec METH à l'amélioration des méthodes, Riccardo Dautaj, Massimiliano Ferrari, Gérard de Montmollin et Veronica Roth.

Résumé

La pondération de la statistique de la valeur ajoutée (WS) a été révisée en profondeur pour l'enquête de l'année 2021. Les améliorations apportées proviennent de l'utilisation de nouvelles données administratives issues de décomptes TVA des entreprises et de cotisations à l'AVS des personnes travaillant en Suisse. Chaque étape de la procédure (modélisation de la non-réponse, robustification, calage) tire avantage de la disponibilité de ces données. Des gains de précision et donc de qualité sont obtenus pour l'estimation de totaux sur des domaines tels que le domaine d'activité économique NOGA. Ces gains sont détaillés et analysés dans ce rapport. Ils proviennent en grande partie du calage du chiffre d'affaires TVA. Par ailleurs, de nouvelles interprétations sont faites concernant la non-réponse à l'enquête. S'il avait déjà été constaté, pour d'autres enquêtes de l'OFS, que les entreprises qui tendent à mieux rémunérer leurs salariés répondent plus souvent, force est également de constater que les entreprises plus productives ont tendance à plus facilement répondre à l'enquête.

1 Introduction

La statistique de la valeur ajoutée (WS) est une enquête annuelle par échantillonnage aléatoire où des unités (entreprises) renseignent des valeurs monétaires telles que leur chiffre d'affaires (t_{010}), leur production brute (t_{017}) ou leur consommation intermédiaire (t_{022}). La différence $t_{029} = t_{017} - t_{022}$ est une variable nommée "valeur ajoutée" qui donne son nom à l'enquête. La WS est une enquête réalisée par l'Office Fédéral de la Statistique (OFS) et coordonnée par la section Statistique monétaire des entreprises (MON). Elle a pour objectif d'observer la structure et l'évolution des activités des entreprises suisses et en particulier d'estimer le total de la valeur ajoutée, qui intervient dans le calcul du produit intérieur brut. Les estimations se basent sur un échantillon aléatoire d'entreprises. Le présent rapport se focalise sur la méthodologie de pondération des données de l'échantillon net (échantillon des répondants) de la WS 2021. Ce processus se décompose en trois étapes clés : le traitement de la non-réponse, la robustification des poids et le calage sur des variables auxiliaires. Pour chacune de ces étapes, la disponibilité de nouvelles informations auxiliaires permet des gains de précision importants sur les estimations.

Ce rapport est organisé comme suit. La section 2 décrit des variables importantes utilisées dans ce rapport et donne des rappels sur la construction du plan de sondage qui sont utiles pour comprendre la méthodologie de pondération. La section 3 décrit la nouvelle méthodologie de modélisation de la réponse à l'enquête puis la nouvelle procédure de robustification et de calage. Des analyses détaillées sur le modèle de réponse et sur les améliorations apportées sont données en section 4. Les résultats montrent des gains de précision substantiels, notamment lorsque les estimations sont réalisées sur des domaines tels que le secteur d'activité économique NOGA.

2 Variables importantes et rappels sur le plan de sondage

2.1 NOGA, grandes régions, nombre d'employés et classes de taille

Le domaine d'activité d'une unité est évalué à l'aide de la nomenclature générale des activités économiques : NOGA [OFS, 2008]. La NOGA est une variable définie sur l'ensemble du cadre WS, avec un niveau de granularité maximal de 6 chiffres. Le degré de granularité le plus souvent utilisé dans ce rapport est la NOGA à deux chiffres, notée NOGA2. En prenant en compte le regroupement effectué entre la NOGA2 08 (industries extractives dans d'autres domaines que les minerais métalliques, les hydrocarbures, la houille et la lignite) et la NOGA2 09 (services de soutien aux industries extractives) et le fait que certaines NOGA2 sont exclues de l'enquête (voir Chevalier [2024], section 2.4), il y a 73 NOGA2 différentes représentées dans le cadre WS. Une description des différentes NOGA2 est donnée dans Chevalier [2024], en annexe A. La NOGA est une variable importante utilisée à chaque étape de la pondération.

Une révision des NOGA2 des unités de l'échantillon net est effectuée par la section MON et il est possible en pratique d'avoir certaines unités qui "changent" de NOGA2 entre l'élaboration du plan de sondage et la pondération de l'enquête. Concrètement, le cadre de sondage de la WS contient une variable "NOGA" utilisée pour le plan de sondage et pour la grande majorité des étapes de la pondération (en particulier : modèle de réponse et calage). Une fois le cadre de sondage établi, la variable NOGA n'est jamais modifiée ou écrasée. Au moment de la pondération, une autre variable est créée, notée ici "NOGA2 MON". Il s'agit de la NOGA2 réévaluée suite à des analyses effectuées par la section MON. La NOGA2 MON est disponible uniquement sur l'échantillon net, c'est-à-dire sur l'échantillon d'entreprises qui répondent à l'enquête. Pour la grande majorité des

entreprises, la NOGA2 MON est égale à la NOGA2 initiale : pour la WS 2021, les deux variables sont différentes pour 219 unités sur un total de 15 186 unités dans l'échantillon net, soit environ 1.4% des unités de l'échantillon net. Si les poids d'extrapolations (tels que calculés selon les procédures décrites dans tout ce rapport) sont pris en compte, ces unités représentent 1.2% des employés en Suisse, 2.1% du chiffre d'affaires TVA (tel que défini en section 2.2) et 2.7% de la valeur ajoutée. La NOGA2 MON est utilisée pour les publications de l'OFS au niveau NOGA2. Dans la pondération, elle intervient notamment lors de l'étape de robustification, voir section 3.3. Un point important est que la NOGA2 MON n'intervient pas dans le calage puisqu'aucun total pour cette variable n'est connu au niveau de l'ensemble du cadre de sondage. Elle n'intervient pas non plus au niveau du modèle de réponse car l'utilisation de cette variable exigerait une disponibilité au niveau de l'échantillon brut (unités interrogées dans l'enquête), ce qui n'est pas le cas. Dans ce rapport, le terme NOGA2 utilisé seul – sans préciser s'il s'agit d'une NOGA2 MON ou d'une NOGA2 initiale (cadre) – fera toujours référence à la NOGA2 cadre.

La grande région, GR, d'une unité est une variable issue du découpage de la géographie Suisse au niveau de granularité NUTS 2¹. Il y a en Suisse 7 grandes régions (décrites par exemple dans Chevalier [2024], section 2.2) : Région lémanique, Espace Mittelland, Suisse du Nord-Ouest, Zurich, Suisse orientale, Suisse centrale et Tessin. La grande région est une variable connue sur tout le cadre de sondage qui est aussi utilisée lors de l'élaboration du plan de sondage, par exemple pour définir des objectifs de précision. Dans le cadre de la pondération, la grande région intervient au niveau du modèle de réponse, voir section 3.2 et du calage, voir section 3.4. Ce rapport présente plusieurs résultats et interprétations par grande région.

Le nombre d'employés d'une unité est renseigné dans le cadre de sondage au moyen de deux variables notamment. La variable `EPT_cadre` contient le nombre d'employés en équivalent plein temps de l'unité. La variable `betot_cadre` rend compte du nombre d'employés, sans tenir compte du taux d'activité. Un employé travaillant par exemple à 40% va ainsi compter pour 0.4 `EPT_cadre` et 1 `betot_cadre`. Ces deux variables sont disponibles sur l'ensemble du cadre de sondage. La variable `EPT_cadre` est la variable principale de calage de l'ancienne méthode de pondération. Pour la révision présentée dans ce rapport cette variable de calage est conservée et de nouvelles variables sont ajoutées, voir section 3.4. Plusieurs variables catégorielles sont créées à partir de `betot_cadre` et sont utilisées notamment pour le modèle de réponse.

- Une classe de taille à 5 modalités, `clt5` : tout comme pour le plan de sondage [Chevalier, 2024], les classes de taille 1 à 5 sont définies pour un `betot_cadre` situé respectivement dans les intervalles [3,9], [10,19], [20,49], [50,249], [250,∞).
- Une classe de taille à 3 modalités, "`clt3`", est utilisée pour le modèle de réponse (voir section 3.2) en croisant `clt3` avec `betot_cadre`. Les 3 modalités correspondent à un nombre d'employés dans les intervalles [3,9], [10,49], [50,∞) respectivement.
- Une classe de taille modifiée, `clt_x`, qui se base à la fois sur le nombre d'employés `betot_cadre` et sur la productivité TVA (telle que définie en section 2.2) est créée lors de l'élaboration du plan de sondage et est définie sur le cadre de sondage. Sa construction exacte est présentée dans le rapport de méthode sur le plan de sondage [Chevalier, 2024], en section 5.2.2. La variable `clt_x` est utilisée pour construire certaines variables de calage, voir section 3.4.2.

2.2 Chiffre d'affaires TVA, productivité TVA et données AVS

L'OFS a accès aux décomptes de taxe sur la valeur ajoutée (TVA) qui permettent de déduire un chiffre d'affaires pour un grand nombre d'unités. Ce chiffre d'affaires est nommé "chiffre

1. <https://www.europarl.europa.eu/factsheets/fr/sheet/99/la-nomenclature-commune-des-unites-territoriales-statistiques-nuts>

d'affaires TVA". Il est disponible pour l'année antérieure à l'année de référence de l'enquête. Ainsi, pour la WS 2021, c'est le chiffre d'affaires TVA de l'année 2020 qui est potentiellement disponible. Lorsqu'aucun décompte n'est disponible, un chiffre d'affaires TVA est tout de même imputé pour l'unité, selon une procédure décrite dans [Saliba \[2022\]](#) et résumée dans [Chevalier \[2024\]](#), section 2.3. Le chiffre d'affaires TVA est ainsi une variable auxiliaire disponible sur tout le cadre de sondage dont l'importance est soulignée dans [Chevalier \[2024\]](#). C'est en grande partie la disponibilité de cette nouvelle variable qui a motivé la révision du plan de sondage et de la pondération de la WS. Compte tenu de son importance, le chiffre d'affaires TVA est fortement contrôlé et éventuellement corrigé. Ces contrôles sont effectués lors de l'élaboration du plan de sondage, voir [Chevalier \[2024\]](#), section 3. Pour la pondération, ce sont ces chiffres d'affaires TVA contrôlés qui sont utilisés. Le chiffre d'affaires TVA intervient à chaque étape de la pondération : modèle de réponse, robustification et calage. Le calage du chiffre d'affaires TVA dans la pondération est peut-être l'une des principales améliorations apportée par cette révision. Celui-ci permet des gains de précision dans la mesure où sa corrélation avec les réponses à l'enquête est forte (voir par exemple les tableaux dans [Chevalier \[2024\]](#), annexe B). Des résultats chiffrés sur les gains de précision apportés par le calage du chiffre d'affaires TVA sont donnés en section 4.3.

La productivité TVA d'une entreprise est définie comme le ratio entre chiffre d'affaires TVA et EPT_cadre. Cette variable intervient dans le modèle de réponse, via la création de variables catégorielles. Le facteur de productivité TVA d'une entreprise, aussi présenté dans [Chevalier \[2024\]](#) section 2.3, est le ratio entre une productivité TVA et une productivité TVA médiane au sein de la NOGA2. Une entreprise avec un facteur de productivité TVA supérieur à 1 est ainsi plus productive que 50% des entreprises de sa NOGA2. Le facteur de productivité TVA intervient dans le modèle de réponse en tant que variable catégorielle. L'interprétation principale est que les entreprises les plus productives ont tendance à répondre davantage. Ce point est traité en détail en section 4.2.4.

L'OFS dispose de données sur les revenus individuels, mises à disposition par les caisses de compensation Assurance-vieillesse et survivants (AVS). Un revenu annuel est disponible pour les personnes ayant cotisé à l'AVS. Pour la pondération de la WS 2021, ce sont les revenus AVS au cours de l'année 2020 qui sont utilisés, car les revenus de l'année 2021 ne sont pas encore disponibles au moment de la pondération. Les données de l'AVS ont déjà été utilisées dans le passé pour la pondération dans des enquêtes, en particulier pour l'enquête sur la structure des salaires, voir [Chevalier & Qualité \[2021\]](#). Pour la WS, les données AVS sont utilisées dans le modèle de réponse avec deux résultats importants, voir section 4.2.3. Premièrement, les entreprises du cadre WS pour lesquelles des revenus AVS sont retrouvés ont tendance à répondre plus souvent lorsqu'elles payent des salaires plus élevés. Cette interprétation n'est pas nouvelle et avait déjà été observée dans l'enquête sur la structure des salaires [[Chevalier & Qualité, 2021](#)]. Deuxièmement, les entreprises du cadre WS pour lesquelles aucun revenu AVS n'est retrouvé ont tendance à répondre moins souvent. Cette seconde interprétation est nouvelle. Des résultats détaillés sont donnés en section 4.2.3.

2.3 Plan de sondage et VIU

La pondération d'une enquête suppose que les unités de l'échantillon brut ont été tirées selon un certain plan de sondage. La méthodologie de construction de celui-ci a été révisée et est décrite dans le rapport de méthodes [Chevalier \[2024\]](#). Le plan de sondage est un plan de Poisson où les unités de l'échantillon brut sont sélectionnées indépendamment, avec un taux de sondage qui dépend de la strate de l'unité. La probabilité de tirage π_i d'une unité i confère naturellement un

poids de sondage $1/\pi_i$ lorsque l'unité i est tirée dans l'échantillon brut. Ce poids ne sera cependant généralement pas égal au poids final utilisé lors de la pondération pour plusieurs raisons. Premièrement, le poids d'une unité est relié à sa probabilité d'inclusion dans l'échantillon net, c'est-à-dire l'échantillon d'entreprises interrogées dans l'enquête (échantillon brut) qui de plus répondent à celle-ci. La réponse à l'enquête n'est pas garantie et des analyses sur les taux de réponse sont fournies en sections 4.1 et 4.3. La modélisation de la probabilité de réponse est traitée en section 3.2. Deuxièmement, les poids obtenus en prenant l'inverse de la probabilité d'inclusion estimée sont encore modifiés par deux opérations successives : la robustification et le calage. Ces opérations visent à améliorer la précision des estimations et assurer leur cohérence avec des informations auxiliaires disponibles.

Pour la pondération de la WS, certaines unités nommées "Very Important Units" (VIU) reçoivent un traitement différent des autres. Les arguments en faveur de l'utilisation de VIU et la procédure pour choisir ces unités sont présentés dans Chevalier [2024] section 5.1. Les VIU sont ainsi choisies lors de l'élaboration du plan de sondage et sont au nombre de 252 pour la WS 2021, voir Chevalier [2024] section 7.3. Lors de la pondération, certaines entreprises sont potentiellement ajoutées à la liste des VIU si elles remplissent les trois conditions suivantes :

- elles font partie de l'enquête de l'année précédente et y ont répondu,
- elles font partie de l'enquête actuelle, sont dans une strate de sondage exhaustive (c'est-à-dire avec un taux de sondage de 100%) et répondent à l'enquête,
- leur réponse t_{010} à l'enquête de l'année précédente dépasse 10% du total de la variable t_{010} dans la NOGA2 de l'entreprise, extrapolé à l'aide des poids obtenus lors de cette pondération (c'est-à-dire la pondération de l'enquête de l'année précédente).

Les entreprises qui n'étaient pas des VIU et le deviennent à partir des règles ci-dessus sont nommées "VIU t_{010} ". Le fait de rajouter des VIU t_{010} après avoir établi le plan de sondage n'est pas nouveau et était déjà effectué dans le passé, selon la même procédure. Dans la mesure du possible, la nouvelle procédure de choix des VIU et les interactions avec la section MON décrites dans Chevalier [2024] visent à limiter le nombre d'entreprises qui "deviennent" des VIU a posteriori (c'est-à-dire après avoir construit le plan de sondage) et ce pour mieux calculer les précisions et coefficients de variation (CV) de l'enquête lors du plan de sondage. Pour la WS 2021, 15 VIU t_{010} sont ajoutées à la liste des VIU, ce qui porte le nombre total de VIU à 267.

Le nombre de VIU t_{010} diminue lors de la WS 2022 en raison de cette volonté de choisir autant que possible les VIU lors de la construction du plan de sondage : pour la WS 2022 il y a 8 VIU t_{010} .

Les VIU sont exclues des différentes étapes de la pondération (robustification, calage) dans le sens où leur poids est fixé directement à 1. Ceci confère implicitement aux VIU une probabilité de réponse de 100% et cette hypothèse forte doit être cohérente avec les taux de réponse observés en pratique. Pour la WS 2021, sur 267 VIU, il y a une seule unité non-répondante, voir table 3. Pour la WS 2022, le nombre de VIU est relativement proche (257) et toutes les VIU ont répondu.

3 Non-réponse, robustification et calage

3.1 Vue d'ensemble de la pondération

L'objectif de la pondération de la WS est d'attribuer un poids aux unités (entreprises) qui répondent à l'enquête en vue d'effectuer des estimations basées sur ces poids. D'après la théorie des sondage (voir Tillé [2019], chapitre 2) ces estimations sont sans biais si les poids des unités sont égaux à l'inverse de leur probabilité d'inclusion dans l'échantillon net. Dans la pratique,

cette probabilité est estimée. Une unité fait partie de l'échantillon net si elle est interrogée dans l'enquête (i.e. présente dans l'échantillon brut) et qu'elle accepte d'y répondre. L'inverse de la probabilité estimée d'inclusion de l'unité i s'écrit :

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

où π_i est la probabilité de tirage de l'unité i dans l'échantillon brut, déterminée lors de l'élaboration du plan de sondage et $\hat{p}_i$ est une estimation de la probabilité de réponse de l'unité i conditionnellement à sa sélection dans l'échantillon brut. Cette probabilité est estimée au moyen d'un modèle de réponse, voir section 3.2.

Le poids d_i^* obtenu après avoir estimé la probabilité de réponse n'est pas le poids final proposé pour la pondération. Celui-ci est en effet modifié par deux processus différents. Premièrement, une opération dite de "robustification" transforme le poids d_i^* en un nouveau poids

$$d_i := d_i^* - \alpha_i(d_i^* - 1), \quad (2)$$

où $0 \leq \alpha_i \leq 1$ est un facteur qui dépend de l'unité i . Lorsque $\alpha_i = 0$, le poids d_i^* n'est pas modifié, ce qui pour la WS 2021 est le cas pour un peu plus de 99% des unités. Lorsque $\alpha_i > 0$, le poids de l'unité est rapproché de la valeur seuil "1" et le poids est dit "robustifié". La robustification la plus forte pour une unité i correspond au cas où $\alpha_i = 1$. Dans ce cas-là, le poids robustifié d_i est égal à 1 ce qui constitue le poids le plus faible possible avec cette nouvelle méthodologie. Les détails et justifications sur ces transformations sont donnés en section 3.3. La robustification permet en théorie d'obtenir des estimations plus précises, mais comporte un risque de biais. L'impact de la robustification sur les estimations est étudié en section 3.3.5. Des résultats chiffrés sont donnés en table 15.

Après la robustification, l'ensemble des poids d_i est disponible pour toutes les unités de l'échantillon net. Ces poids dits "après robustification" sont modifiés par une procédure de calage qui donne des poids finaux, notés w_i , utilisés pour les estimations. Le calage vise à améliorer l'estimation en permettant de retrouver des totaux connus sur des variables auxiliaires bien choisies. Dans le cadre de la révision, le calage a été revu, avec notamment l'ajout de nouvelles variables. Ceci est décrit en section 3.4. Les gains de précisions obtenus grâce au calage sont présentés et discutés en section 4.3.

Les unités qui reçoivent un poids à l'aide du modèle de réponse, d'une robustification et d'un calage sont les unités de l'échantillon net en excluant les VIU répondantes (266 unités pour la WS 2021). Les VIU répondantes ont, comme expliqué précédemment, un poids fixé à 1 et sont exclues des différentes étapes de la pondération. Les VIU non-répondantes ne reçoivent pas de poids puisqu'elles ne sont pas dans l'échantillon net. Elles interviennent cependant – contrairement aux VIU répondantes – dans l'estimation des coefficients du modèle de réponse ainsi que dans le calcul des totaux à caler lors de l'étape de calage, voir section 3.4.

3.2 Modèle de réponse

Un modèle de réponse est utilisé pour estimer la probabilité de réponse $\hat{p}_i$ de l'équation (1). Le modèle utilisé s'inspire de la révision effectuée sur une autre enquête de l'OFS : l'enquête sur la structure des salaires, voir [Chevalier & Qualité \[2021\]](#), section 3.2. Il s'agit d'un modèle de régression logistique où la quantité $\text{logit}(\hat{p}_i) := \log[\hat{p}_i / (1 - \hat{p}_i)]$ dépend linéairement de plusieurs prédicteurs selon l'équation suivante,

$$\text{logit}(\hat{p}_i) = \beta_0 + \sum_{j=1}^{n_{\text{pred}}} \beta_j x_{i,j}, \quad (3)$$

où n_{pred} est le nombre de prédicteurs sans compter la constante, $\beta_0, \dots, \beta_{n_{\text{pred}}}$ sont les coefficients du modèle à estimer et $x_{i,j}$ est la valeur du j -ème prédicteur pour l'unité i . Le logiciel SAS version 8.3 et la procédure "logistic" sont utilisés pour estimer les coefficients du modèle. Ce dernier se base sur l'échantillon brut (22 912 unités pour la WS 2021) et l'échantillon net (15 186 unités) en excluant les VIU répondantes (266 unités) du modèle. Les taux de réponse globaux en incluant et excluant les VIU répondantes sont respectivement de 66.3% et 65.9%. Les prédicteurs du modèle sont décrits ci-dessous. La procédure "logistic" est utilisée avec une option "stepwise" qui permet de sélectionner ou non les prédicteurs selon leur pouvoir prédictif dans une liste fournie. Dans le cas présent, tous les prédicteurs proposés sont sélectionnés et l'option "stepwise" est enlevée pour les enquêtes WS des années suivantes, de sorte à conserver le même modèle d'année en année. Pour la WS 2021, les prédicteurs sont décrits ci-dessous par ordre de sélection.

1. La constante (intercept, 1 coefficient β_0).
2. La classe de taille `c1t5`, basée sur `betot_cadre`. Cette variable est décrite en section 2. La variable ayant 5 modalités, 4 prédicteurs sont créés, avec 4 coefficients β_j à estimer. Les résultats suggèrent que les grandes entreprises répondent plus fréquemment. Les coefficients estimés par le modèle et des interprétations sont donnés en section 4.2.2.
3. La `NOGA2`, sans regroupements. Le nombre de modalités, lorsqu'il n'y a pas de regroupement de `NOGA2`, est de 74, ce qui correspond à 73 coefficients à estimer. Les 73 coefficients estimés ne sont pas reproduits dans ce rapport, mais des taux de réponse par `NOGA2` sont donnés en tables 9 et 10.
4. Une variable catégorielle nommée "`pct_sup50_cat`" basée sur les données de revenus AVS. La variable est une variable catégorielle construite de façon relativement complexe. Pour construire les catégories, la démarche suivante est adoptée :
 - si aucun revenu AVS n'est trouvé pour l'entreprise, celle-ci est mise dans une catégorie à part entière, nommée par convention "99",
 - si des revenus AVS sont trouvés, chaque revenu AVS est comparé au revenu AVS médian sur toute la Suisse. Le pourcentage de personnes dont la rémunération est au dessus de cette médiane est calculé pour chaque entreprise. Si ce pourcentage est compris entre 5% (inclus) et 95% (exclus), il est arrondi au 10% le plus proche; ce qui mène à la création de 9 catégories distinctes nommées "10%", "20%", ..., "90%". Si le pourcentage est strictement inférieur à 5%, l'unité est mise dans une catégorie "0%", mais celle-ci est croisée avec la variable `c1t3` définie en section 2.1. Il y a ainsi 3 modalités, $0\% \times 1$, $0\% \times 2$ et $0\% \times 3$, si `betot_cadre` est dans l'intervalle $[3,9]$, $[10,49]$ ou $[50,\infty)$ respectivement. Enfin, si le pourcentage de personnes payées au dessus du revenu AVS suisse médian est supérieur ou égal à 95%, l'unité est classée dans une catégorie "100%" qui est elle aussi croisée avec `c1t3`, ce qui mène à la création des catégories $100\% \times 1$, $100\% \times 2$ et $100\% \times 3$.

Les raisons qui poussent à l'utilisation de catégories aussi complexes sont multiples. D'une part, il apparaît que les entreprises qui ne sont pas retrouvées dans le fichier de revenus AVS (catégorie "99") répondent beaucoup moins à l'enquête, ce qui justifie la création d'une catégorie à part entière à prendre en compte dans le modèle logistique. Le fait de ne pas retrouver l'entreprise dans le fichier de revenus AVS utilisé par l'OFS est ainsi une information en soi qui peut être exploitée pour améliorer les prédictions quant à la propension de l'unité à répondre. D'autre part, lorsque l'entreprise est retrouvée dans le fichier AVS, les catégories proposées ici reprennent les idées développées dans le modèle de réponse de l'enquête sur la structure des salaires (voir [Chevalier & Qualité \[2021\]](#), section 3.2.2) et en particulier le fait que les entreprises qui ont tendance à mieux rémunérer leurs employés (selon les données AVS) ont tendance à mieux répondre à

l'enquête. Les taux de réponse sont ainsi plus élevés dans les catégories "90%", "80%" que dans les catégories "10%", "20%". Par rapport à l'enquête sur la structure des salaires, une subtilité supplémentaire est introduite et se base sur l'idée que les entreprises dans la catégorie "0%" ou bien "100%" ne se comportent pas forcément de la même façon si elles ont peu (par exemple : 3) d'employés ou au contraire beaucoup d'employés. Autrement dit, le fait de faire partie de la catégorie 0% ou 100% ("aucun", respectivement "tous" les salariés payés au dessus du revenu AVS médian) est une information plus forte lorsque l'entreprise a beaucoup d'employés, avec des effets attendus (positifs ou négatifs) plus forts sur les taux de réponse. La variable catégorielle ainsi créée a $1 + 9 + (2 \times 3) = 16$ modalités, ce qui correspond à 15 coefficients à estimer. Ces coefficients sont reproduits en table 6 et des interprétations sont données en section 4.2.3.

5. Une variable catégorielle à 10 modalités (donc 9 coefficients estimés) basée sur le facteur de productivité TVA (voir section 2.2 pour une définition). Le logarithme du facteur de productivité TVA est calculé, puis transformé en une variable catégorielle en considérant les 10 intervalles : $(-\infty, -2.75)$, $[-2.75, -2.25)$, $[-2.25, -1.5)$, $[-1.5, -0.5)$, $[-0.5, 0.1)$, $[0.1, 0.3)$, $[0.3, 0.5)$, $[0.5, 0.7)$, $[0.7, 0.9)$, $[0.9, \infty)$. Le fait d'utiliser une variable catégorielle au lieu d'une variable continue permet de modéliser des effets non-linéaires du facteur de productivité TVA sur le taux de réponse. Les intervalles qui servent à définir les modalités ont été choisis en estimant, à l'aide d'un lissage par noyau, les taux de réponse en fonction du logarithme du facteur de productivité TVA, voir figure 3. La prise en compte de cette variable catégorielle dans le modèle de réponse permet d'intégrer le fait que les entreprises les plus productives ont tendance à répondre davantage. Ce prédicteur est analysé en détail en section 4.2.4 avec des résultats chiffrés en table 7.
6. La variable numérique `betot_cadre` croisée avec la variable catégorielle `c1t3`. Le croisement d'une variable discrète (`betot_cadre`) avec une variable catégorielle (`c1t3`) permet de modéliser des effets non-linéaires du nombre d'employés sur la probabilité de réponse. Le fait d'intégrer cette variable, alors même que la variable de classe de taille `c1t5` est déjà présente dans le modèle, permet de modéliser un effet plus fin de `betot_cadre` sur la probabilité de réponse, notamment pour les très grandes entreprises. Des analyses en ce sens sont effectuées en section 4.2.2, voir notamment la figure 2.
7. La grande région (GR) croisée avec `c1t3`. Il s'agit d'une variable catégorielle avec $7 \times 3 = 21$ modalités et au total $6 \times 3 = 18$ coefficients estimés. La grande région est le prédicteur sélectionné en dernier dans le modèle, mais celui-ci a une influence largement significative (p-valeur inférieure à 0.0001). Des analyses sont fournies en section 4.2.5 et des résultats chiffrés en table 8.

Les prédicteurs introduits dans le modèle correspondent à respectivement 1, 4, 73, 15, 9, 3 et 18 coefficients à estimer, ce qui fait un total de 123 paramètres pour le modèle de réponse. Ceci est à mettre en perspective avec le nombre d'observations utilisées pour calibrer le modèle, qui est égal à la taille de l'échantillon brut sans les "Very Important Units" (VIU) répondantes, c'est-à-dire 22 646 unités pour la WS 2021. Le nombre de paramètres à estimer est inférieur à celui de l'ancienne méthode. En effet, l'ancienne méthode estime dans chaque strate un taux de réponse empirique en comptant le nombre de répondants dans la strate et en divisant par la taille de l'échantillon brut dans cette strate. Les strates ayant été définies comme le croisement entre les variables `NOGA2` (avec regroupements, mais tout de même environ 70 modalités) et `c1t5` (5 modalités), l'ancienne méthodologie estime ainsi plus de 300 paramètres. Dans la littérature, Harrell [2001] recommande, sur la base de simulations numériques, d'utiliser des modèles logistiques dont le nombre de paramètres est inférieur à 10% (idéalement 5%) du nombre d'observations de sorte à éviter des problèmes de surajustement (overfitting).

Par rapport au modèle de réponse développé dans l'enquête sur la structure des salaires [Chevalier & Qualité, 2021], une variable importante a été testée mais pas retenue pour la WS. Il s'agit de la variable qui mesure le pourcentage, selon l'AVS, d'employés de sexe masculin. Dans l'enquête sur la structure des salaires, cette variable est rendue catégorielle avec 11 modalités et avait une influence légère, mais significative, sur les taux de réponse, voir Chevalier & Qualité [2021], section 3.2.2. L'interprétation était que les sociétés qui employaient davantage de femmes avaient des taux de réponses légèrement meilleurs. Pour la WS, l'effet de cette variable n'est pas statistiquement significatif et ainsi la variable n'est pas intégrée au modèle de réponse.

3.3 Robustification

3.3.1 Motivations et calculs de variance

Comme expliqué en section 3.1, le modèle de réponse permet d'estimer une probabilité de réponse pour chaque unité de l'échantillon brut, et à fortiori pour les unités de l'échantillon net. Ceci permet d'estimer l'inverse de la probabilité d'inclusion, selon l'équation (1).

L'intérêt d'une opération dite de "robustification" des poids peut être motivé à l'aide de calculs de variance. Soit y une variable relevée dans l'enquête, disponible sur l'échantillon net S . Le total de la variable y sur tout le cadre de sondage est noté T_y et est inconnu. Celui-ci est classiquement estimé à partir des réponses sur l'échantillon net, à l'aide de la somme pondérée

$$\hat{T}_y = \sum_{i \in S} d_i^* y_i, \quad (4)$$

où les poids d_i^* sont donnés par l'équation (1). Dans le cas d'un plan de sondage de Poisson et dans le cas où la probabilité de réponse de l'unité i est égale à $\hat{p}_i$, le biais de l'estimateur de l'équation (4) est nul et sa variance est classiquement estimée par

$$\widehat{\text{Var}}(\hat{T}_y) = \sum_{i \in S} d_i^* (d_i^* - 1) y_i^2. \quad (5)$$

En présence de calage, l'équation (5) s'applique en remplaçant la réponse y_i par le résidu ε_i d'une régression linéaire appliquée sur les unités de l'échantillon net S dans laquelle les réponses sont prédites à l'aide de toutes les variables de calage (voir section 3.4.2 pour une description complète de ces variables). À partir de l'équation (5) il apparaît que les unités qui ont un résidu au carré ε_i^2 élevé affectent fortement la variance des estimations si leur poids d_i^* est strictement supérieur à 1.

Comme discuté dans le rapport de méthode [Chevalier, 2024], un indicateur de l'influence d'une unité sur les estimations et leur variance est le biais conditionnel (voir Chevalier [2024], section 3.1 et les références citées). Pour un plan de sondage de Poisson et en présence de calage, le biais conditionnel d'une unité répondante i est

$$\text{biais}_i^{\text{cond}} = (d_i^* - 1) \varepsilon_i. \quad (6)$$

Il est possible de tirer des interprétations de l'équation (6). Une unité i qui donne une réponse supérieure à ce qui est "attendu" compte tenu des variables de calage (résidu $\varepsilon_i > 0$) aura un biais conditionnel positif si son poids est strictement supérieur à 1. À l'inverse, une unité avec $\varepsilon_i < 0$ (c'est-à-dire qui répond une valeur inférieure à ce qui est attendu) aura un biais conditionnel négatif, toujours si son poids est strictement supérieur à 1. Un article [Beaumont et al.,

[2013] s'appuie sur le biais conditionnel pour proposer une méthode dite de "robustification" des poids. Celle-ci vise à diminuer le poids de certaines unités répondantes en fonction de leur biais conditionnel. La méthode proposée dans Beaumont et al. [2013] n'est pas reprise telle quelle dans ce rapport, mais inspire fortement l'algorithme de robustification des poids proposé pour la WS. Cet algorithme est décrit dans la section suivante.

3.3.2 Algorithme de robustification

L'estimateur "robuste" du total dans un domaine proposé par Beaumont et al. [2013] consiste à reprendre l'estimateur classique de l'équation (4) et à soustraire la moitié de la somme du plus grand (c'est-à-dire, le plus positif) et du plus petit (c'est-à-dire, le plus négatif) biais conditionnel $\text{biais}_i^{\text{cond}}$, parmi les unités répondantes. Cette approche revient à agir sur les poids d_i^* de deux unités, avec les biais conditionnels positifs et négatifs les plus forts, en transformant ces d_i^* en de nouveaux poids $d_i^* - \frac{1}{2}(d_i^* - 1)$. Ces résultats sont obtenus en section 3.4 de Beaumont et al. [2013] par des calculs de minimisation d'un biais conditionnel maximal.

La transformation appliquée sur les poids des unités avec le plus petit et plus grand biais conditionnel rapproche ces poids de la valeur 1 de la moitié de leur distance à 1. Cette transformation est appliquée sur les poids de deux unités seulement. Les poids des autres unités demeurent inchangés. L'idée de robustifier les poids en considérant des paires d'unités ayant des biais conditionnels opposés est intéressante. Intuitivement, le poids d'une unité qui répond "davantage que prévu" est réduit, ainsi que celui d'une unité qui répond "moins que prévu" de sorte à créer un effet de compensation. Ceci devrait aider à contrôler le biais si les résidus ε_i sont distribués de façon à peu près symétrique autour de zéro.

La méthode proposée dans Beaumont et al. [2013] s'applique dans une situation où une seule réponse y est donnée par les unités et où l'estimation porte sur un seul total : par exemple un total sur toute la Suisse. En pratique, les estimations ne sont pas seulement effectuées sur toute la Suisse, mais aussi par domaines : par exemple par NOGA2, grande région ou canton. De plus, ces estimations sont faites pour plusieurs réponses à l'enquête telles que le chiffre d'affaires (t010), la production brute (t017), les consommations intermédiaires (t022) ou la valeur ajoutée (t029 = t017 - t022). L'algorithme de robustification proposé dans cette révision se base ainsi sur les biais conditionnels des unités, mais avec plusieurs différences par rapport à Beaumont et al. [2013]. Ces différences sont résumées ci-dessous.

- L'algorithme de robustification est appliqué à chaque NOGA2 MON, voir section 2.1 pour un rappel de définition. Cela signifie qu'à l'intérieur de chaque NOGA2 MON, des paires d'unités peuvent être sélectionnées et avoir un poids (de chaque unité composant la paire) robustifié. Le choix d'appliquer l'algorithme au niveau NOGA2 MON et pas au niveau de la NOGA2 "classique" du cadre de sondage est justifié en section 3.3.4.
- Il n'y a pas forcément une unique paire d'unités dans chaque NOGA2 MON dont le poids est robustifié. Le nombre de paires robustifiées est décidé par un calcul décrit en section 3.3.3. Il peut potentiellement prendre toute valeur positive ou nulle et varie selon la NOGA2 MON.
- Plusieurs réponses à l'enquête sont considérées. Pour une unité fixée, il suffit qu'une seule des réponses considérées satisfasse certains critères pour que le poids de l'unité soit robustifié. Le poids obtenu est utilisé pour les extrapolations par rapport à chacune des réponses. Il n'y a pas des poids différents associés aux différentes réponses à l'enquête.
- La robustification d'un poids d_i^* ne consiste pas forcément à appliquer la transformation $d_i = d_i^* - \frac{1}{2}(d_i^* - 1)$. En cas de robustification, la distance du poids d_i^* à 1 peut selon le cas

être réduite de 50% (transformation identique à [Beaumont et al. \[2013\]](#)) où de 100% (poids mis à exactement 1). Les détails sont donnés en section [3.3.3](#). D'une manière générale, la robustification consiste à appliquer la transformation $d_i = d_i^* - \alpha_i(d_i^* - 1)$ où le coefficient $0 \leq \alpha_i \leq 1$ est calculé par l'algorithme. La robustification du poids d'une unité sera dite "partielle" si $\alpha_i = 0.5$ et "totale" si $\alpha_i = 1$.

- Certains critères supplémentaires sont pris en compte par l'algorithme de robustification. En particulier, les entreprises dont la NOGA2 MON est différente de la NOGA2 du cadre de sondage ont un traitement légèrement différent, voir section [3.3.4](#).

En considérant une réponse y à l'enquête et un domaine qui sera, sauf exception, la NOGA2 MON l'algorithme de robustification a la forme suivante :

1. Une NOGA2 MON est sélectionnée. Pour cette NOGA2 MON, le total de la variable y , noté $\hat{T}_y^{\text{prov}}$, est provisoirement estimé à l'aide de la somme des réponses dans la NOGA2 MON pondérée par les poids d_i^* actuels, où d_i^* est donné par l'équation (1).
2. Une régression linéaire, spécifique à la NOGA2 MON, est effectuée sur toutes les unités de l'échantillon net S . La variable régressée est égale à la réponse y si l'entreprise fait partie de la NOGA2 MON sélectionnée et 0 sinon. Les régresseurs sont les variables de calage et sont décrits en section [3.4.2](#).
3. À l'aide de la régression linéaire, les résidus ε_i sont calculés, ainsi que les biais conditionnels $(d_i^* - 1)\varepsilon_i$.
4. Les biais conditionnels des entreprises faisant partie de la NOGA2 MON considérée sont triés par ordre croissant et sont divisés par le total estimé dans la NOGA2 MON $\hat{T}_y^{\text{prov}}$ (voir : étape 1). Les biais conditionnels sont ainsi exprimés en pour cent de $\hat{T}_y^{\text{prov}}$, ce qui simplifie les étapes suivantes où des seuils interviennent.
5. Les entreprises de la NOGA2 MON sélectionnée sont classées par paires de la façon suivante : l'entreprise avec le biais conditionnel le plus positif est appariée avec l'entreprise avec le biais conditionnel le plus négatif. L'entreprise avec le second biais conditionnel le plus positif est appariée avec l'entreprise qui a le second biais conditionnel le plus négatif, etc. Pour chaque paire, le plus grand biais conditionnel, en valeur absolue, est considéré. Ce biais conditionnel est associé à la paire et décide de la robustification des poids. Si le biais conditionnel de la paire dépasse un seuil (détaillé en section [3.3.3](#)), alors la paire est robustifiée, c'est-à-dire que les deux poids des deux entreprises de la paire sont robustifiés. Sinon, les poids des unités de la paire ne sont pas modifiés. Les paires ainsi formées sont constituées uniquement d'entreprises faisant partie de la NOGA2 MON considérée. Les entreprises en dehors de cette NOGA2 MON n'ont en général pas un biais conditionnel nul, mais elles ne sont pas prises en compte pour former les paires.
6. Une fois la NOGA2 MON traitée, l'algorithme boucle en passant à la NOGA2 MON suivante, jusqu'à ce que toutes les NOGA2 MON soient traitées.

3.3.3 Variables de l'enquête considérées par l'algorithme

L'algorithme de robustification décrit en section [3.3.2](#) est appliqué successivement pour quatre variables de l'enquête : t010, t017, t022, t053. Il comporte plusieurs paramètres (seuils).

- Le chiffre d'affaires selon l'enquête, t010, est considéré en premier. Pour cette variable, et uniquement celle-ci, l'algorithme considère seulement la NOGA2 MON 46 (commerce de gros, à l'exception des automobiles et des motocycles) et aucune des autres. Ceci est dû au fait qu'une robustification est effectuée par la suite pour toutes les NOGA2 MON en

considérant la variable t_{017} qui est très similaire à la variable t_{010} dans toutes les NOGA2 MON, à l'exception de la NOGA2 MON 46. Effectuer une robustification basée sur t_{010} dans toutes les NOGA2 MON créerait ainsi un effet de redondance, ce qui est évité en considérant (pour la variable t_{010}) seulement le domaine où t_{010} et t_{017} sont potentiellement très différentes. Ces différences entre t_{010} et t_{017} observées uniquement dans la NOGA2 MON 46 sont liées au point suivant : la production brute t_{017} inclut la marge, c'est-à-dire la différence entre les ventes (variable 501, relevée dans l'enquête) et les achats (variable 601). La production brute est ainsi inférieure au chiffre d'affaires t_{010} , car elle prend en compte les coûts d'achat. Pour les autres NOGA2 MON, la production brute correspond au chiffre d'affaires total t_{010} , auquel s'ajoutent d'autres produits, sans déduction des achats. La production brute est ainsi généralement supérieure au chiffre d'affaires, car elle ne prend pas en compte les coûts d'achat et peut de plus inclure d'autres éléments de revenu.

Une paire d'unités construite selon l'algorithme est robustifiée avec un coefficient $\alpha_i = 1/2$ (robustification partielle) si le biais conditionnel de la paire dépasse 0.15% de $\hat{T}_{t_{010}}^{\text{prov}}$. Si de plus ce biais conditionnel dépasse 0.2% de $\hat{T}_{t_{010}}^{\text{prov}}$, alors α_i est fixé à 1 (robustification totale), ce qui veut dire que les poids des unités de la paire sont fixés à exactement 1. Des justifications sur les choix des différents seuils sont données dans le dernier paragraphe de cette section.

- Une fois la robustification par rapport à la variable t_{010} terminée, les nouveaux poids sont utilisés pour effectuer une nouvelle robustification par rapport à la variable t_{017} (sans recalcul du total $\hat{T}_{t_{017}}^{\text{prov}}$). Les entreprises dont le poids a changé lors de l'étape précédente (que ça soit une robustification totale ou non) ne peuvent plus subir de modifications supplémentaires sur leur poids dans le cadre de robustifications successives. Ces entreprises ne sont ainsi plus classées dans des paires en vue d'une possible robustification. Pour la variable t_{017} , toutes les NOGA2 MON sont considérées. La paire avec le biais conditionnel le plus élevé parmi toutes les paires est robustifiée (robustification partielle) si son biais conditionnel dépasse 1% de $\hat{T}_{t_{017}}^{\text{prov}}$. Les autres paires sont robustifiées (robustification partielle) si le biais conditionnel dépasse 2% de $\hat{T}_{t_{017}}^{\text{prov}}$. Si de plus le biais conditionnel dépasse 3% de $\hat{T}_{t_{017}}^{\text{prov}}$, la robustification devient totale ($\alpha_i = 1$).
- La robustification suivante est faite en considérant la variable t_{022} (consommations intermédiaires) et toutes les NOGA2 MON. Les entreprises dont le poids a déjà été robustifié ne peuvent plus l'être à nouveau. La robustification des poids est partielle si le biais conditionnel dépasse 2% de $\hat{T}_{t_{022}}^{\text{prov}}$ et totale s'il dépasse 3% de $\hat{T}_{t_{022}}^{\text{prov}}$.
- La dernière robustification porte sur la variable t_{053} (masse salariale) et sur toutes les NOGA2 MON. Les seuils de robustification partielle et totale sont de 2% et 3% de $\hat{T}_{t_{053}}^{\text{prov}}$ respectivement.

L'algorithme de robustification est ainsi appliqué pour 4 variables différentes de l'enquête avec des seuils et des domaines choisis pour contrôler le nombre d'entreprises robustifiées. La variable t_{017} est la seule variable pour laquelle un seuil différencié (seuil de 1%) est appliqué à la paire avec le biais conditionnel le plus élevé par rapport aux autres paires (seuil de 2%). Ceci est effectué dans le but de privilégier dans la mesure du possible la robustification d'une seule paire par NOGA2 MON de sorte à se rapprocher de l'algorithme décrit dans [Beaumont et al. \[2013\]](#). Les robustifications avec les variables t_{010} , t_{022} et t_{053} sont jugées comme un peu moins importantes que t_{017} et se font sans seuils différenciés. Finalement, l'ordre appliqué dans ces variables ainsi que les seuils visent à avoir un nombre d'entreprises robustifiées un peu plus élevé en raison de leur réponse t_{017} par rapport à t_{010} , t_{022} et t_{053} .

3.3.4 Robustification des unités qui changent de NOGA

Une dernière règle au sein de l'algorithme de robustification a été introduite. Si une paire robustifiée (donc : avec un biais conditionnel qui dépasse les seuils décrits en section 3.3.3) contient au moins une entreprise dont la NOGA2 MON est différente de la NOGA2 du cadre, alors le coefficient α_i des deux unités de cette paire est fixé à 1, ce qui veut dire que la robustification est totale pour les deux unités de la paire.

Cette règle a été introduite pour avoir plus facilement des poids de 1 sur les entreprises qui ont à la fois un fort biais conditionnel et une NOGA2 qui change dans le temps. De telles entreprises sont en effet potentiellement problématiques car les publications (extrapolations) de l'OFS sont réalisées au niveau NOGA2 MON alors que le calage est effectué au niveau des NOGA2 du cadre. Ainsi, les entreprises dont la NOGA2 MON et cadre sont différentes et qui ont des poids strictement supérieurs à 1 "représentent" d'une certaine façon d'autres entreprises de la même NOGA2 cadre, alors que leur réponse est décomptée, pour les publications, dans leur NOGA2 MON. Ce point problématique est évité autant que possible par la règle décrite ici et par la constatation suivante. Les entreprises dont la NOGA2 a changé ont tendance à avoir un résidu ε_i élevé dans la régression linéaire décrite en section 3.3.2 (étape 2), du fait d'une réponse qui a lieu dans une NOGA2 MON différente de celle attendue par les variables de calage. Le résidu élevé génère un biais conditionnel élevé si le poids avant robustification est strictement supérieur à 1, ce qui favorise la robustification. Ce point justifie l'intérêt d'appliquer l'algorithme de robustification au niveau des NOGA2 MON et non au niveau des NOGA2 cadre.

Pour la WS 2021, les algorithmes de robustification utilisant les variables t_{010} , t_{017} , t_{022} , t_{053} mènent à la robustification de respectivement 14, 76, 16 et 4 unités, soit 110 unités au total. Sur ces 110 unités, il y a 52 robustifications totales et 58 robustifications partielles. La robustification est totale en raison de cette règle spécifique liée au changement de NOGA2 pour 9 paires d'unités. Le nombre d'unités dont les poids sont robustifiés en raison de la réponse t_{017} est plus important que pour les variables t_{010} , t_{022} et t_{053} . Comme expliqué en section 3.3.3, l'idée sous-jacente était d'accorder davantage d'importance à t_{017} par rapport aux autres variables. De futurs changements sont cependant possibles. Par exemple, une robustification qui utiliserait directement la variable $t_{029} = t_{017} - t_{022}$ serait envisageable. La variable t_{029} est en effet la valeur ajoutée et c'est cette variable qui donne son nom à l'enquête. Une robustification utilisant la variable t_{029} pourrait limiter l'impact d'unités dont la valeur ajoutée est singulière (très élevée ou au contraire très faible) et qui ne sont pas déjà robustifiée sur la base de leur réponse t_{017} ou t_{022} .

L'ancienne méthode de pondération intégrait déjà un mécanisme de robustification qui précédait le calage. Ce mécanisme menait aussi à des poids qui étaient rapprochés de 1 (sans jamais descendre sous le seuil de 1) lorsque les réponses s'écartaient d'une valeur attendue, calculée selon une régression. Lors de la WS précédente (WS 2020), avec l'ancienne méthodologie, environ 10% des poids de l'échantillon net ont été robustifiés. En comparaison, le nouvel algorithme modifie environ 0.7% des poids de l'échantillon net, ce qui est une propriété désirable pour garder un contrôle renforcé sur les risques de biais.

3.3.5 Impact de la robustification sur les estimations

Le fait de robustifier les poids a un impact sur les coefficients de variations (CV) des estimateurs du total des réponses à l'enquête (voir section 4.3) mais aussi sur ces estimateurs eux même. Pour mesurer l'impact de l'algorithme de robustification dans une situation où un calage est effectué après celle-ci, des poids d'extrapolation ont été calculés avec le même modèle de

réponse que celui décrit en section 3.2, sans appliquer de robustification, et en effectuant le calage décrit en section 3.4. Ces poids permettent d'estimer des totaux $\hat{T}_y^{\text{norob}}$ des différentes variables d'enquêtes y qui peuvent être comparés aux totaux $\hat{T}_y$ estimés en présence de robustification.

Pour une variable d'enquête y , l'impact relatif de la robustification est défini comme la quantité $(\hat{T}_y - \hat{T}_y^{\text{norob}})/\hat{T}_y$. Ces impacts sont donnés par NOGA2 en annexe de ce rapport, voir table 15, pour les variables d'enquête t010 (chiffre d'affaires selon l'enquête), t017 (production brute), t022 (consommations intermédiaires), t053 (masse salariale) et t029 (valeur ajoutée). Pour certaines NOGA2, l'impact de la robustification sur les totaux des variables t017 ou t022 est fort, typiquement supérieur à 5% mais reste faible pour la valeur ajoutée t029 = t017 - t022. L'impact sur la variable t029 demeure faible pour toutes les NOGA2 à l'exception de la NOGA2 36 (captage, traitement et distribution d'eau). Cette dernière NOGA2 a cependant des valeurs monétaires relativement faibles par rapport aux autres NOGA2 (nettement moins de 0.1% de la production brute ou de la valeur ajoutée totale estimée en Suisse).

3.4 Calage

3.4.1 Motivations du calage

Les étapes de la pondération décrites précédemment (modèle de réponse et robustification des poids) permettent d'obtenir des poids, notés d_i , pour chaque unité i de l'échantillon net S . Ces poids pourraient en théorie être utilisés pour fournir des estimations approximativement sans biais, avec une variance estimée donnée par l'équation (5). Tout comme pour la pondération d'autres enquêtes de l'OFS (par exemple l'enquête sur la structure des salaires, voir [Chevalier & Qualité \[2021\]](#)), la présence d'informations auxiliaires peut être utilisée avantageusement pour faire baisser les variances des estimateurs tout en gardant un biais faible. Cette utilisation fait intervenir un calage des poids d_i qui sont transformés en des poids calés w_i . Ceux-ci permettent de retrouver des totaux connus T_x , sur le cadre de sondage, sur des variables auxiliaires x , c'est-à-dire que l'équation suivante s'applique :

$$\sum_S w_i x_i = T_x. \quad (7)$$

Les variables de calage sont ici écrites sous la forme d'un vecteur x de longueur p , où p est le nombre de variables de calage. Pour la WS, les totaux T_x sont calculés en considérant le cadre de l'enquête et en excluant les VIU répondantes, puisque ces dernières ne font pas partie du calage ($w_i = 1$). Les VIU non-répondantes ne reçoivent pas de poids mais entrent en compte dans le calcul du total T_x .

Comme expliqué en section 3.3.1, un estimateur de la variance du total d'une réponse à l'enquête y est donné par l'équation (5). En présence de robustification et de calage, les poids avant robustification d_i^* sont remplacés par les poids après robustification (mais avant calage) d_i et les réponses y_i sont remplacées par des résidus ε_i obtenus à l'aide d'une régression linéaire pondérée, ce qui donne l'équation :

$$\widehat{\text{Var}}(\hat{T}_y) = \sum_{i \in S} d_i(d_i - 1)\varepsilon_i^2. \quad (8)$$

Dans la régression linéaire pondérée, l'unité i reçoit un poids qui dépend du poids après robustification d_i , du g-weight obtenu avec le calage (voir section 3.4.3 pour une définition) et du facteur q_i , décrit en section 3.4.3. La variable régressée est la réponse y si le total est estimé sur toute

la Suisse, ou bien la réponse y multipliée par la variable indicatrice d'appartenance de l'unité i au domaine D considéré, si le total est estimé sur un domaine D , par exemple une NOGA2. Les régresseurs sont les p variables de calage.

L'équation (8) montre que la variance estimée du total peut être réduite si les carrés des résidus ε_i^2 sont proches de zéro, autrement dit si les p variables de calage permettent de bien prédire linéairement la réponse y_i dans le domaine D considéré. Lors de la précédente révision de cette enquête, un calage était déjà effectué. Pour la présente révision, la disponibilité de nouvelles informations auxiliaires permet d'améliorer le calage et la méthode utilisée s'inspire fortement des révisions apportées dans l'enquête sur la structure des salaires, voir [Chevalier & Qualité \[2021\]](#). D'une part, de nouvelles variables de calage sont introduites, voir section 3.4.2. D'autre part, le nouveau calage permet de garantir des poids w_i supérieurs ou égaux à 1, ce qui est une propriété désirable, voir section 3.4.3.

3.4.2 Variables de calage

L'ancienne méthode de pondération utilise EPT_cadre comme principale variable de calage, voir [Assoulin \[2014\]](#), section 3.3. Le calage est effectué par NOGA2, ce qui signifie que les poids permettent, dans chaque NOGA2, de retrouver les totaux sur le cadre de la variable EPT_cadre. Le choix de cette variable est justifié par la bonne corrélation entre EPT_cadre et plusieurs réponses à l'enquête, voir les tableaux par NOGA2 dans [Chevalier \[2024\]](#), annexe B avec notamment les corrélations EPT_cadre et les réponses t010 (chiffre d'affaires selon l'enquête) et t029 (valeur ajoutée).

Pour la révision, le calage de la variable EPT_cadre par NOGA2 est conservé. De plus, les variables suivantes sont ajoutées au calage :

- chiffre d'affaires TVA par NOGA2 (73 variables de calage) et par grande région (7 variables de calage),
- nombre d'unités par NOGA2 (73 variables de calage),
- revenus AVS total sur toute la Suisse (1 variable de calage),
- uniquement pour la NOGA2 46 : nombre d'unités dans chaque classe de taille c1t_x distincte (i.e. 1, 2, 3, 4, 5, 9, 9.1 : 7 variables de calage).

Le chiffre d'affaires TVA calé est celui de l'année précédente à la date de référence de l'enquête (donc : le chiffre d'affaires TVA de 2020 pour la WS 2021) avec les éventuelles corrections implémentées lors de l'élaboration du plan de sondage, voir [Chevalier \[2024\]](#), section 3. L'utilisation de cette variable dans le calage est naturelle, compte tenu de sa forte corrélation avec les réponses à l'enquête. En plus des NOGA2, le calage du chiffre d'affaires TVA se fait au niveau des grandes régions pour améliorer la précision des estimateurs sur des domaines faisant intervenir cette variable.

Le calage du nombre d'unités par NOGA2 n'était pas réalisé dans le passé mais était presque atteint du fait de la méthodologie appliquée pour prendre en compte la non-réponse. Le modèle de réponse utilisé était en effet basé sur des taux de réponse empirique par NOGA2 croisé avec la classe de taille c1t5 de l'entreprise. Avec le nouveau modèle de réponse qui prend en compte de nouvelles variables auxiliaires (chiffre d'affaires TVA, revenus AVS), mais qui n'utilise pas le croisement NOGA2 $\times$ c1t5, le calage du nombre d'unités n'est pas atteint sans ajouter cette variable au calage.

Le revenu AVS total d'une entreprise est une variable bien corrélée avec l'une des réponses à l'enquête : la variable t053 (masse salariale selon l'enquête). Il serait intéressant d'effectuer un

calage de cette variable au niveau NOGA2 et pas uniquement sur toute la Suisse. Cependant, le nombre important de variables de calages qui interviennent déjà au niveau NOGA2 (nombre d'unités, EPT_cadre et chiffre d'affaires TVA) complique cette tâche. Ainsi, pour le moment, le calage à l'aide de cette variable est réalisé uniquement sur la Suisse entière, ce qui correspond à une seule variable de calage au lieu de 73.

Enfin, le calage du nombre d'unités dans la NOGA2 46 par classe de taille `clt_x` a été introduit compte tenu des spécificités observées dans cette NOGA2. Le plan de sondage comporte déjà un certain nombre d'exceptions liées à cette NOGA2, voir [Chevalier \[2024\]](#). Pour la pondération, le calage du nombre d'unités par classe de taille `clt_x` est introduit de sorte à ce que les répondants de la NOGA2 46 "représentent" les autres entreprises (de la NOGA2 46) de classe de taille similaire. En particulier, il est souhaité que les entreprises en classe de taille 9 et 9.1 qui répondent représentent les autres entreprises dans cette NOGA2 de classe de taille 9 et 9.1 respectivement. Il est rappelé que, dans la NOGA2 46, les classes de taille 9 et 9.1 correspondent aux entreprises avec un facteur de productivité TVA supérieur à 10 et un chiffre d'affaires TVA avant traitement respectivement inférieur et supérieur à CHF 1 milliard. Dans la NOGA2 46, les entreprises à forte productivité TVA peuvent afficher des différences importantes entre leur réponses `t010` et `t017`, d'où l'intérêt d'un calage par classe de taille. Un calage par croisement entre NOGA2 et classe de taille serait en théorie possible pour toutes les NOGA2 et pas seulement la NOGA2 46. Cependant, ceci mènerait à un nombre de variables de calage bien plus important et des cellules de calage avec peu d'unités, ce qui rend la convergence du calage plus difficile voire impossible et augmente le risque de biais introduit par le calage.

Le nombre total de variables de calage est ici de 234. Dans la pratique, le calage peut être difficile à atteindre dans certaines NOGA2. Cela se produit notamment lorsque le calage doit faire évoluer le total pondéré du chiffre d'affaires TVA et de EPT_cadre dans des sens opposés, c'est à dire lorsque le calage doit faire augmenter la somme pondérée du chiffre d'affaires TVA tout en faisant diminuer la somme pondérée de EPT_cadre (par exemple, si des unités à très forte productivité TVA ne répondent pas), ou inversement. La difficulté dans ce cas est la corrélation importante entre chiffre d'affaires TVA et EPT_cadre qui peut rendre le calage impossible. Si le calage n'est pas atteint dans une NOGA2 donnée, il est nécessaire de renoncer à caler (pour cette NOGA2 uniquement) l'une de ces deux variables : soit EPT_cadre, soit le chiffre d'affaires TVA. Dans ce cas, le choix se porte toujours en faveur du chiffre d'affaires TVA qui est en moyenne mieux corrélé aux réponses que EPT_cadre. Cela signifie que, pour cette NOGA2, le chiffre d'affaires TVA sera calé, mais pas EPT_cadre. Pour la WS 2021, ceci se produit pour deux NOGA2 distinctes parmi 73 NOGA2. Il s'agit des NOGA2 12 (fabrication de produits à base de tabac) et 35 (production et distribution d'électricité, de gaz, de vapeur et d'air conditionné). Ces deux NOGA2 sont de taille modérée (environ 20 000 EPT au total sur toute la Suisse).

3.4.3 Méthode de calage

Le calage consiste à transformer les poids avant calage d_i en poids calés w_i de sorte à satisfaire l'équation de calage (7) et ce en modifiant le moins possible les poids avant calage (voir [Deville & Särndal \[1992\]](#)). Pour la WS 2021, le calage est réalisé en utilisant le programme Calmar de l'Insee [\[Sautory & Le Guennec, 2003\]](#) en utilisant la distance "logit".

Le calage utilise des facteurs individuels de poids q_i (voir [Deville & Särndal \[1992\]](#), équation 1.2) fixés à $1 / \text{EPT_cadre}$. Par ailleurs, en s'inspirant des innovations proposées dans l'enquête sur la structure des salaires (voir [Chevalier & Qualité \[2021\]](#), section 3.4), les poids après calage sont forcément supérieurs ou égaux à 1. Ceci est implémenté en reprenant l'idée de caler les "poids moins 1", au lieu de caler les poids sans transformation préalable.

Les g-weights sont définis comme le ratio w_i/d_i et dépendent de l'unité i considérée. Lors du calage, des bornes multiplicatives sont imposées. Celles-ci obligent l'algorithme de calage à multiplier le poids d_i par un facteur compris entre deux bornes définies par l'utilisateur. Ces bornes permettent généralement d'avoir un contrôle sur les g-weights et en particulier de fixer des valeurs minimales et maximales pour ceux-ci. Dans le cas présent, les bornes s'appliquent au "poids moins 1", c'est-à-dire au ratio $(w_i - 1)/(d_i - 1)$. Pour la WS 2021, les poids $(d_i - 1)$ sont multipliés par un facteur compris en 0 et 6, la valeur 0 correspondant à un poids calé w_i de 1.

Le fait d'appliquer un facteur multiplicatif sur le poids $(d_i - 1)$ revient à garantir un poids supérieur ou égal à 1. Une conséquence importante est que les poids initialement proches de la valeur 1 ne peuvent pas être fortement modifiés par le calage. Pour certaines unités, le poids avant calage d_i est égal à 1, par exemple : suite à une robustification totale. Dans ce cas-là, le poids calé w_i ne peut pas s'écarter de la valeur 1 puisque le calage applique un facteur multiplicatif sur une distance à 1 qui est nulle. Ce point peut être vu comme un avantage dans la mesure où l'algorithme de robustification vise justement à rapprocher de 1 les poids des unités avec de forts biais conditionnels, voir section 3.3. Dans ce cas-là, le fait que le calage ne puisse pas inverser l'effet produit par la robustification est un point positif. L'inconvénient principal du calage des poids moins 1 est le même que celui observé et décrit dans l'enquête sur la structure des salaires [Chevalier & Qualité, 2021]. La marge de manœuvre réduite qui est donnée à l'algorithme de calage, en ne lui permettant pas de transformer fortement les poids proches de 1, rend le calage plus difficile et fait que celui-ci n'est pas toujours atteint. Pour la WS 2021, hormis les difficultés décrites en section 3.4.2 pour caler à la fois le chiffre d'affaires TVA et EPT_cadre, le calage des poids moins 1 fonctionne et il n'est pas nécessaire de recourir à des méthodes plus complexes telles que celles décrites dans Chevalier & Qualité [2021], notamment le calage des poids moins 1 pour certaines unités et le calage des poids non-modifiés pour d'autres.

Les g-weights w_i/d_i ont une moyenne de 0.9994, proche de 1. Le calage atteint ainsi le but de modifier les poids d'entrée (robustifiés) le moins possible, en moyenne. Parmi les 14 920 unités de l'échantillon net dont le poids est calé (c'est à dire les 15 186 unités répondantes sans les 266 VIU répondantes), 20 ont un g-weight inférieur à 0.25 (valeur minimale : 0.011) et 17 ont un g-weight supérieur à 2.5 (valeur maximale : 4.70). La table 1 donne des statistiques sur les poids à chaque étape de la pondération. Les poids de sondage sont obtenus en prenant l'inverse de la probabilité de sélection dans l'échantillon brut, $1/\pi_i$. Les poids après non-réponse (Poids après NR) sont obtenus par la formule de l'équation (1). Les poids après robustification (Poids après rob.) correspondent aux poids d_i robustifiés, avant d'appliquer l'algorithme de calage. Enfin, les poids finaux utilisés pour les publications sont les poids après calage, w_i .

4 Résultats

4.1 Taille de l'échantillon net

Les tables 2, 3 et 4 donnent des statistiques sur le nombre d'unités présentes dans l'échantillon brut et net de la WS 2021 par NOGA2, classe de taille c1t5, grande région et canton. La valeur dite "espérée" de la taille de l'échantillon net correspond au nombre d'unités espérées dans l'échantillon net en appliquant les hypothèses de taux de réponse mises en place lors de l'élaboration du plan de sondage, c'est-à-dire respectivement 50%, 58%, 67%, 80% et 90% en classe de taille c1t5=1, 2, 3, 4 et 5. Le taux de réponse attendu des VIU est de 100%. Les écarts-types, σ , sur

TABLE 1 Statistiques (moyenne, minimum, percentiles à 10%, 25%, 50%, 75%, 90%, maximum) sur les poids à chaque étape de la pondération.

	Poids sondage	Poids après NR	Poids après rob.	Poids après calage
moyenne	5.77	10.9	10.8	10.7
minimum	1	1	1	1
p10	1	1.09	1.09	1.07
p25	1	1.19	1.19	1.17
p50	1	1.58	1.57	1.55
p75	5.04	8.95	8.68	9.11
p90	25	40.6	40.6	39.8
maximum	25	440	440	439

les valeurs espérées sont calculés sous une hypothèse d'indépendance (de la propension à répondre) entre les différentes entreprises.

Les probabilités de réponse estimées à l'aide du modèle de la section 3.2 n'interviennent pas dans ces tables. Les chiffres présentés permettent de contrôler la validité des hypothèses effectuées sur les taux de réponse (en fonction de la classe de taille) lors de l'élaboration du plan de sondage et d'expliquer partiellement les différences entre les précisions attendues lors de l'élaboration du plan et les précisions estimées lors de la pondération. La colonne "net-esp." donne l'écart entre la taille réalisée et espérée de l'échantillon net, exprimé en nombre d'écarts-types σ . Pour les table 2 et 4, cette colonne permet d'identifier les NOGA2 et cantons où la propension à répondre est meilleure ou moins bonne que ce qui était espéré en faisant des hypothèses uniquement basées sur la classe de taille. Pour la table 3, cette colonne permet de vérifier la plausibilité des hypothèses effectuées lors du plan de sondage (taux de réponse par $c1t5$ significativement différent ou non des hypothèses) et (table de droite) d'identifier si des disparités se dégagent par grande région.

Les tables 2 et 4 montrent une influence de la NOGA2 et du canton sur le taux de réponse puisque ceux-ci sont parfois significativement différents de ce qui était espéré en utilisant uniquement des hypothèses basées sur la classe de taille des entreprises. Ceci justifie l'introduction de ces variables dans le modèle de réponse, même si la grande région est utilisée à la place du canton. La table 3 montre que les entreprises en classe de taille $c1t5=1$ et 5 répondent un peu plus souvent que prévu. Les taux de réponse utilisés dans le plan de sondage pour ces classes de taille (respectivement 50% et 90%) sont légèrement plus pessimistes que les taux de réponse observés (respectivement 52.3% et 92.1%) à hauteur de 3.9 et 2.4 écart-types. À l'inverse, les autres classes de taille répondent un peu moins que prévu, avec un écart maximal de 1.7 écart-types pour la classe de taille 4. Ces écarts peuvent être considérés comme modérés dans le sens où ils ne sont pas significatifs statistiquement sous une hypothèse de normalité. Ainsi, les taux de réponse utilisés lors de l'élaboration du plan de sondage sont globalement raisonnables, voire légèrement conservateurs (taille d'échantillon net légèrement supérieure à la taille espérée). Les résultats par grande région de la table 3 suggèrent d'utiliser cette variable dans le modèle de réponse, ce qui est fait en pratique. Des résultats sur l'influence des grandes régions sur la probabilité de réponse sont donnés en section 4.2.5.

TABLE 2 Taille de l'échantillon brut (en nombre d'unités) par NOGA2. La valeur "espérée" correspond à l'espérance du nombre d'unités dans l'échantillon net compte tenu des hypothèses sur les taux de réponse établies lors du plan de sondage. Un calcul de l'écart-type σ est également effectué. La valeur "net" correspond à la taille observée de l'échantillon net. La valeur "net-esp." correspond à la différence entre la taille réalisée et espérée de l'échantillon net. Cette différence est exprimée en nombre d'écart-types σ .

NOGA2	brut espéré $\pm 2\sigma$		net net-esp.		NOGA2	brut espéré $\pm 2\sigma$		net net-esp.	
0809	115	71 $\pm$ 10	69	-0.4	50	100	61 $\pm$ 9	63	0.3
10	600	424 $\pm$ 22	333	-8.4	51	82	50 $\pm$ 9	39	-2.7
11	147	91 $\pm$ 11	69	-3.9	52	232	163 $\pm$ 13	180	2.6
12	10	8 $\pm$ 3	6	-1.3	53	147	90 $\pm$ 11	73	-3.0
13	129	83 $\pm$ 11	93	1.9	55	451	323 $\pm$ 18	302	-2.3
14	92	53 $\pm$ 9	46	-1.5	56	861	531 $\pm$ 27	428	-7.5
15	73	42 $\pm$ 8	45	0.8	58	143	96 $\pm$ 11	102	1.1
16	356	222 $\pm$ 18	189	-3.7	59	311	175 $\pm$ 17	153	-2.6
17	120	78 $\pm$ 10	65	-2.6	60	74	48 $\pm$ 8	55	1.8
18	188	120 $\pm$ 13	105	-2.3	61	123	78 $\pm$ 10	95	3.4
19	6	5 $\pm$ 1	6	1.3	62	624	421 $\pm$ 22	487	6.0
20	220	156 $\pm$ 13	184	4.3	63	170	105 $\pm$ 12	102	-0.5
21	164	116 $\pm$ 11	131	2.7	68	739	438 $\pm$ 26	505	5.2
22	223	158 $\pm$ 13	150	-1.2	69	394	229 $\pm$ 19	238	1.0
23	164	110 $\pm$ 12	116	1.1	70	560	372 $\pm$ 21	425	5.0
24	129	89 $\pm$ 10	84	-0.9	71	749	482 $\pm$ 25	578	7.6
25	464	340 $\pm$ 18	272	-7.4	72	248	166 $\pm$ 14	182	2.2
26	498	369 $\pm$ 19	338	-3.4	73	223	149 $\pm$ 14	172	3.4
27	195	139 $\pm$ 12	161	3.7	74	272	159 $\pm$ 16	151	-1.0
28	460	348 $\pm$ 18	346	-0.2	75	172	95 $\pm$ 13	98	0.4
29	112	67 $\pm$ 10	51	-3.1	77	185	116 $\pm$ 13	121	0.7
30	119	71 $\pm$ 10	60	-2.1	78	294	219 $\pm$ 14	200	-2.6
31	147	97 $\pm$ 11	79	-3.2	79	153	100 $\pm$ 11	105	0.9
32	181	122 $\pm$ 12	117	-0.9	80	177	118 $\pm$ 12	98	-3.4
33	169	103 $\pm$ 12	91	-2.0	81	472	322 $\pm$ 19	302	-2.1
35	398	255 $\pm$ 18	310	6.0	82	227	151 $\pm$ 14	150	-0.1
36	93	51 $\pm$ 9	63	2.6	85	397	282 $\pm$ 17	272	-1.2
37	116	71 $\pm$ 10	71	0.0	87	337	258 $\pm$ 15	279	2.9
38	171	112 $\pm$ 12	124	2.1	88	320	219 $\pm$ 16	227	1.0
39	19	11 $\pm$ 4	11	-0.1	90	483	284 $\pm$ 21	259	-2.4
41	543	382 $\pm$ 20	345	-3.6	91	50	31 $\pm$ 6	33	0.7
42	187	131 $\pm$ 12	121	-1.6	92	32	25 $\pm$ 4	27	0.9
43	1 215	782 $\pm$ 32	675	-6.7	93	348	228 $\pm$ 17	201	-3.2
45	491	298 $\pm$ 21	329	3.0	94	238	158 $\pm$ 14	149	-1.3
46	2 066	1 400 $\pm$ 41	1 705	14.9	95	135	73 $\pm$ 11	54	-3.4
47	1 142	732 $\pm$ 31	750	1.2	96	332	192 $\pm$ 17	182	-1.1
49	535	387 $\pm$ 20	389	0.2	Total	22 912	15 099 $\pm$ 137	15 186	1.3

TABLE 3 Taille de l'échantillon brut et net (espéré et réalisé) en nombre d'unités par classe de taille et grande région.

c1t5	brut	espéré $\pm$ 2σ	net	net-esp.	GR	brut	espéré $\pm$ 2σ	net	net-esp.
1	7 546	3 773 $\pm$ 87	3 944	3.9	1	4 024	2 611 $\pm$ 58	2 522	-3.1
2	3 045	1 766 $\pm$ 54	1 752	-0.5	2	4 474	2 966 $\pm$ 60	3 010	1.5
3	3 598	2 411 $\pm$ 56	2 375	-1.3	3	2 916	1 955 $\pm$ 48	1 953	-0.1
4	7 284	5 827 $\pm$ 68	5 770	-1.7	4	4 257	2 858 $\pm$ 58	2 917	2.0
5	1 172	1 055 $\pm$ 21	1 079	2.4	5	3 252	2 132 $\pm$ 52	2 178	1.8
VIU	267	267 $\pm$ 0	266		6	2 869	1 872 $\pm$ 49	1 967	3.9
					7	1 120	704 $\pm$ 31	639	-4.2

TABLE 4 Taille de l'échantillon brut et net (espéré et réalisé) en nombre d'unités par canton.

canton	brut	espéré $\pm$ 2σ	net	net-esp.
AG	1 578	1 053 $\pm$ 36	1 065	0.7
AI	53	34 $\pm$ 7	33	-0.2
AR	129	83 $\pm$ 10	91	1.6
BE	2 583	1 707 $\pm$ 46	1 747	1.8
BL	696	467 $\pm$ 24	462	-0.4
BS	642	435 $\pm$ 22	426	-0.8
FR	661	434 $\pm$ 23	439	0.5
GE	1 471	962 $\pm$ 35	946	-0.9
GL	126	84 $\pm$ 10	78	-1.1
GR	543	353 $\pm$ 21	351	-0.2
JU	198	133 $\pm$ 13	122	-1.8
LU	1 157	769 $\pm$ 31	791	1.4
NE	433	291 $\pm$ 19	294	0.3
NW	136	87 $\pm$ 11	93	1.0
OW	116	75 $\pm$ 10	80	1.0
SG	1 458	960 $\pm$ 35	981	1.2
SH	265	172 $\pm$ 15	188	2.2
SO	599	401 $\pm$ 22	408	0.6
SZ	504	324 $\pm$ 21	340	1.6
TG	678	448 $\pm$ 24	456	0.7
TI	1 120	704 $\pm$ 31	639	-4.2
UR	74	49 $\pm$ 8	48	-0.1
VD	1 812	1 181 $\pm$ 39	1 158	-1.2
VS	741	469 $\pm$ 25	418	-4.0
ZG	882	568 $\pm$ 27	615	3.4
ZH	4 257	2 858 $\pm$ 58	2 917	2.0

4.2 Interprétations sur le modèle de réponse

4.2.1 Statistiques descriptives et odds ratios

Le modèle de réponse décrit en section 3.2 est analysé en détail dans cette section et des interprétations sont données. Des statistiques descriptives sur les probabilités de réponse obtenues avec le modèle sont données en figure 1. Cette figure montre que les probabilités de réponse

prédites par le modèle varient fortement avec la classe de taille `clt5` et que le modèle est capable d'attribuer des probabilités de réponse un peu plus fortes pour les unités répondantes (boîtes rouges) que non-répondantes (boîtes grises).

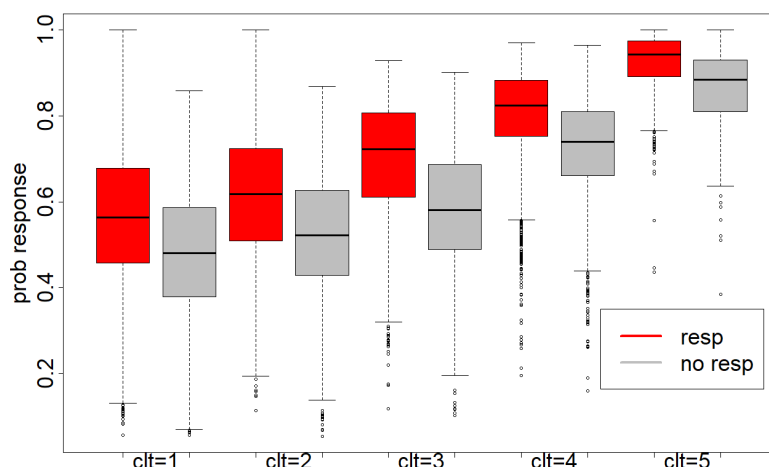


FIGURE 1 Probabilités de réponse prédite par le modèle par classe de taille `clt5` pour les unités répondantes (boîtes rouges) et non-répondantes (boîtes grises).

Les analyses et interprétations sur le modèle de réponse données dans cette section se font en analysant les coefficients β_j estimées dans l'équation (3) et en particulier en étudiant l'odds ratio, $\exp(\beta_j)$. Concrètement, une variable catégorielle avec n_m modalités est classiquement prise en compte dans le modèle (3) en créant $n_m - 1$ variables indicatrices (donc : $n_m - 1$ colonnes dans la matrice de design de la régression logistique), égales à 0 ou 1 selon que l'observation considérée est dans la catégorie correspondante. Chaque variable indicatrice est associée à une modalité. L'unique modalité qui n'a pas de variable indicatrice associée, et donc de coefficient β_j correspondant dans l'équation (3), est la modalité de "référence" et peut être choisie arbitrairement sans influence sur les probabilités de réponse obtenues dans le modèle (3). Les observations avec la modalité de référence ont une valeur de 0 pour l'ensemble des $n_m - 1$ indicatrices de la variable catégorielle considérée. À l'inverse, les observations avec l'une des $n_m - 1$ modalités pour lesquelles un coefficient β_j est associé ont une valeur de 1 pour la modalité considérée et 0 sinon. Dès lors, une application simple de l'équation (3) permet de calculer un rapport de probabilité, $\exp(\beta_j)$, qui compare les observations qui ont la modalité associée au coefficient β_j par rapport aux observations qui ont la modalité de référence. Le rapport ne porte pas directement sur la probabilité de réponse $\hat{p}_i$ mais sur les "odds" $\hat{p}_i/(1 - \hat{p}_i)$, d'où le terme usuel odds ratio.

Les coefficients β_j estimés dans le modèle (3) sont interprétés à l'aide de l'odds ratio correspondant $\exp(\beta_j)$. Un odds ratio supérieur à 1 (respectivement inférieur à 1) indique des probabilités de réponse en moyenne plus élevées (respectivement, plus faibles) pour les observations qui ont la modalité considérée, par rapport à celles qui ont la modalité de référence. Les odds ratios permettent ainsi une comparaison entre les différentes modalités d'une variable catégorielle. Une variable catégorielle fortement influente sur la probabilité de réponse se traduit par l'existence de modalités avec des odds ratios significativement différents de 1.

Les sections suivantes analysent les variables catégorielles sélectionnées dans le modèle (3) et donnent des interprétations à la lumière des odds ratios obtenus. Tous les paramètres et odds ratios du modèle sont donnés et interprétés à l'exception de la constante (intercept) et

des 73 paramètres relatifs à la variable catégorielle NOGA2. Pour des interprétations par NOGA2, les tables 9 et 10 contiennent des résultats sur le taux de réponse empirique par NOGA2.

4.2.2 Interprétations relatives à la classe de taille

La première variable catégorielle sélectionnée dans SAS par la procédure “logistic” avec l’option “stepwise” est la classe de taille à 5 modalités, *c1t5* décrite en fin de section 2.1. La classe de taille *c1t5*=5 (c’est-à-dire les entreprises avec *betot_cadre* supérieur ou égal à 250) est sélectionnée automatiquement par SAS comme modalité de référence. Comme des taux de réponse plus élevés sont observés pour les entreprises de plus de 250 employés, il serait logique d’observer des odds ratios inférieurs à 1 pour l’ensemble de classes de taille 1, 2, 3 et 4. Cependant, la situation est complexifiée par le fait que d’autres variables faisant intervenir *betot_cadre* sont intégrées au modèle. En particulier, la variable quantitative *betot_cadre* est directement intégrée au modèle, croisée avec la variable (catégorielle) *c1t3* pour permettre des effets linéaires différents selon la *c1t3*. Un odds ratio peut être calculé pour les 3 variables ainsi créées : il correspond à la progression relative des odds $\hat{p}_i/(1 - \hat{p}_i)$ pour chaque employé (*betot*) supplémentaire.

La table 5 donne les coefficients β_j et odds ratios associés, pour les prédicteurs mentionnés ici. Comme l’interprétation individuelle de ces coefficients est assez difficile, un graphique est donné en figure 2. Celui-ci présente le taux de réponse empirique observé avec les données de la WS 2021 (courbe noire) en fonction de *betot_cadre*. La courbe rouge (model1) donne les taux de réponse moyens prédits par le modèle de réponse et la courbe bleue (model2) donne ces taux de réponse pour un modèle de régression logistique qui n’intègre pas le prédicteur *betot_cadre* × *c1t3*. Ces trois courbes sont obtenues à l’aide d’un lissage par noyau Gaussien (estimation par noyau, voir par exemple Parzen [1962]) avec un paramètre de lissage qui grandit progressivement avec la valeur de *betot_cadre* (linéairement, entre 3 et 3’000). La figure 2

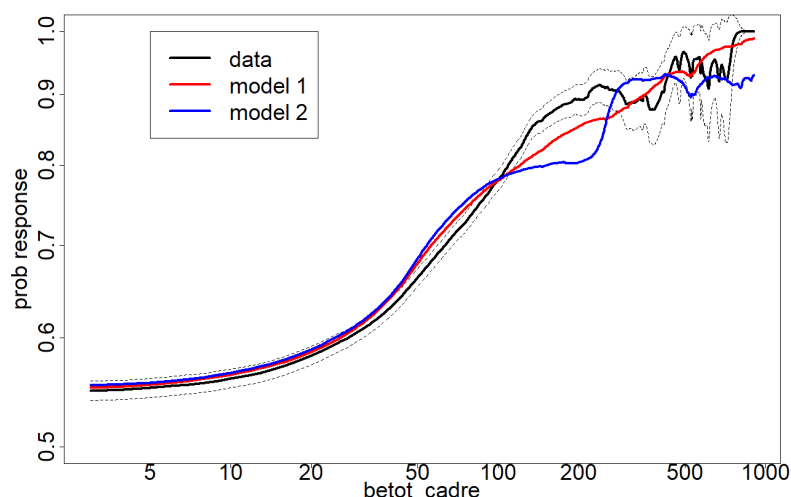


FIGURE 2 Taux de réponse empirique et modélisé en fonction du nombre d’employés (*betot_cadre*). La courbe noire est calculée à partir des données de la WS 2021 (réponse à l’enquête). Les lignes pointillées noires correspondent aux intervalles de confiances (probabilité de réponse ± 2 écart-types). La courbe rouge correspond aux probabilités de réponse anticipées par le modèle utilisé dans ce rapport, et la courbe bleue aux probabilités de réponse d’un modèle sans le prédicteur *betot_cadre* × *c1t3*. Ces trois courbes sont obtenues à l’aide d’un lissage par noyau Gaussien.

TABLE 5 Coefficients – en lien avec la variable `betot_cadre` – estimés dans le modèle de réponse de l'équation (3). Les odds ratios associés aux variables `c1t5 = 1, 2, 3, 4` s'entendent par rapport à une unité en classe de taille 5. Les autres odds ratios s'entendent comme une progression des odds pour chaque `betot_cadre` supplémentaire. La dernière colonne donne la moyenne des probabilités de réponse prédites par le modèle pour chaque `c1t5`. Cette moyenne correspond aussi au taux de réponse empirique dans chaque `c1t5`. Le taux de réponse moyen dans la modalité de référence (`c1t5=5`) est de 92.1%.

modalité ou variable	coefficient	odds ratio	moyenne $\hat{p}_i$
<code>c1t5 = 1</code>	-0.65	0.52	52.3%
<code>c1t5 = 2</code>	-0.65	0.52	57.5%
<code>c1t5 = 3</code>	-0.68	0.51	66.0%
<code>c1t5 = 4</code>	0.17	1.19	79.2%
<code>betot_cadre × c1t3=1</code>	0.048	1.049	
<code>betot_cadre × c1t3=2</code>	0.016	1.016	
<code>betot_cadre × c1t3=3</code>	0.004	1.004	

justifie le choix de prédicteurs qui prévoient une variation continue des probabilités de réponse avec `betot_cadre`. Avec un choix uniquement basé sur des variables catégorielles (par exemple, en choisissant uniquement la variable `c1t5` dans le modèle, voir courbe bleue sur la figure 2), le modèle produit une courbe avec des probabilités de réponse qui se stabilisent au delà de 250 employés, ce qui ne correspond pas aux observations (courbe noire). Les données de la WS 2021 montrent effectivement des probabilités de réponse qui continuent d'augmenter pour les très grandes entreprises, sans se stabiliser au delà du seuil de 250 employés. Par ailleurs, les 3 prédicteurs `betot_cadre × c1t3` sont hautement significatifs dans le modèle de réponse utilisé.

Finalement, la figure 2 permet également d'identifier de potentielles futures améliorations sur le modèle de réponse, par exemple en croisant la variable `betot_cadre` avec une variable légèrement plus complexe que `c1t3`. La probabilité de réponse prédite par le modèle utilisé ("model 1" sur la figure 2) progresse effectivement avec une pente à peu près constante au delà de `betot_cadre=50`, alors que la courbe du taux de réponse empirique ("data") montre une croissance assez faible des probabilités de réponse au delà de `betot_cadre=250`.

4.2.3 Interprétations relatives aux revenus AVS

La variable catégorielle basée sur les revenus AVS décrite en section 3.2 est sélectionnée dans le modèle de réponse en raison de sa forte significativité. Pour cette variable, la modalité de référence est la modalité "99" qui correspond aux entreprises pour lesquelles aucun revenu AVS n'est retrouvé. Il est possible de ne pas trouver de revenu AVS pour certaines entreprises en raison du décalage temporel qui existe entre la table de revenus AVS et la date de l'enquête. Pour la WS 2021, la table de revenus AVS porte sur des revenus en 2020. Les entreprises récemment installées peuvent ainsi être présentes dans l'échantillon brut de l'enquête sans que des revenus AVS soient retrouvés.

Comme expliqué précédemment, les entreprises pour lesquelles aucun revenu AVS n'est retrouvé répondent nettement moins à l'enquête que les autres. Comme la modalité correspondante est la modalité de référence, les odds ratios associés aux autres modalités ont ainsi tendance à être supérieurs à 1. La table 6 détaille les effectifs associés à chaque modalité de la variable catégorielle basée sur les revenus AVS. Les coefficients estimés, odds ratios associés

et probabilités de réponse moyennes, selon le modèle, sont également donnés. La modalité "99" de référence est présente pour 519 entreprises, ce qui représente 2.29% des entreprises de l'échantillon brut qui sont prises en compte dans le modèle de réponse (22 646 entreprises, c'est-à-dire la taille de l'échantillon brut sans les VIU).

TABLE 6 Effectif (en nombre d'unités), coefficient et odds ratio obtenus dans le modèle de l'équation (3) pour les différentes modalités de la variable "pct_sup50_cat" présentée en section 3.2. La variable pct_sup50_cat est une variable catégorielle qui dépend des revenus AVS de l'unité. Les odds ratios sont tous calculés par rapport à la modalité de référence "99", qui correspond aux entreprises pour lesquelles aucun revenu AVS n'est trouvé. La dernière colonne donne la moyenne des probabilités de réponse prédites par le modèle pour chaque modalité. Cette moyenne correspond aussi au taux de réponse empirique. Le taux de réponse moyen dans la modalité de référence (modalité "99") est de 23.1%, ce qui explique les odds ratios tous supérieurs à 1.

modalité	effectif	coefficient	odds ratio	moyenne $\hat{p}_i$
0.1	1 695	1.38	3.98	61.0%
0.2	1 861	1.51	4.51	64.3%
0.3	2 171	1.61	4.98	60.4%
0.4	1 464	1.71	5.54	66.7%
0.5	1 957	1.79	6.01	66.2%
0.6	1 883	1.75	5.75	68.7%
0.7	2 364	1.96	7.11	74.0%
0.8	2 622	1.99	7.35	77.6%
0.9	1 900	1.85	6.37	80.7%
0 × c1t3=1	2050	1.43	4.18	47.4%
0 × c1t3=2	429	1.13	3.08	44.3%
0 × c1t3=3	305	0.95	2.59	63.9%
1 × c1t3=1	647	1.77	5.87	62.8%
1 × c1t3=2	381	1.75	5.73	74.3%
1 × c1t3=3	398	1.23	3.41	82.9%

Plusieurs interprétations importantes peuvent être faites à partir de la table 6.

- Pour les modalités "0.1", "0.2", jusqu'à "0.9" (qui correspondent respectivement à environ 10%, 20%, ..., 90% des salariés dont le revenu AVS est supérieur au revenu AVS médian sur toute la Suisse), une tendance à la hausse des odds ratios est constatée. Les odds ratios les plus élevés sont associés aux modalités correspondantes aux entreprises avec un taux élevé de salariés dont le revenu AVS est supérieur au revenu AVS médian. Ce lien positif entre des revenus AVS élevés et une probabilité de réponse également plus élevée avait déjà été constaté dans l'enquête Suisse sur la structure des salaires, voir [Chevalier & Qualité \[2021\]](#), section 3.2.2. Lors de l'élaboration du modèle de réponse, il a été envisagé de construire le prédicteur "pct_sup50_cat" en comparant les revenus AVS des entreprises au revenu AVS médian de la NOGA2 de l'entreprise (au lieu du revenu AVS médian sur toute la Suisse). Cette démarche n'a cependant pas permis d'améliorer le modèle de réponse.
- Les entreprises avec strictement moins de 5% de salariés avec un revenu AVS supérieur au revenu AVS médian sont classées dans la modalité 0 × c1t3=1, 0 × c1t3=2 ou 0 × c1t3=3 selon la valeur de c1t3. Comme anticipé en section 3.2, le fait d'avoir un faible taux (moins de 5%) de salariés avec un revenu AVS supérieur au revenu AVS médian produit un effet plus prononcé chez les grandes entreprises (odds ratio de 2.59 lorsque c1t3=3,

ce qui est une valeur bien inférieure à celle observée pour les modalités “0.1”, “0.2”, etc.) que pour les petites entreprises (odds ratio de 4.18 lorsque $c_{1t3}=1$, une valeur proche de celle observée pour la modalité “0.1”).

- De manière similaire, les entreprises avec strictement plus de 95% de salariés avec un revenu AVS supérieur au revenu AVS médian sont classées dans la modalité $1 \times c_{1t3}=1$, $1 \times c_{1t3}=2$ ou $1 \times c_{1t3}=3$ selon la valeur de c_{1t3} . Les odds ratios les plus élevés sont observés chez les petites et moyennes entreprises ($c_{1t3}=1$ et 2). Ce point est plus difficile à interpréter car une diminution des odds ratios est observée entre la modalité “0.9” (entre 85% et 95% d’employés payés au dessus du revenu AVS médian) et la modalité “1” (plus de 95% d’employés payés au dessus du revenu AVS médian). Cette diminution est cependant modérée pour les petites et moyennes entreprises. Pour les grandes entreprises, il est possible qu’une majorité de l’effet sur les probabilités de réponse soit déjà pris en compte par les variables de classe de taille détaillées en section 4.2.2.
- Tous les odds ratios présents dans la table 6 sont supérieurs à 1. Les probabilités de réponse sont ainsi chaque fois plus élevées par rapport à la modalité “99” de référence. Parmi les entreprises ayant la modalité “99”, les taux de réponse empiriques en classe de taille $c_{1t5}=1, 2, 3, 4, 5$ sont respectivement de 17%, 22%, 26%, 52% et 50%. Ces taux sont sensiblement inférieurs aux taux observés par classe de taille pour l’ensemble de l’échantillon brut : 52%, 58%, 66%, 79%, 92%, voir table 3. Des améliorations supplémentaires sur le modèle de réponse pourraient être obtenues en retravaillant la définition exacte de la modalité “99”. En effet ce n’est pas directement le fait qu’une entreprise soit absente des données de revenus AVS qui cause la non-réponse de celle-ci, mais probablement d’autres raisons liées à ce point, telles qu’une installation récente en Suisse. Il existe d’autres causes potentielles pour ne pas retrouver de revenus dans la table de revenus AVS, telles que les fusions d’entreprises ou possibles changements de numéro d’entreprise. Ces événements, ont potentiellement un impact différent sur la probabilité de réponse des entreprises. Ceci pourrait être étudié dans de futurs travaux sur le modèle de réponse.

4.2.4 Interprétations relatives à la productivité TVA

Une variable catégorielle basée sur le logarithme du facteur de productivité TVA est intégrée au modèle de réponse. Le facteur de productivité TVA est défini en section 2.2 comme le ratio entre la productivité TVA de l’entreprise et la productivité TVA médiane observée dans la NOGA2 de l’entreprise. Les modalités de la variable catégorielle sont décrites en section 3.2 et ont été obtenues en observant des taux de réponse empiriques en fonction du logarithme du facteur de productivité TVA à l’aide d’un lissage par noyau Gaussien (voir figure 3) avec un paramètre de lissage h fixé à 0.1. La variable construite a 10 modalités au total et la modalité de référence est celle correspondante aux facteurs de productivité TVA les plus élevés.

La table 7 donne le coefficient, l’odds ratio et la probabilité de réponse moyenne associés à chaque modalité. L’interprétation principale qu’il est possible de tirer est que les odds ratios ont tendance à progresser avec le facteur de productivité TVA, de manière non-linéaire. Les odds ratios présentés sont tous inférieurs à 1 car la modalité de référence est associée aux entreprises avec les facteurs de productivité les plus élevés et qui répondent plus souvent. L’ancienne méthodologie n’intégrait pas les chiffres d’affaires TVA au modèle de réponse ou au calage car celui-ci n’était pas disponible pour toutes les unités du cadre de sondage. Ainsi, implicitement, il était supposé que les entreprises qui ne répondent pas ont des productivités TVA comparables à celles qui répondent. Les résultats présentés ici montrent que ce n’est pas le cas. En associant des probabilités de réponse plus basses aux entreprises peu productives, la nouvelle pondéra-

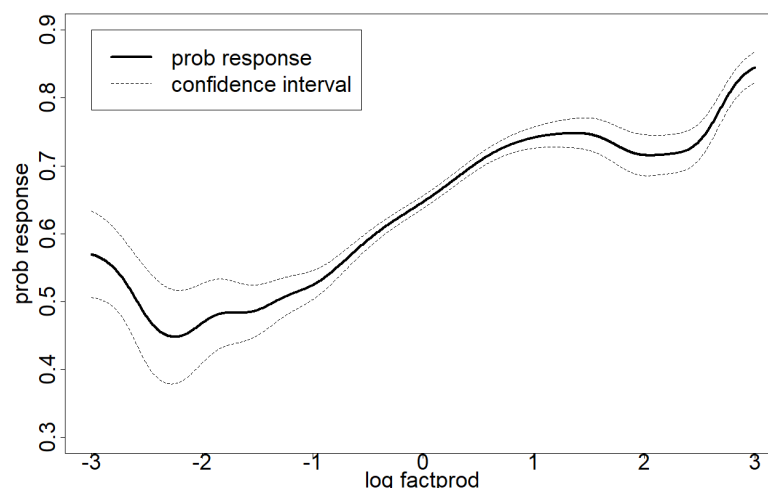


FIGURE 3 Probabilité de réponse empirique (courbe noire) en fonction du logarithme du facteur de productivité. Les probabilités sont calculées à l'aide d'un lissage par noyau Gaussien. Les lignes pointillées noires correspondent aux intervalles de confiances (probabilité de réponse ± 2 écart-types).

tion va tendre à donner des poids plus élevés aux entreprises répondantes et peu productives car celles-ci représentent mieux les entreprises qui ne répondent pas à l'enquête.

Il est possible d'établir un parallèle entre la conclusion obtenue ici (les entreprises répondantes ont en moyenne de meilleures productivités TVA que les non-répondantes) et une conclusion importante obtenue dans l'enquête sur la structure des salaires. Dans [Chevalier & Qualité \[2021\]](#) il avait été constaté que les entreprises qui tendaient à mieux rémunérer les salariés (selon les données AVS) répondaient davantage à l'enquête. Ignorer ce point revenait ainsi à affirmer que les entreprises non-répondantes payent des salaires en moyenne identiques aux entreprises répondantes, ce qui pouvait conduire in fine à une sur-estimation des salaires. Pour la WS, un phénomène comparable se produit car les entreprises répondantes donnent une image biaisée de la productivité des entreprises Suisse, avec ainsi un risque de sur-estimation des valeurs monétaires. Le nouveau modèle de réponse intègre désormais le fait que les entreprises plus productives répondent davantage, réduisant ainsi le risque mentionné.

4.2.5 Interprétations relatives aux grandes régions

La variable catégorielle "grande région" (7 modalités) est sélectionnée en dernier dans le modèle par la procédure "logistic" de SAS avec l'option "stepwise". Les données montrent que la grande région est une information utile pour prédire la probabilité de réponse, même si son effet est moins fort que pour les autres variables décrites précédemment.

Dans le modèle retenu, la grande région est croisée avec la classe de taille à 3 modalités `c1t3`, pour des raisons qui vont être expliquées ci-dessous. Pour faciliter les interprétations, il est utile de connaître les résultats obtenus si la grande région est utilisée dans le modèle de réponse sans ce croisement avec `c1t3`. Dans ce cas-là, SAS sélectionne la grande région 7 (Tessin) comme modalité de référence. Les odds ratios obtenus pour les grandes régions 1 à 6 sont respectivement de 1.07, 1.36, 1.11, 1.15, 1.30, 1.23. Ceci indique que les entreprises des grandes régions 2 (Espace Mitteland) et 5 (Suisse orientale) tendent à répondre un peu plus que le Tessin. Pour les autres grandes régions, la différence avec le Tessin est positive mais plus modérée.

TABLE 7 Effectif (en nombre d'unités), coefficient et odds ratio obtenus dans le modèle de l'équation (3) pour les différentes modalités de la variable "log_factprod_cat" présentée en section 3.2. Les odds ratios sont tous calculés par rapport à la modalité de référence "1", qui correspond aux entreprises avec le plus fort facteur de productivité TVA. La dernière colonne donne la moyenne des probabilités de réponse prédites par le modèle pour chaque modalité. Le taux de réponse moyen dans la modalité de référence est de 76.4%.

modalité	effectif	coefficient	odds ratio	moyenne $\hat{p}_i$
-3	220	-0.87	0.42	58.2%
-2.5	98	-1.31	0.27	44.9%
-2	316	-1.07	0.34	46.8%
-1	2 807	-0.90	0.41	53.2%
0	7 890	-0.58	0.56	62.6%
0.2	2 744	-0.49	0.61	67.1%
0.4	2 175	-0.38	0.68	69.7%
0.6	1 578	-0.24	0.79	73.1%
0.8	999	-0.24	0.79	74.0%

Fort de ces résultats, l'idée de croiser la grande région avec c1t3 est de comprendre si cette tendance à mieux répondre dans certaines grandes régions s'applique à toutes les entreprises ou uniquement à un sous-ensemble identifiable. La table 8 fournit des éclaircissements et montre que la tendance à mieux répondre dans certaines grandes régions s'applique uniquement pour les entreprises avec un betot_cadre supérieur ou égal à 10 (c1t3=2), avec un effet encore renforcé si betot_cadre est supérieur ou égal à 50 (c1t3=3). Les odds ratios obtenus sont en effet proches de 1 et parfois même strictement inférieurs à 1 pour tous les croisements entre grande région et c1t3 lorsque c1t3=1. Les résultats de la table 8 justifient ainsi l'intérêt de croiser la grande région avec c1t3.

TABLE 8 Odds ratios et probabilités de réponse moyennes dans le modèle de l'équation (3) pour les différents croisements grande région $\times$ c1t3. Les odds ratios sont calculés par rapport à des entreprises de classe de taille comparable et dans la grande région 7 (Tessin). Les taux de réponse moyens dans cette grande région pour les classes de taille c1t3=1, 2, 3 sont respectivement de 49.2%, 51.7% et 72.3%.

grande région	odds ratio c1t3=1	odds ratio c1t3=2	odds ratio c1t3=3	moy. $\hat{p}_i$ c1t3=1	moy. $\hat{p}_i$ c1t3=2	moy. $\hat{p}_i$ c1t3=3
1	1.00	1.18	1.10	50.1%	59.4%	77.4%
2	1.16	1.62	1.45	52.7%	64.8%	80.2%
3	0.96	1.12	1.39	51.1%	60.3%	82.3%
4	0.97	1.28	1.35	52.5%	64.1%	82.5%
5	1.08	1.36	1.75	52.2%	62.4%	83.6%
6	1.03	1.30	1.58	57.1%	64.0%	82.9%

4.3 Coefficients de variation d'estimateur du total

Des coefficients de variation (CV) de l'estimation du total de la réponse t017 (production brute) pour la WS 2021 sont présentés dans cette section. Ces coefficients sont définis de manière générale comme le ratio entre l'écart-type (estimé) de l'estimateur du total et l'estimateur du total lui-même. Les tables 9, 10, 11, 12 reproduisent les CV visés et attendus publiés dans le rapport

de méthodes [Chevalier \[2024\]](#). Le CV visé est un objectif de précision minimale à atteindre. Le CV attendu est calculé une fois le plan de sondage élaboré en reprenant les hypothèses sur le taux de réponse élaborées à ce moment là, voir section 4.1. Ces deux CV sont calculés lors du plan de sondage et ne sont donc pas calculés à partir de réponses à l'enquête. Pour avoir une idée des précisions qui seront obtenues sans disposer de réponses à l'enquête, une variable dite "proxy" est utilisée lors du plan de sondage et cette variable est à présent le chiffre d'affaires TVA contrôlé, voir [Chevalier \[2024\]](#), section 4.1. La colonne "taille net" donne la taille de l'échantillon net dans la NOGA2 (cadre), grande région ou canton considéré. Cette information est fournie pour expliquer des CV parfois plus élevés dans les NOGA2, grandes régions ou cantons dont le nombre d'unités est plus faible.

Cette section présente les CV pour la réponse t_{017} alors que d'autres réponses à l'enquête sont importantes, notamment la réponse t_{029} (valeur ajoutée). La raison est que la variable proxy utilisée lors du plan de sondage (chiffre d'affaires TVA contrôlé) est fortement corrélée à la réponse t_{010} (voir [Chevalier \[2024\]](#), Annexe B) et que les réponses t_{010} et t_{017} sont relativement proches dans la majorité des NOGA2. Les CV obtenus pour la variable t_{017} peuvent ainsi être comparés avec les CV anticipés lors du plan de sondage. Les CV obtenus pour la réponse t_{029} peuvent aussi être négatifs car le total de la variable t_{029} peut être négatif, ce qui rend ces CV parfois moins pertinents. Dans la plupart des domaines, ces CV sont néanmoins intéressants et sont donnés en annexe de ce rapport, voir tables 13 et 14.

Les CV "cal. simple" (calage simple), "no rob." (sans robustification) et "obtenus", donnés dans les tables 9, 10, 11, 12, sont des CV de l'estimateur du total de t_{017} sous plusieurs hypothèses. Les variances sont calculés en utilisant l'équation (8), à partir de poids et de variables de calage qui varient selon les hypothèses.

- Le CV avec calage simple est calculé en utilisant les poids d_i obtenus avant le calage (mais après la robustification) et en calculant les résidus ε_i en utilisant les variables de calage de l'ancienne méthode de pondération, c'est-à-dire uniquement la variable EPT_cadre croisée avec la NOGA2.
- Le CV obtenu utilise les mêmes poids d_i mais calcule les résidus ε_i en utilisant toutes les variables de calage décrites en section 3.4.2. La comparaison entre CV avec calage simple et CV obtenu donne un aperçu de l'impact des changements proposés dans cette révision concernant le calage. Les CV obtenus par NOGA2 sont légèrement différents des CV donnés dans les publications officielles de l'OFS pour deux raisons principales. Premièrement la méthode de calcul de la variance est légèrement différente de celle utilisée pour les publications, mais donne en règle générale des résultats très proches (différence de CV inférieure à 0.2%). Deuxièmement, les CV sont calculés par NOGA2 cadre pour pouvoir être efficacement comparés aux CV calculés lors du plan de sondage. À l'inverse, les CV publiés par l'OFS le sont au niveau de la NOGA2 MON.
- Le CV sans robustification est calculé en utilisant des poids différents, puisqu'il s'agit ici des poids d_i^* de l'équation (1). Les poids d_i^* ne subissent pas de robustification. Les résidus ε_i de l'équation (8) sont calculés avec toutes les variables de calage. La comparaison entre CV sans robustification et CV obtenu donne un aperçu de l'impact sur le CV de la robustification des poids décrite dans ce rapport.

Les deux dernières colonnes des tables 9, 10, 11, 12 donnent les taux de réponse attendus compte tenu des hypothèses effectuées lors du plan de sondage et le taux de réponse obtenu lors de l'enquête. Ces colonnes reprennent, sous forme de pourcentage, des informations présentées dans les tables 2 et 3. Ces informations sont données dans le but d'expliquer partiellement les différences qui pourraient exister entre le CV obtenu (lors de la pondération) et le CV attendu (lors du plan de sondage). Des résultats par NOGA2 cadre sont donnés dans les tables 9

TABLE 9 Coefficients de variation (CV), exprimés en pour cent, de l'estimateur du total de la réponse τ_{017} (production brute) par NOGA2 pour les NOGA2 08 et 09 (regroupées) à 49. "Taille net" est la taille en nombre d'unités de l'échantillon net. Les CV visés et attendus sont calculés lors de l'élaboration du plan de sondage. Les CV "calage simple", "sans robustification" et "obtenus" sont calculés lors de la pondération. Le taux de réponse attendu (en pour cent) est calculé compte tenu d'hypothèses sur les taux de réponse effectuées lors de l'élaboration du plan de sondage. Ce taux peut être comparé au taux de réponse obtenu pour expliquer partiellement les différences entre CV obtenu et CV attendu.

NOGA2	taille net	CV visé	CV attendu	CV cal. simple	CV no rob.	CV obtenu	CV obt. -CV att.	Tx rep. attendu	Tx rep. obtenu
0809	69	6.8	5.0	18.5	1.7	1.7	-3.4	61.9	60.0
10	333	2.8	2.2	2.0	0.8	0.8	-1.4	70.6	55.5
11	69	3.4	3.4	3.4	1.2	1.2	-2.2	62.2	46.9
12	6	1.2	1.1	0.9	0.0	0.0	-1.0	75.9	60.0
13	93	3.7	3.7	3.5	1.6	1.6	-2.1	64.3	72.1
14	46	3.9	3.9	10.9	2.2	2.2	-1.7	57.7	50.0
15	45	2.5	2.5	1.6	0.6	0.6	-1.9	57.1	61.6
16	189	4.0	4.0	3.3	2.9	2.0	-2.0	62.4	53.1
17	65	2.9	2.8	2.7	8.6	1.8	-1.0	65.2	54.2
18	105	3.9	3.9	4.7	1.5	1.5	-2.4	63.6	55.9
19	6	0.5	0.3	0.0	0.0	0.0	-0.3	84.7	100.0
20	184	4.4	4.1	4.1	1.0	1.0	-3.1	71.1	83.6
21	131	2.7	2.5	2.7	0.5	0.3	-2.2	70.8	79.9
22	150	4.4	4.3	2.7	1.4	1.4	-3.0	70.9	67.3
23	116	6.0	4.6	4.0	0.7	0.7	-3.8	66.9	70.7
24	84	3.0	2.9	1.6	0.7	0.7	-2.2	68.6	65.1
25	272	3.3	3.3	3.5	1.2	1.2	-2.1	73.3	58.6
26	338	2.6	2.2	5.1	0.7	0.7	-1.5	74.2	67.9
27	161	2.9	2.3	3.0	1.0	1.0	-1.3	71.2	82.6
28	346	2.2	2.1	2.9	0.7	0.7	-1.4	75.7	75.2
29	51	5.1	5.0	8.9	1.0	1.0	-4.0	59.5	45.5
30	60	2.2	2.1	2.4	1.3	1.3	-0.8	59.3	50.4
31	79	3.2	3.2	3.2	2.4	2.4	-0.8	65.8	53.7
32	117	4.0	3.9	10.1	3.3	1.9	-2.0	67.5	64.6
33	91	3.1	3.1	4.4	1.8	1.7	-1.4	61.0	53.8
35	310	3.7	3.6	1.9	1.7	1.7	-1.9	64.0	77.9
36	63	3.6	3.6	2.1	1.1	0.9	-2.7	54.4	67.7
37	71	3.8	3.8	3.3	1.1	1.1	-2.7	61.2	61.2
38	124	9.0	6.6	5.7	2.5	2.0	-4.6	65.2	72.5
39	11	10.8	10.6	9.9	3.0	2.5	-8.2	58.7	57.9
41	345	4.0	4.0	3.5	9.6	1.3	-2.7	70.3	63.5
42	121	2.6	2.5	8.3	7.4	7.2	4.7	70.0	64.7
43	675	3.4	2.3	2.2	1.2	1.2	-1.1	64.4	55.6
45	329	6.3	5.6	2.3	1.5	1.5	-4.0	60.8	67.0
46	1 705	3.1	2.8	1.8	1.7	1.3	-1.4	67.8	82.5
47	750	2.5	2.0	1.1	0.8	0.8	-1.2	64.1	65.7
49	389	3.2	2.9	3.7	0.8	0.8	-2.1	72.3	72.7

TABLE 10 Coefficients de variation (CV), exprimés en pour cent, de l'estimateur du total de la réponse τ_{017} (production brute) pour les NOGA2 50 à 96 et pour la Suisse entière.

NOGA2	taille net	CV visé	CV attendu	CV cal. simple	CV no rob.	CV obtenu	CV obt. -CV att.	Tx rep. attendu	Tx rep. obtenu
50	63	1.3	1.2	0.6	0.2	0.2	-1.0	61.4	63.0
51	39	4.2	4.2	5.1	2.1	2.1	-2.1	61.4	47.6
52	180	10.1	4.8	5.2	3.9	3.0	-1.8	70.2	77.6
53	73	4.2	4.1	1.5	0.6	0.6	-3.5	61.3	49.7
55	302	3.4	3.4	2.7	2.5	1.9	-1.5	71.6	67.0
56	428	3.8	2.5	3.1	1.9	1.9	-0.7	61.6	49.7
58	102	4.0	3.8	5.3	1.1	1.1	-2.8	67.3	71.3
59	153	6.4	6.4	5.4	2.6	2.3	-4.1	56.3	49.2
60	55	4.7	4.7	2.0	0.6	0.4	-4.2	64.6	74.3
61	95	3.9	3.2	2.8	0.9	0.9	-2.3	63.2	77.2
62	487	5.6	5.6	3.5	1.7	1.4	-4.2	67.4	78.0
63	102	5.2	5.2	2.7	1.5	1.5	-3.7	61.7	60.0
68	505	4.3	4.3	4.3	3.8	3.3	-1.0	59.2	68.3
69	238	4.0	4.0	3.3	2.1	1.6	-2.4	58.1	60.4
70	425	9.0	4.9	4.8	3.4	2.6	-2.3	66.5	75.9
71	578	5.5	5.4	4.0	6.6	1.9	-3.5	64.4	77.2
72	182	5.0	4.8	4.6	2.1	1.9	-2.9	66.9	73.4
73	172	6.0	5.5	6.3	3.2	2.3	-3.1	66.6	77.1
74	151	6.0	6.0	6.0	3.9	3.6	-2.4	58.4	55.5
75	98	3.1	3.1	2.7	1.8	1.5	-1.6	55.4	57.0
77	121	12.0	8.9	14.2	1.7	1.5	-7.4	62.9	65.4
78	200	4.0	3.3	3.8	2.1	1.7	-1.6	74.4	68.0
79	105	9.8	5.1	6.6	3.5	3.1	-2.0	65.4	68.6
80	98	2.6	2.6	1.9	1.3	1.3	-1.3	66.8	55.4
81	302	3.8	3.6	3.5	1.7	1.7	-1.9	68.2	64.0
82	150	6.0	4.7	6.4	4.5	3.2	-1.6	66.5	66.1
85	272	4.0	4.0	4.0	4.0	3.4	-0.6	71.1	68.5
87	279	3.3	3.3	1.5	1.4	1.4	-1.9	76.4	82.8
88	227	3.6	3.6	3.4	5.1	3.3	-0.3	68.5	70.9
90	259	3.4	3.4	5.2	4.8	4.8	1.4	58.8	53.6
91	33	10.5	10.5	4.2	3.1	2.6	-7.9	61.5	66.0
92	27	1.1	1.0	1.2	0.8	0.8	-0.1	78.7	84.4
93	201	4.0	3.4	2.7	1.7	1.4	-2.0	65.5	57.8
94	149	6.0	5.0	8.1	4.1	3.4	-1.6	66.3	62.6
95	54	4.0	4.0	5.1	3.3	3.1	-0.9	54.2	40.0
96	182	4.0	4.0	2.7	2.9	1.8	-2.2	57.8	54.8
Suisse	15 186		0.84	0.69	0.49	0.33	-0.51	65.9	66.3

et 10. La dernière ligne de la table 10 contient aussi des chiffres pour toute la Suisse. Plusieurs conclusions peuvent être tirées :

- Les CV obtenus sont généralement plus faibles que les CV attendus et ce même dans les cas où les taux de réponse obtenus sont plus faibles qu'anticipé lors du plan de sondage. La raison principale est que les CV attendus (lors du plan de sondage) ne tiennent pas compte du bénéfice apporté par le calage du chiffre d'affaires TVA, voir [Chevalier \[2024\]](#),

TABLE 11 Coefficients de variation (CV), exprimés en pour cent, de l'estimateur du total de la réponse τ_{017} (production brute) par canton.

canton	taille net	CV visé	CV attendu	CV cal. simple	CV no rob.	CV obtenu	CV obt. -CV att.	Tx rep. attendu	Tx rep. obtenu
AG	1 065	5.0	2.0	2.4	2.5	2.3	0.3	66.7	67.5
AI	33	35.0	32.1	15.4	17.1	15.3	-16.8	63.8	62.3
AR	91	25.0	12.2	10.6	10.4	10.3	-1.9	64.0	70.5
BE	1 747	5.0	2.8	2.2	1.6	1.5	-1.3	66.1	67.6
BL	462	10.0	5.1	3.2	3.0	3.0	-2.1	67.1	66.4
BS	426	5.0	2.2	3.3	1.5	1.5	-0.7	67.8	66.4
FR	439	9.0	5.0	4.4	8.2	3.3	-1.7	65.6	66.4
GE	946	9.0	5.2	1.4	1.4	1.2	-3.9	65.4	64.3
GL	78	15.0	11.3	6.9	6.8	6.8	-4.5	66.3	61.9
GR	351	8.0	5.5	2.4	2.3	2.3	-3.2	65.0	64.6
JU	122	10.0	8.5	7.3	7.3	7.3	-1.2	67.3	61.6
LU	791	9.0	4.3	2.6	3.2	2.4	-1.9	66.5	68.4
NE	294	10.0	4.5	2.1	2.4	2.0	-2.5	67.2	67.9
NW	93	15.0	11.3	12.6	11.9	11.9	0.5	64.3	68.4
OW	80	20.0	16.2	8.5	8.6	8.6	-7.5	64.7	69.0
SG	981	5.0	4.3	2.7	2.3	2.3	-2.0	65.8	67.3
SH	188	10.0	5.7	5.6	5.4	5.3	-0.3	64.7	70.9
SO	408	5.0	3.1	2.8	3.1	2.7	-0.3	67.0	68.1
SZ	340	10.0	7.7	6.2	6.0	6.0	-1.8	64.2	67.5
TG	456	9.0	5.4	4.2	4.1	4.1	-1.3	66.0	67.3
TI	639	9.0	5.2	3.8	3.0	3.0	-2.2	62.9	57.1
UR	48	18.0	15.7	12.7	12.2	12.2	-3.5	65.6	64.9
VD	1 158	5.0	3.3	2.9	2.2	2.1	-1.2	65.2	63.9
VS	418	5.0	4.6	5.1	5.0	5.0	0.3	63.2	56.4
ZG	615	10.0	5.9	3.0	3.0	2.5	-3.4	64.4	69.7
ZH	2 917	5.0	2.5	2.3	1.9	1.5	-1.0	67.1	68.5

TABLE 12 Coefficients de variation (CV), exprimés en pour cent, de l'estimateur du total de la réponse τ_{017} (production brute) par grande région.

grande région	taille net	CV visé	CV attendu	CV cal. simple	CV no rob.	CV obtenu	CV obt. -CV att.	Tx rep. attendu	Tx rep. obtenu
1	2 522	4.0	3.1	1.4	0.8	0.7	-2.4	64.9	62.7
2	3 010	4.0	1.8	1.4	1.5	0.6	-1.2	66.3	67.3
3	1 953	4.0	1.4	2.1	0.8	0.8	-0.6	67.1	67.0
4	2 917	4.0	2.5	2.3	1.9	1.5	-1.0	67.1	68.5
5	2 178	4.0	2.1	1.8	1.3	1.3	-0.8	65.6	67.0
6	1 967	4.0	3.3	1.9	1.5	1.1	-2.2	65.3	68.6
7	639	6.0	5.2	3.8	3.0	3.0	-2.2	62.9	57.1

section 4.1. De ce fait, le CV attendu a tendance à davantage se rapprocher du CV "calage simple" qui lui aussi ne tient pas compte du calage du chiffre d'affaires TVA. La prise en compte du calage du chiffre d'affaires TVA dès l'élaboration du plan de sondage – pour le calcul de CV – est une question assez difficile qui pourrait faire l'objet de futurs travaux.

- La comparaison entre CV “calage simple” et CV obtenu montre des gains parfois importants obtenus avec le nouveau calage. Ces gains sont liés à la forte corrélation entre réponse à l’enquête et chiffre d’affaires TVA (voir par exemple [Chevalier \[2024\]](#), Annexe B).
- Les CV “sans robustifications” sont parfois identiques aux CV obtenus, mais aussi parfois bien moins bons (c’est-à-dire plus élevés). Ces gains en CV doivent être tempérés par le fait que la robustification introduit potentiellement un biais. Pour la WS 2021, il est rappelé que la robustification des poids concerne 110 unités, ce qui représente 0.7% des poids de l’échantillon net. L’ancienne méthodologie appliquait une robustification des poids pour environ 10% des unités.

La table [11](#) donne des CV par cantons. Certains cantons comportent relativement peu d’entreprises dans l’échantillon net ce qui peut expliquer les disparités dans les CV obtenus. Dans le cas des cantons, le bénéfice du calage et de la robustification est moins fort que pour les NOGA2. Ceci s’explique par le fait que les nouvelles variables de calage telles que le chiffre d’affaires TVA sont croisées avec la NOGA2, mais pas avec le canton. Dans le cas du chiffre d’affaires TVA, un croisement par grande région est effectué. L’algorithme de robustification intervient lui aussi au niveau des NOGA2 et non au niveau canton. Le principal mécanisme en place qui vise à améliorer la précision de l’enquête au niveau canton est l’introduction, lors de la création du plan de sondage, de Very Important Units (VIU) cantons, voir [Chevalier \[2024\]](#), section 5.1.3.

La table [12](#) donne les CV par grandes régions. Les gains apportés par le nouveau calage sont plus marqués au niveau grande région qu’au niveau canton puisque le calage du chiffre d’affaires TVA est réalisé au niveau grande région. Les CV obtenus sont globalement proches de 1% à l’exception de la grande région 7 (Tessin) qui comporte moins d’unités dans l’échantillon net, ce qui cause des CV légèrement moins bons.

5 Conclusions et perspectives

La révision de la pondération de la statistique de la valeur ajoutée (WS) améliore la précision des résultats publiés par l'OFS grâce notamment à la modélisation améliorée de la non-réponse, la procédure révisée de robustification des poids et l'utilisation de nouvelles informations auxiliaires pour le calage. Les chiffres d'affaires TVA et les informations sur les revenus AVS sont de nouvelles sources de données qui permettent des gains particulièrement visibles dans le cas d'estimation de totaux par NOGA2 et par grande région. Il résulte de ces travaux une amélioration globale de la qualité qui pourrait être utilisée pour diminuer la charge sur les entreprises et/ou améliorer la précision de l'enquête. Les améliorations apportées proviennent également de choix méthodologiques effectués lors de la construction du plan de sondage, avec par exemple un taux de réponse des VIU qui tend à s'améliorer et à converger vers 100%. Le travail effectué ouvre la voie à des améliorations futures. Par exemple, le modèle de réponse pourrait encore être affiné par une meilleure compréhension des mécanismes de non-réponse des entreprises récemment installées en Suisse ou qui subissent des événements tels que des fusions, rachats, etc. L'utilisation du chiffre d'affaires TVA pourrait également être optimisée en prenant davantage en compte la méthode de calcul de celui-ci (chiffre d'affaires TVA original ou imputé) dans les calculs. Des analyses supplémentaires pourraient être effectuées sur les comparaisons entre ancien et nouvel algorithme de robustification. Finalement, la révision du plan de sondage et de la pondération ouvrent la voie à des travaux supplémentaires pour améliorer les estimations sur des domaines tels que les cantons ou grandes régions, ou encore sur des degrés de granularité plus fins pour les NOGA, par exemple la NOGA à 3 chiffres.

Bibliographie

- ASSOULIN, D. (2014). Methodenbericht : Wertschöpfungsstatistik, Revision 2009 : Statistische Datenaufbereitung und Hochrechnung. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- BEAUMONT, J.-F. & HAZIZA, D. & RUIZ-GAZEN, A. (2013). A unified approach to robust estimation in finite population sampling. *Biometrika* **100**(3), 555–569.
- CHEVALIER, C. (2024). Rapport de méthodes : Statistique de la valeur ajoutée 2021. Révision du plan de sondage. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- CHEVALIER, C. & QUALITÉ, L. (2021). Rapport de méthodes : Enquête suisse sur la structure des salaires 2018. Révision de la pondération. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- DEVILLE, J.-C. & SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* **87**, 376–382.
- HARRELL, F. E. (2001). Regression modeling strategies : with applications to linear models, logistic regression, and survival analysis. Vol. 608. New York : Springer.
- OFFICE FÉDÉRAL DE LA STATISTIQUE (OFS) (2008). NOGA 2008. Nomenclature générale des activités économiques. Introduction. Office Fédéral de la Statistique OFS. <https://dam-api.bfs.admin.ch/hub/api/dam/assets/344515/master>
- PARZEN, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics* **33**(3), 1065–1076.
- SALIBA, J. (2022). Robust regression, MissForest and calibration combined with non-linear optimization with constraints to impute VAT turnover. *UNECE, Conference of European Statisticians, Expert meeting on Statistical Data Editing*, 3–7 October 2022. https://unece.org/sites/default/files/2022-10/SDE2022_S4_Switzerland_Saliba_D.pdf
- 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 Coefficients de variation pour la valeur ajoutée t029

TABLE 13 CV estimés (calage simple, sans robustification et obtenus) de l'estimateur du total de la valeur ajoutée t029, par NOGA2. La valeur "NA" correspond à des CV négatifs, donc pas pertinents. Ces CV sont négatifs car le total de la valeur ajoutée est négatif dans ces NOGA2.

NOGA2	CV cal. simple	CV no rob.	CV obtenu	NOGA2	CV cal. simple	CV no rob.	CV obtenu
0809	20.3	3.6	3.4	50	1.0	0.7	0.7
10	1.3	1.2	1.2	51	2.0	1.7	1.7
11	3.2	2.1	2.1	52	4.2	4.5	4.0
12	0.7	0.3	0.3	53	1.2	1.0	1.0
13	2.5	2.1	2.1	55	4.0	3.4	3.2
14	8.6	2.8	2.8	56	3.7	2.6	2.6
15	3.3	2.6	2.6	58	4.8	2.1	2.1
16	2.2	1.6	1.6	59	9.4	6.6	6.5
17	1.7	2.0	1.6	60	1.1	1.0	0.8
18	2.9	2.4	2.4	61	1.8	1.3	1.3
19	0.0	0.0	0.0	62	2.4	1.9	1.8
20	1.5	0.7	0.7	63	1.6	1.3	1.3
21	1.2	1.4	1.2	68	4.6	4.5	3.9
22	2.9	2.1	2.1	69	3.7	3.0	2.3
23	2.9	1.9	1.9	70	3.3	3.0	2.9
24	1.0	0.9	0.9	71	2.7	2.5	2.4
25	3.6	2.5	2.5	72	10.9	11.8	10.7
26	2.9	1.2	1.2	73	5.1	4.0	3.9
27	1.9	1.6	1.6	74	6.0	5.8	5.0
28	2.3	0.9	0.9	75	3.1	2.4	2.0
29	3.8	2.1	2.1	77	11.5	4.5	4.4
30	2.0	1.1	1.1	78	2.2	2.7	1.9
31	2.6	2.2	2.2	79	NA	NA	NA
32	7.9	2.3	2.0	80	2.0	1.6	1.6
33	3.9	3.3	3.3	81	3.0	1.7	1.7
35	2.1	2.1	2.1	82	5.6	8.0	5.9
36	4.9	5.5	3.3	85	4.6	4.6	4.2
37	3.0	1.8	1.8	87	1.6	1.6	1.5
38	4.4	4.0	3.8	88	4.8	5.8	4.7
39	13.4	6.1	5.6	90	82.5	77.3	77.3
41	1.7	1.7	1.4	91	NA	NA	NA
42	13.4	13.4	13.3	92	1.0	0.7	0.7
43	1.6	1.3	1.3	93	4.5	4.1	4.0
45	2.2	2.0	2.0	94	5.0	5.8	5.0
46	2.1	1.9	1.7	95	6.8	6.8	6.7
47	1.2	1.1	1.1	96	2.7	2.5	2.0
49	6.1	3.2	3.2				

TABLE 14 CV estimés (calage simple, sans robustification et obtenus) de l'estimateur du total de la valeur ajoutée t029, par cantons, grandes régions et pour toute la Suisse.

canton	CV cal. simple	CV no rob.	CV obtenu	grande région	CV cal. simple	CV no rob.	CV obtenu
AG	2.8	2.8	2.7	1	1.7	1.3	1.2
AI	16.7	22.1	16.7	2	1.2	1.0	0.9
AR	13.6	13.5	13.5	3	1.3	1.0	1.0
BE	1.8	1.8	1.6	4	1.8	1.4	1.3
BL	3.4	3.3	3.3	5	2.4	1.9	1.9
BS	1.3	1.2	1.2	6	2.1	1.7	1.6
FR	4.5	3.6	3.5	7	3.9	3.3	3.3
GE	2.3	2.5	2.1	Suisse	0.6	0.5	0.5
GL	7.8	7.8	7.8				
GR	5.0	4.9	4.9				
JU	8.1	8.0	8.0				
LU	3.0	3.0	2.9				
NE	1.7	2.1	1.7				
NW	9.9	9.6	9.3				
OW	9.6	9.5	9.5				
SG	3.0	2.7	2.7				
SH	8.1	8.0	7.9				
SO	4.3	4.9	4.2				
SZ	4.3	4.2	4.2				
TG	5.3	5.2	5.2				
TI	3.9	3.3	3.3				
UR	11.3	11.1	11.1				
VD	2.8	2.5	2.5				
VS	5.3	5.2	5.2				
ZG	3.8	3.6	3.5				
ZH	1.8	1.4	1.3				

B Impact de la robustification sur l'estimation des totaux

TABLE 15 Impact (imp.), exprimés en pour cent, de la robustification sur l'estimation des totaux par NOGA2 pour les réponses t010 (chiffre d'affaires), t017 (production brute), t022 (consommations intermédiaires), t053 (masse salariale) et t029 (valeur ajoutée). La valeur "NA" est indiquée pour la variable t029 lorsque le total estimé de cette variable est négatif dans la NOGA2 considérée.

NOGA2	imp. t010	imp. t017	imp. t022	imp. t053	imp. t029	NOGA2	imp. t010	imp. t017	imp. t022	imp. t053	imp. t029
0809	0.12	0.14	-0.32	0.66	0.89	50	0.00	0.00	0.00	0.02	0.01
10	0.00	0.00	0.00	0.00	-0.01	51	0.00	0.00	0.00	0.01	0.01
11	0.00	0.00	0.00	0.00	0.00	52	-1.12	-1.08	-1.21	-1.34	-0.71
12	0.00	0.00	0.00	-0.01	0.00	53	0.00	0.00	0.00	0.00	0.00
13	-0.01	-0.01	-0.01	0.00	0.00	55	-0.74	-0.72	-0.98	-0.44	-0.43
14	-0.01	-0.01	-0.01	0.01	0.00	56	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	58	0.00	0.00	0.00	0.01	0.01
16	-0.91	-0.93	-1.62	0.12	0.09	59	0.96	0.95	0.95	0.07	0.95
17	-6.02	-6.18	-9.23	-0.89	-0.63	60	0.29	0.12	0.40	0.10	-0.21
18	0.01	0.00	0.01	0.00	0.00	61	0.01	0.01	0.03	0.00	-0.01
19	0.00	0.00	0.00	0.00	0.00	62	-0.34	-0.07	-0.35	-0.06	0.28
20	0.00	0.00	-0.01	-0.01	0.00	63	0.00	0.04	0.08	0.01	0.00
21	-0.01	-0.09	-0.14	0.03	0.01	68	-1.15	-0.79	-0.32	0.11	-1.05
22	0.02	0.02	0.02	0.03	0.02	69	-0.24	-0.29	-0.70	-0.27	-0.10
23	0.00	0.00	0.00	0.00	0.01	70	-0.36	-0.58	-0.84	0.04	0.05
24	0.00	0.00	0.00	0.00	0.00	71	-1.66	-1.38	-3.28	-0.15	0.43
25	0.00	0.00	-0.01	0.00	0.01	72	-0.40	-0.42	-0.75	0.04	1.01
26	0.00	0.00	0.00	-0.01	0.00	73	0.13	0.13	-0.03	0.35	0.38
27	0.00	0.00	0.00	0.01	0.01	74	-0.29	-0.30	0.70	-0.12	-1.36
28	0.04	0.04	-0.02	-0.06	0.13	75	0.43	0.41	0.17	0.26	0.60
29	0.27	0.30	0.58	-0.15	-0.35	77	0.55	0.59	0.54	0.52	0.64
30	-0.01	-0.01	-0.01	0.00	0.00	78	-0.32	-0.39	-1.85	-0.22	-0.15
31	0.01	0.01	0.02	0.00	0.00	79	0.31	0.34	0.29	-1.94	NA
32	-1.40	-1.38	-1.93	-0.83	-0.38	80	0.01	0.01	0.00	0.01	0.01
33	0.37	0.36	0.34	0.25	0.41	81	0.00	0.00	0.00	0.00	0.00
35	0.00	0.00	0.00	0.00	0.01	82	-2.92	-3.03	-5.27	-1.30	-0.37
36	0.81	0.84	-2.78	0.97	6.28	85	-0.99	-0.92	-1.42	-0.47	-0.56
37	0.00	0.00	-0.01	0.00	0.00	87	-0.14	-0.29	-0.36	-0.21	-0.26
38	-0.60	-0.59	-0.81	0.15	-0.18	88	-0.79	-1.42	-3.55	-0.54	-0.42
39	0.14	0.04	-0.94	0.61	1.39	90	0.01	0.01	0.01	0.01	0.03
41	-11.23	-10.80	-16.19	0.59	0.38	91	0.28	0.30	-0.23	-0.06	NA
42	0.06	0.05	0.06	0.01	0.05	92	0.00	0.00	0.00	0.00	0.00
43	0.00	0.00	-0.01	0.01	0.01	93	-0.36	-0.34	-0.61	-0.63	0.46
45	-0.01	0.00	-0.01	0.00	0.00	94	-1.24	-1.10	-1.19	-0.51	-0.94
46	-2.08	0.01	-0.34	0.26	0.33	95	-0.04	-0.06	0.14	-0.16	-0.28
47	0.01	0.01	0.05	0.00	0.00	96	-1.55	-1.62	-4.76	-0.27	0.19
49	0.01	0.01	-0.02	0.01	0.04	Suisse	-1.41	-0.44	-0.76	-0.03	0.07

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. (2025). Statistique de la valeur ajoutée 2021. Révision de la pondération. Numéro OFS : 338-0092

Bianchi, B., Leuenberger, M., Massiani, A., Panchard, C. (2025). Enquête PIAAC 2022-2023. Plan de sondage, pondération, estimation de variance et analyse du biais dû à la non-réponse. Numéro OFS : 338-0091

Meskaldji, D.-E., Panchard, C., Qualité, L., Saliba, J. (2024). Enquêtes thématiques 2018-2023. Plans de sondage, tirages, pondération et estimation de variance. Numéro OFS : 338-0090

Chevalier, C. (2024). Statistique de la valeur ajoutée 2021. Révision du plan de sondage. Numéro OFS : 338-0089

Nedyalkova, D., Assoulin, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalents. FSO number : 338-0088

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Kilchmann, D. (2007). Beherbergungsstatistik Campingplätze. Stichprobenrahmen und Schätzverfahren 2005/06. BFS-Nummer : 338-0047
- Gabler, S., Häder, S. (2007). Haushalts- und Personenerhebungen. Machbarkeit von Random Digit Dialing in der Schweiz. BFS-Nummer : 338-0046
- Ferrez, J., Graf, M. (2007). Enquête suisse sur la structure des salaires. Programmes R pour l'intervalle de confiance de la médiane. Numéro OFS : 338-0045
- Renaud, A. (2007). Harmonisation de la scolarité obligatoire en Suisse (HarmoS). Design général de l'enquête et échantillon des écoles. Numéro OFS : 338-0044
- Potterat, J. (2007). Betriebszählung 2005. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0043
- Hulliger, B. (2006). Umweltschutzausgaben der Unternehmen 2003, Stichprobenplan, Datenaufbereitung und Schätzverfahren. BFS-Nummer : 338-0042
- Renfer, J.-P. (2006). Enquête sur les chiffres d'affaires du commerce de détail. Plan d'échantillonnage et méthodes d'estimation. Numéro OFS : 338-0041
- Salamin, P.-A. (2006). Statistique de l'aide sociale dans le domaine de l'asile. Plan de sondage et extrapolations pour l'enquête pilote 2005. Numéro OFS : 338-0040
- Renaud, A. (2006). Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004. Numéro OFS : 338-0039
- Graf, M. (2006). Swiss Earnings Structure Survey 2002-2004. Compositional data in a stratified two-stage sample : Analysis and precision assessment of wage components. FSO number : 338-0038
- Potterat, J. (2006). Pensionskassenstatistik 2004. Statistische Methoden zur Schätzung der provisorischen Ergebnisse. BFS-Nummer : 338-0037
- Potterat, J. (2006). Kosten und Nutzen der Berufsbildung aus Sicht der Betriebe im Jahr 2004. Stichprobenplan, Gewichtung und Schätzverfahren. BFS-Nummer : 338-0036
- Kilchmann, D. (2006). Vierteljährliche Wohnbaustatistik. Stichprobenplan, statistische Datenaufarbeitung und Schätzverfahren 2005. BFS-Nummer : 338-0035
- Kilchmann, D. (2006). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2004. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. BFS-Nummer : 338-0034
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 2. BFS-Nummer : 338-0033
- Kilchmann, D., Eichenberger, P., Potterat, J. (2005). Volkszählung 2000. Statistische Einsetzungsverfahren Band 1. BFS-Nummer : 338-0032
- Graf, M., Matei, A. (2005). Enquête suisse sur la structure des salaires 2002. La précision du salaire brut standardisé médian. Numéro OFS : 338-0031
- Graf, E., Renfer, J.-P. (2005). Enquête suisse sur la santé 2002. Plan d'échantillonnage, pondération et estimation de la précision. Numéro OFS : 338-0030
- Potterat, J. (2005). Mietpreis-Strukturerhebung 2003. Gewichtung und Schätzverfahren. BFS-Nummer : 338-0029
- Potterat, J. (2005). Landwirtschaftliche Betriebszählung 2003. Schätzverfahren für die Zusatzerhebung. BFS-Nummer : 338-0028
- Renaud, A. (2004). Coverage estimation for the Swiss population census 2000. Estimation methodology and results. FSO number : 338-0027
- Kilchmann, D. (2004). Revision des Schweizerischen Lohnindex. Schätzmethoden der Lohnindices und deren Varianzschätzer. BFS-Nummer : 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS : 338-0025

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 à cet effet plusieurs moyens et canaux et diffuse ses informations statistiques par domaines thématiques.

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

Sélection de publications

Annuaire statistique de la Suisse



L'Annuaire statistique, que publie l'Office fédéral de la statistique (OFS), constitue depuis 1891 l'ouvrage de référence de la statistique suisse. Il résume les principaux résultats statistiques concernant la population, la société, l'État, l'économie et l'environnement de la Suisse.

Statistique: 175 ans au service de l'État fédéral moderne



Cette brochure a été éditée à l'occasion de la 10 000e publication de l'Office fédéral de la statistique. Reprenant dix publications et événements appartenant au passé, elle retrace l'histoire de l'OFS, de ses publications et de la statistique en Suisse. La brochure compte 52 pages et est disponible en quatre langues (français, allemand, italien et anglais).

www.statistique.ch – sources d'informations des plus utiles

Le portail Statistique suisse est un outil moderne et attrayant qui vous donne accès aux informations statistiques actuelles. Nous attirons ci-après votre attention sur les offres les plus prisées.

Catalogue des publications



Presque tous les documents publiés par l'OFS depuis 1860 sont disponibles gratuitement sous forme électronique sur le portail Statistique suisse (www.statistique.ch). Les publications sur papier peuvent être commandées par téléphone au +41 58 463 60 60 ou par e-mail à l'adresse suivante order@bfs.admin.ch.
www.statistique.ch → Statistiques → Catalogue

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

Aperçu des jeux de données publiés par l'OFS



Pour donner une meilleure vue d'ensemble de tous les jeux de données publiés en continu par l'OFS, nous les avons regroupés sur une page distincte. Cette page contient également un aperçu des données de la statistique fédérale déjà accessibles sur la plateforme opendata.swiss ainsi que des données de la National Summary Data Page (NSDP).
data.bfs.admin.ch

Tableau de bord des votations Suisse



Le tableau de bord de l'OFS utilise les données publiques ouvertes (OGD) pour présenter rapidement les résultats des votations sous une forme visuelle et conviviale. Toutes les données sont librement accessibles et disponibles en cinq langues.
abstimmungen.admin.ch/fr/overview

Informations complémentaires

Service de renseignement

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



www.statistique.ch



Facebook



X / Twitter



Youtube



LinkedIn

La pondération de la statistique de la valeur ajoutée (WS) a été révisée en profondeur pour l'enquête de l'année 2021. Les améliorations apportées proviennent de l'utilisation de nouvelles données administratives issues de décomptes TVA des entreprises et de cotisations à l'AVS des personnes travaillant en Suisse. Chaque étape de la procédure (modélisation de la non-réponse, robustification, calage) tire avantage de la disponibilité de ces données. Des gains de précision et donc de qualité sont obtenus pour l'estimation de totaux sur des domaines tels que le domaine d'activité économique NOGA. Ces gains sont détaillés et analysés dans ce rapport. Ils proviennent en grande partie du calage du chiffre d'affaires TVA. Par ailleurs, de nouvelles interprétations sont faites concernant la non-réponse à l'enquête. S'il avait déjà été constaté, pour d'autres enquêtes de l'OFS, que les entreprises qui tendent à mieux rémunérer leurs salariés répondent plus souvent, force est également de constater que les entreprises plus productives ont tendance à plus facilement répondre à l'enquête.

En ligne

www.statistique.ch

Imprimés

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

Numéro OFS

338-0092

ISBN

978-3-303-00769-3

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



Système d'indicateurs MONET 2030

www.statistique.ch → Statistiques → Développement durable
→ Système d'indicateurs MONET 2030

**La statistique
compte pour vous.**

www.la-statistique-compte.ch