

2021

00

Bases
statistiques
et généralités

Neuchâtel 2024

Statistique de la valeur ajoutée 2021

Révision du plan de sondage

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 du plan de sondage

Rédaction Clément Chevalier, OFS
Éditeur Office fédéral de la statistique (OFS)

Neuchâtel 2024

Éditeur: Office fédéral de la statistique (OFS)

Renseignements: Clément Chevalier, OFS, 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 2024
La reproduction est autorisée, sauf à des fins commerciales, si la source est mentionnée.

Numéro OFS: 338-0089

ISBN: 978-3-303-00761-7

Table des matières

Préambule et Résumé	5
1 Introduction	7
2 Variables importantes et cadre de sondage	7
2.1 Nombre d'employés et classe de taille	7
2.2 NOGA, canton et grande région	8
2.3 Chiffre d'affaires TVA et productivité	8
2.4 Unités dans le cadre	12
3 Contrôles des chiffres d'affaires TVA	12
3.1 Motivation et vue d'ensemble	12
3.2 Plafonnement automatique avant les contrôles	13
3.3 Unités avec un grand chiffre d'affaires TVA qui répondent à l'enquête	14
3.4 Unités avec un grand chiffre d'affaires TVA qui n'ont jamais répondu à l'enquête	16
3.5 Unités qui répondent des valeurs élevées aux enquêtes précédentes	18
3.6 Fortes variations dans les chiffres d'affaires TVA	18
4 Hypothèses et calculs liés au plan de sondage	19
4.1 Variable proxy, variance et coefficient de variation	19
4.2 Hypothèses sur le taux de réponse des unités	21
5 Stratification et unités très importantes	21
5.1 Very Important Units	21
5.1.1 Justifications et historique	21
5.1.2 VIU betot	22
5.1.3 VIU NOGA et VIU cantons	23
5.1.4 Candidats VIU et application pratique de ces procédures	25
5.2 Stratification du plan de sondage	26
5.2.1 Ancienne méthode de stratification	26
5.2.2 Nouvelle méthode de stratification, classe de taille modifiée	27
6 Stratégie de calcul des taux de sondage	28
6.1 Objectifs de précision	28
6.1.1 Présentation des objectifs	28
6.1.2 Objectifs de précision au niveau de la NOGA	29
6.1.3 Objectifs de précision sur d'autres domaines	30
6.2 Calculs, contraintes et algorithme d'allocation	31
6.3 Algorithme de tirage coordonné	33
7 Résultats de l'allocation	34
7.1 Cadre, nombre d'unités, stratification	34
7.2 Objectifs de précision par NOGA2	34
7.3 Résultats de l'allocation	34
8 Conclusions et perspectives	40
Bibliographie	41
Annexes	42
A Liste des NOGA2	42
B Corrélations entre variables auxiliaires et réponse à l'enquête t010	45
Rapports de méthodes de la section méthodes statistiques de l'OFS	47

Préambule

La statistique de la valeur ajoutée (en allemand, Wertschöpfungsstatistik, abrégé en "WS") est le résultat d'une enquête par échantillonnage réalisée chaque année par l'Office Fédéral de la Statistique (OFS). L'enquête sur la valeur ajoutée relève des informations auprès d'entreprises Suisses (unités). Ces informations permettent de calculer pour chaque unité répondante des indicateurs tels que le chiffre d'affaires (t010), la production brute (t017), les consommations intermédiaires (t022), la masse salariale (t053) ou la valeur ajoutée (t029). L'objectif principal de l'enquête est d'estimer le total de la variable t029 sur des domaines donnés, tels que les domaines d'activités économiques NOGA. Les résultats de ces estimations sont notamment utilisées dans le calcul du produit intérieur brut (PIB). La section des méthodes statistiques (METH) de l'OFS a été sollicitée afin de réviser le plan de sondage et la pondération de la WS 2021. Le présent rapport se focalise uniquement sur le plan de sondage de la WS, c'est-à-dire sur la construction de l'échantillon des unités à enquêter. La nouvelle méthodologie optimise l'utilisation de données administratives issues notamment de décomptes de la taxe sur la valeur ajoutée (TVA). Ceci permet une meilleure identification des unités les plus importantes pour l'enquête, une meilleure anticipation des précisions obtenues lorsque les réponses à l'enquête sont disponibles et plus généralement une amélioration de l'estimation des caractéristiques cibles. L'auteur remercie les membres de la section METH qui ont relu attentivement le rapport, Daniel Assoulin, Raphaël Jauslin, Daniel Kilchman, Lionel Qualité, 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, Veronica Roth.

Résumé

Le plan de sondage de la statistique de la valeur ajoutée (WS) a été révisé pour l'enquête WS 2021. Des données administratives issues des décomptes de la taxe sur la valeur ajoutée (TVA) sont utilisées de manière extensive pour produire une stratification adaptée et quantifier la précision de l'enquête. Cette révision du plan de sondage permet d'améliorer la précision de l'enquête tout en conservant une taille d'échantillon pratiquement constante. L'amélioration provient d'une sélection plus importante d'unités dont le chiffre d'affaires TVA est élevé. Cette sélection se fait via le choix de "Very Important Units" (VIU) ainsi qu'à travers une stratification revue, qui fait intervenir le chiffre d'affaires TVA. Le calcul des taux de sondage dans chaque strate s'inspire de la méthode développée lors de la révision du plan de sondage de l'enquête Suisse sur la structure des salaires (ESS). Cette méthode permet l'intégration d'objectifs de précision multiples. Le gain d'efficacité obtenu se fait néanmoins au prix de contrôles plus importants sur les chiffres d'affaires TVA, afin de garantir une bonne corrélation entre ceux-ci et les réponses à l'enquête. Pour la WS 2021, l'allocation mène à un échantillon brut de 22 912 unités. Les précisions attendues sont meilleures ou à minima égales à celles ciblées avec l'ancienne méthode, et ce pour chaque catégorie d'activité des entreprises. Des objectifs de précision supplémentaires sont implémentés au niveau des grandes régions, cantons ou par catégories d'activités avec une nomenclature plus fine.

1 Introduction

La statistique de la valeur ajoutée (WS) est une enquête annuelle par échantillonnage au cours de laquelle des unités (entreprises) sont sélectionnées dans un cadre de sondage constitué d'entreprises d'au moins 3 employés et satisfaisant un certain nombre de conditions. La dernière révision du plan de sondage a été effectuée pour l'enquête WS 2009 et est décrite dans [Assoulin \[2013\]](#). Le type de plan de sondage utilisé pour la révision reste inchangé : il s'agit d'un plan de Poisson stratifié, dans lequel un taux de sondage est calculé pour chaque strate. Si ce principe est repris dans la révision, des changements importants sont apportés au niveau de la stratification et de la méthode de calcul des taux de sondage.

Dans l'ancienne méthode, le calcul des taux de sondage est basé sur une allocation de Neyman optimale [\[Neyman, 1934\]](#) en fixant des objectifs de précision par domaine d'activité économique (NOGA). Ces calculs reposent sur une variable dite "proxy", connue sur tout le cadre de sondage, qui devrait être idéalement bien corrélée aux réponses à l'enquête et dont la variabilité dans chaque strate devrait représenter au mieux la variabilité des réponses à l'enquête. Dans l'ancienne méthode, la variable choisie est le nombre d'employés en équivalent plein temps, `EPT_cadre`. Pour la révision, cette variable proxy est le chiffre d'affaires TVA – une variable qui n'était pas encore disponible lors de la dernière révision de la WS. Les avantages liés à l'utilisation du chiffre d'affaires TVA sont décrits tout au long de ce rapport. Ceux-ci peuvent se résumer en deux points principaux. Premièrement, les précisions de l'enquête sont mieux estimées lors de l'établissement du plan de sondage, c'est-à-dire que la différence entre les précisions obtenues une fois les réponses à l'enquête disponibles et les précisions anticipées lors de l'établissement du plan tend à se réduire. Deuxièmement, le nouveau choix de la variable proxy doit permettre de mieux sélectionner dans l'échantillon les unités les plus importantes, ce qui améliore la qualité de l'enquête.

Le rapport est organisé comme suit. La section 2 présente les variables importantes utilisées et résume la méthode de construction du cadre de sondage. La section 3 présente les nombreux contrôles effectués sur les chiffres d'affaires TVA, afin de garantir que leur utilisation soit la plus efficace possible. Les hypothèses de calcul et certains calculs simples de précision qui introduisent la suite des travaux sont présentés en section 4. Par la suite, la stratification et le choix d'unités jugées importantes sont décrits en section 5. La stratégie d'allocation est décrite en section 6 et les résultats de celle-ci sont donnés en section 7.

2 Variables importantes et cadre de sondage

Certaines variables jouent un rôle clé dans l'élaboration du plan et du cadre de sondage. En particulier, l'utilisation du chiffre d'affaires TVA (voir section 2.3) est nouvelle et constitue l'une des clés pour améliorer le plan de sondage de la WS.

2.1 Nombre d'employés et classe de taille

L'OFS dispose d'un registre des entreprises et des établissements (REE). Celui-ci est mis à jour en continu et décrit un peu plus de 700 000 unités à l'aide d'un certain nombre de variables. Pour la WS 2021 – enquête réalisée en 2022 au cours de laquelle l'OFS récolte des données concernant l'année 2021 – c'est l'état de ce registre en décembre 2021 qui est pris en compte. L'une des variables est le nombre d'employés de l'unité et sera nommée `betot_cadre` dans ce rapport. À partir de la variable `betot_cadre`, une variable de classe de taille, `c1t` est directement construite. Les unités sont respectivement en classe de taille 1, 2, 3, 4 et 5 si `betot_cadre` est

respectivement dans les intervalles [3,9], [10, 19], [20, 49], [50, 249] et [250,∞). Cette séparation, avec des intervalles similaires, est reprise dans plusieurs statistiques de l'OFS¹. Les unités avec strictement moins de 3 employés sont exclues du cadre de sondage de la WS, comme expliqué en section 2.4. La classe de taille est une variable importante pour élaborer le plan de sondage. Celle-ci sera utilisée, sous une forme légèrement modifiée, pour la stratification, voir section 5.2.

La variable `betot_cadre` compte le nombre de personnes qui travaillent dans une unité donnée, mais ne tient pas compte du taux d'activité de ces personnes. Le nombre d'employés en équivalent plein temps, noté `EPT_cadre` dans ce rapport, est une variable importante et disponible sur tout le cadre de sondage. Celle-ci est construite dans le cadre de la Statistique Structurale des Entreprises (STATENT², voir Nedyalkova & Assoulin [2024]). Comme expliqué en section 1, `EPT_cadre` était une variable importante pour l'allocation d'échantillon avec l'ancienne méthode. La nouvelle méthode s'appuie davantage sur le chiffre d'affaires TVA. Dans le cadre de la révision, la variable `EPT_cadre` est utilisée principalement pour calculer la productivité d'une unité, en divisant le chiffre d'affaires de l'unité par `EPT_cadre`, voir section 2.3.

2.2 NOGA, canton et grande région

La catégorie d'activité d'une entreprise est définie par son classement dans la nomenclature générale des activités économiques (NOGA, voir OFS [2008]). Concrètement, chaque unité du cadre de sondage est rangée dans une NOGA à l'aide d'une variable à 6 chiffres. Les deux premiers chiffres donnent la NOGA2, qui sera la plus couramment utilisée dans ce rapport. Les NOGA à trois et quatre chiffres qui permettent de distinguer les activités économiques selon une nomenclature plus fine seront moins mentionnées (voir cependant la section 6.1.3 pour une utilisation de la NOGA à trois chiffres) et si le mot "NOGA" est utilisé seul alors il fera toujours référence à la NOGA2. Une description succincte des NOGA2 est donnée en annexe A. Pour la WS 2021, la NOGA 09 ne contient que 6 unités au niveau du cadre de sondage. Pour plusieurs calculs cette NOGA est fusionnée avec la NOGA 08 (195 unités). Ce regroupement de NOGA est le seul regroupement effectué. En tenant compte de ce regroupement, le cadre de sondage de la WS 2021 contient des unités faisant partie de 73 NOGA2 différentes. La NOGA2 ainsi définie est, avec la classe de taille, l'une des variables utilisées pour la construction des strates.

La grande région (GR) d'une unité est une autre variable utilisée pour l'élaboration d'objectifs de précision de l'enquête (voir section 6.1). La grande région correspond au découpage de la Suisse au niveau NUTS 2³. La liste des grandes régions, avec entre parenthèses les cantons qui les composent, est la suivante : Région lémanique (GE, VS, VD), Espace Mittelland (BE, FR, JU, NE, SO), Suisse du Nord-Ouest (AG, BL, BS), Zurich (ZH), Suisse orientale (AI, AR, GL, GR, SG, SH, TG), Suisse centrale (LU, NW, OW, SZ, UR, ZG) et Tessin (TI). Pour la WS 2021, des objectifs de précision de l'enquête sont fixés à la fois par canton et par grande région, voir section 6.1. Le canton de l'unité intervient aussi dans le choix des "Very Important Units" (VIU), voir section 5.

2.3 Chiffre d'affaires TVA et productivité

L'OFS dispose, pour certaines unités, de décomptes TVA qui permettent de déduire le chiffre d'affaires annuel de l'unité. Un tel chiffre d'affaires issu des décomptes TVA sera nommé "chiffre d'affaires TVA" dans ce rapport. Le chiffre d'affaires TVA est une variable particulièrement utile

1. par exemple, dans la Statistiques des groupes d'entreprises : <https://www.bfs.admin.ch/asset/fr/31706906>

2. <https://www.bfs.admin.ch/bfs/fr/home/statistiques/industrie-services/enquetes/statent.html>

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

car il est relié à plusieurs réponses à l'enquête, notamment les variables τ_{010} (chiffre d'affaires) et τ_{017} (production brute). Le chiffre d'affaires TVA, lorsqu'il est disponible, est celui de l'année précédant l'enquête. Ainsi, pour la WS 2021, le chiffre d'affaires TVA le plus récent à disposition est celui de l'année 2020.

Les décomptes officiels de TVA ne sont pas disponibles pour toutes les unités du cadre de sondage. Le chiffre d'affaires TVA est obtenu à partir de décomptes TVA pour environ 85.6% des unités. Ces chiffres d'affaires TVA seront appelés "chiffres d'affaires TVA originaux" dans ce rapport. Pour environ 12.3% des unités, le chiffre d'affaires TVA est dit "imputé", c'est-à-dire prédit à l'aide d'un modèle. Ces unités sont éventuellement des unités exemptes de TVA et/ou des unités dont le chiffre d'affaires est inférieur à CHF 100 000. Finalement, environ 2% des unités font partie d'un groupe d'unités plus grand, pour lequel l'OFS dispose du chiffre d'affaires TVA original seulement au niveau du groupe TVA d'unités. Pour ces dernières, un chiffre d'affaires TVA dit "réparti" est estimé à l'aide d'une méthode qui répartit le chiffre d'affaires TVA agrégé entre les unités du groupe TVA. Les pourcentages d'unités avec un chiffre d'affaires TVA original, réparti ou imputé dans chaque NOGA2 figurent dans la table 1.

Les méthodes de calcul des chiffres d'affaires TVA imputés et répartis sont détaillées dans [Saliba \[2022\]](#) et sont mentionnées succinctement ici. En premier lieu, toutes les unités dont le chiffre d'affaires TVA est manquant reçoivent un chiffre d'affaires TVA imputé. Les unités qui ne font pas partie d'un groupe conservent ce chiffre sans nouvelle modification, alors que celles qui font partie d'un groupe voient leur chiffre d'affaires TVA encore modifié de sorte à assurer une cohérence avec le chiffre d'affaires TVA du groupe. La méthode d'imputation dépend de la classe de taille de l'unité. Pour les unités satisfaisant l'inégalité $\text{betot_cadre} \leq 20$, un algorithme nommé "MissForest" est utilisé [\[Stekhoven & Bühlmann, 2012\]](#). Celui-ci est basé sur des méthodes de classification par "random forest" et se base sur les chiffres d'affaires TVA originaux observés sur d'autres unités ainsi que sur d'autres informations telles que la NOGA2 ou betot_cadre , voir [Saliba \[2022\]](#) section III B pour les détails. Pour les unités avec strictement plus de 20 employés, une régression linéaire robuste est appliquée. Celle-ci se base sur les mêmes données que l'algorithme MissForest. Finalement, si l'unité dont le chiffre d'affaires TVA est manquant fait partie d'un groupe d'entreprises dont le chiffre d'affaires TVA agrégé est connu, les données imputées pour chaque unité doivent être cohérentes avec le chiffre agrégé. Pour y parvenir, un algorithme de répartition effectue un calage en multipliant les chiffres d'affaires TVA des composantes du groupe par des facteurs, pas forcément identiques, et les plus proches possible de 1. Un mécanisme est également mis en place pour éviter que les chiffres d'affaires TVA répartis mènent à des valeurs de productivité (telle que définie dans cette section, ci-dessous) trop importantes par rapport aux productivités observées dans la NOGA2. Le mode de calcul du chiffre d'affaires TVA (original, réparti ou imputé) est important et sera pris en compte notamment lors des contrôles des unités avec un chiffre d'affaires TVA élevé qui ne répondent pas à l'enquête, voir section 3.4.

Une variable importante dérivée du chiffre d'affaires TVA est la productivité. Cette variable sera régulièrement utilisée dans ce rapport. La productivité, dans le contexte de ce rapport, est définie comme le ratio entre chiffre d'affaires TVA et EPT_cadre . Elle peut s'exprimer en millions de CHF par équivalent plein temps. La productivité sera également une variable importante lors des contrôles pour juger de la plausibilité des chiffres d'affaires TVA, voir section 3.4.

Pour déterminer si la productivité d'une unité est élevée ou non en comparaison avec les autres unités du même secteur, le facteur de productivité est une variable calculée pour chaque unité du cadre de sondage de la façon suivante : dans chaque NOGA2, la médiane de la productivité est calculée en prenant en compte les productivités de toutes les unités de la NOGA2 dans le cadre

de sondage. Par la suite, le facteur de productivité d'une unité est simplement le ratio entre sa productivité et la productivité médiane. Une unité avec un facteur de productivité supérieur (respectivement : inférieur) à 1 est ainsi plus (respectivement : moins) productive que la moitié des unités de sa NOGA2. Le facteur de productivité d'une unité est une variable qui sera utilisée lors des contrôles des chiffres d'affaires TVA (voir section 3.6) ainsi que pour la construction des strates du plan de sondage (voir section 5.2).

Le chiffre d'affaires TVA est sans doute l'une des variable les plus importantes utilisées dans cette enquête, pour des raisons détaillées tout au long de ce rapport. D'une manière générale, il est important que le chiffre d'affaires TVA soit linéairement corrélé aux réponses à l'enquête. Dans la grande majorité des cas, les analyses ont montré une très bonne corrélation entre la variable τ_{010} (chiffre d'affaires au cours de l'année, relevé dans l'enquête) et le chiffre d'affaires TVA, voir table 10 pour des données chiffrées. Pour un nombre limité d'unités cependant, le chiffre d'affaires TVA ne prédit pas bien la réponse à l'enquête dans le sens où celle-ci peut être d'un ordre de grandeur très différent et varier d'un facteur mille dans un sens comme dans l'autre. Ce type d'occurrence pénalise la corrélation entre τ_{010} et chiffre d'affaires TVA et limite les bénéfices attendus de par son utilisation. Pour limiter au maximum de telles occurrences, de nombreux contrôles sur les chiffres d'affaires TVA sont effectués. Ces contrôles font partie de l'étape de construction du cadre de sondage et sont détaillés en section 3.

TABLE 1 Proportion d'unités avec des chiffres d'affaires TVA 2020 (umsatz, "ums") originaux, répartis et imputés.

NOGA2	unités	ums orig.	ums rép.	ums imp.	NOGA2	unités	ums orig.	ums rép.	ums imp.
0809	201	94.5%	2.5%	3.0%	50	108	72.2%	3.7%	24.1%
10	2 596	92.7%	1.6%	5.7%	51	102	79.4%	2.9%	17.6%
11	288	86.5%	2.8%	10.8%	52	886	86.5%	6.4%	7.1%
12	10	70.0%	30.0%	0.0%	53	233	88.8%	4.3%	6.9%
13	270	94.4%	1.1%	4.4%	55	3 262	89.9%	2.1%	8.0%
14	236	82.6%	2.1%	15.3%	56	14 522	90.4%	1.0%	8.6%
15	86	93.0%	2.3%	4.7%	58	563	87.4%	4.3%	8.3%
16	2 873	96.9%	0.3%	2.7%	59	627	82.5%	0.6%	16.9%
17	125	95.2%	3.2%	1.6%	60	78	78.2%	7.7%	14.1%
18	892	96.3%	1.7%	2.0%	61	210	88.6%	3.3%	8.1%
19	6	83.3%	16.7%	0.0%	62	5 262	91.1%	1.4%	7.5%
20	441	90.0%	5.7%	4.3%	63	432	82.2%	4.6%	13.2%
21	211	73.5%	11.8%	14.7%	68	5 445	58.1%	2.7%	39.2%
22	532	94.5%	3.4%	2.1%	69	6 810	84.3%	7.0%	8.7%
23	561	95.2%	2.5%	2.3%	70	4 275	80.7%	7.3%	12.0%
24	167	92.2%	4.8%	3.0%	71	9 207	94.4%	1.7%	3.9%
25	3 947	97.0%	1.3%	1.7%	72	804	72.8%	3.1%	24.1%
26	1 320	90.2%	6.8%	3.0%	73	1 261	92.1%	1.5%	6.3%
27	546	93.4%	3.7%	2.9%	74	2 001	89.6%	0.6%	9.8%
28	1 383	93.3%	4.0%	2.7%	75	672	92.4%	2.7%	4.9%
29	131	94.7%	0.8%	4.6%	77	631	90.0%	2.7%	7.3%
30	139	89.9%	5.8%	4.3%	78	1 186	88.9%	5.3%	5.8%
31	394	97.0%	1.3%	1.8%	79	852	75.2%	1.9%	22.9%
32	1 160	92.5%	1.9%	5.6%	80	440	85.9%	0.7%	13.4%
33	1 197	96.7%	0.6%	2.7%	81	5 652	89.7%	0.6%	9.6%
35	440	93.2%	4.3%	2.5%	82	811	82.4%	3.3%	14.3%
36	111	92.8%	0.9%	6.3%	85	3 873	23.3%	1.9%	74.8%
37	181	91.7%	3.9%	4.4%	87	1 076	42.8%	3.9%	53.3%
38	479	94.2%	2.9%	2.9%	88	2 392	5.5%	1.9%	92.6%
39	19	78.9%	5.3%	15.8%	90	1 024	25.6%	0.4%	74.0%
41	3 690	89.5%	2.1%	8.4%	91	65	75.4%	7.7%	16.9%
42	535	91.4%	3.7%	4.9%	92	32	56.3%	37.5%	6.3%
43	18 030	95.9%	0.9%	3.2%	93	3 258	62.9%	0.9%	36.2%
45	6 812	96.3%	0.5%	3.2%	94	1 081	44.3%	2.1%	53.6%
46	9 949	90.7%	2.6%	6.8%	95	308	92.5%	0.3%	7.1%
47	14 508	93.7%	1.0%	5.4%	96	4 226	83.5%	0.7%	15.8%
49	3 671	88.4%	2.0%	9.6%	Total	161 804	85.6%	2.0%	12.3%

2.4 Unités dans le cadre

La construction du cadre de sondage reprend largement les éléments décrits dans le rapport de méthodes élaboré lors de la précédente révision de la WS [Assoulin, 2013]. La liste des unités présentes dans le cadre est obtenue en partant d'un extrait du Registre des Entreprises et des Etablissement (REE), lequel contient des informations sur environ 700 000 unités basées en Suisse. Sur la totalité de ces unités, les variables importantes mentionnées dans ce rapport (telles que `betot_cadre`, `EPT_cadre` ou le chiffre d'affaires TVA) sont disponibles, sans qu'il y ait de données manquantes. Le cadre final de l'enquête, nommé "cadre WS" dans ce rapport, contient 161 804 unités et une description par NOGA2 est donnée dans les tables 2 et 3. Pour faire partie du cadre WS, une unité de l'extrait du REE doit satisfaire les conditions suivantes :

- l'unité doit faire partie du secteur privé ou être une entreprise publique. Les administrations publiques sont exclues,
- certaines NOGA2 sont exclues de l'enquête. En particulier, l'unité doit faire partie du secteur secondaire ou tertiaire,
- en fonction de la NOGA2, seules certaines formes juridiques sont acceptées pour faire partie du cadre WS (les détails sont disponibles auprès de la section Statistique monétaire des entreprises),
- l'unité doit avoir au minimum 3 employés selon l'extrait du REE. La variable déterminante est ici `betot_cadre`.

La règle d'un minimum de 3 employés n'est pas nouvelle. Une discussion sur ce seuil figure dans Assoulin [2013], section 2.1.

3 Contrôles des chiffres d'affaires TVA

3.1 Motivation et vue d'ensemble

Comme mentionné en section 2.3, le chiffre d'affaires TVA est une variable auxiliaire qui possède une bonne corrélation avec plusieurs variables d'enquête importantes, telles que t_{010} (chiffre d'affaires), t_{017} (production brute), t_{022} (consommations intermédiaires) ou t_{029} (valeur ajoutée : $t_{029} = t_{017} - t_{022}$). Des corrélations observées par NOGA2 pour la WS 2021 entre la réponse t_{010} et le chiffre d'affaires TVA de l'année 2020 sont données en annexe, en table 10.

Un des aspects de la révision du plan de sondage de la WS 2021 consiste à utiliser le chiffre d'affaires TVA comme variable "proxy" pour les calculs de variance, voir section 4.1. De plus, toujours depuis l'enquête WS 2021, le chiffre d'affaires TVA est également utilisé dans la pondération au niveau de la modélisation de la non-réponse et du calage. Un rapport de méthodes qui décrit les révisions dans la pondération est en cours de préparation.

Pour améliorer la précision des estimations, la corrélation entre le chiffre d'affaires TVA et la réponse à l'enquête doit être forte, et ce à l'intérieur de chaque NOGA2. Ce point est justifié en section 4, équation (3). Le coefficient de corrélation linéaire de Pearson étant une grandeur fortement influencée par les valeurs extrêmes, il doit être interprété avec précaution. Si aucun contrôle n'est fait, un petit nombre d'unités peut causer une baisse drastique de la corrélation entre chiffre d'affaires TVA et chiffre d'affaires selon l'enquête. Ceci limiterait les bénéfices attendus du fait de l'utilisation du chiffre d'affaires TVA. L'équation (3) permet de déduire un certain nombre de situations à éviter. Par exemple, les unités avec une faible probabilité de faire partie de l'échantillon net de la WS et un écart important, en valeur absolue, entre chiffre d'affaires TVA et chiffre d'affaires selon l'enquête sont susceptibles de générer des variances importantes et ainsi des précisions dégradées.

Un indicateur utile de l'influence d'une unité donnée sur une estimation et sa variance est le biais conditionnel [Moreno-Rebollo et al., 1999, Favre-Martinoz et al., 2018]. Pour cette enquête, un estimateur du total sous un plan de sondage de Poisson, en présence de non-réponse et de calage, est utilisé. Dans ce contexte, pour une unité i répondante, le biais conditionnel est

$$\text{biais}_i^{\text{cond}} = \left(\frac{1}{\pi_i p_i} - 1 \right) \varepsilon_i,$$

où les quantités π_i , p_i et ε_i sont respectivement le taux de sondage de l'unité i , sa probabilité de réponse et un résidu de régression linéaire, voir section 4.1. Si l'unité i ne répond pas, son biais conditionnel est égal à $-\varepsilon_i$. Les unités avec un fort écart entre chiffre d'affaires (t_{010}) et chiffre d'affaires TVA ont un résidu ε_i élevé en valeur absolue, ce qui augmente la variance – sauf dans le cas d'une unité répondante qui aurait à la fois un taux de sondage π_i et une probabilité de réponse p_i de 1.

Comme quelques unités peuvent influencer fortement la précision de l'enquête du fait d'un écart entre chiffre d'affaires (t_{010}) et chiffre d'affaires TVA, les chiffres d'affaires TVA doivent être contrôlés. Il est cependant impossible de vérifier individuellement les données des quelque 160 000 unités du cadre de sondage et des choix doivent être faits sur les unités à contrôler en priorité. Si l'objectif est de garantir une bonne corrélation, il est naturel de s'intéresser en priorité aux valeurs les plus extrêmes au niveau du chiffre d'affaires TVA et au niveau de la réponse à l'enquête (t_{010}). Ainsi, la "partie 1" des contrôles (voir sections 3.3 et 3.4) s'intéresse aux unités qui ont un chiffre d'affaires TVA supérieur à un seuil, qui a été fixé à CHF 100 millions. La "partie 2" (voir section 3.5) s'intéresse aux unités qui ont déjà répondu à l'enquête des valeurs t_{010} supérieures à CHF 100 millions et qui ne seraient pas déjà couvertes dans la partie 1. Enfin, la dernière partie des contrôles (voir section 3.6) se focalise sur les unités dont le chiffre d'affaires TVA a subi une forte variation relative d'une année à l'autre.

Une partie des contrôles effectués se base sur les réponses aux enquêtes précédentes et peut mener à des chiffres d'affaires TVA modifiés sur la base de ces réponses. Cette pratique est potentiellement délicate si l'unité répondante n'avait pas une très forte probabilité d'inclusion dans l'échantillon net. Ce point est discuté en section 3.3.

Les contrôles sur le chiffre d'affaires TVA améliorent la corrélation de celui-ci avec la réponse t_{010} comme le montre la table 10. Cette amélioration est également constatée pour d'autres variables cibles que t_{010} . En particulier, la corrélation entre le chiffre d'affaires TVA et la valeur ajoutée t_{029} s'améliore globalement grâce aux contrôles, voir table 11. Pour la NOGA2 46 (commerce de gros hors automobiles et motocycles⁴), la table 10 montre que les contrôles sur le chiffre d'affaires TVA tendent à faire diminuer la corrélation de celui-ci avec la réponse t_{010} . Ce résultat ne doit pas inquiéter pour deux raisons. Premièrement, la corrélation observée est très dépendante d'un nombre réduit d'unités pour lesquelles, lors de la WS 2021, t_{010} s'avère être en ligne avec le chiffre d'affaires TVA. Ceci n'est pas le cas chaque année et il est déjà arrivé d'observer une corrélation proche de zéro du fait d'une ou deux unités de la NOGA2 46 qui influent fortement la corrélation. Deuxièmement, la réponse t_{010} est une réponse importante à l'enquête, mais la valeur ajoutée t_{029} l'est encore davantage. La table 11 montre que les contrôles dans la NOGA2 46 améliorent la corrélation entre chiffre d'affaires TVA et t_{029} .

3.2 Plafonnement automatique avant les contrôles

Avant d'effectuer des contrôles individuels à proprement parler, une procédure automatique de plafonnement des chiffres d'affaires TVA est mise en place. Des tests ont en effet montré que

4. <https://www.kubb-tool.bfs.admin.ch/fr/code/46>

ces plafonnements augmentaient la corrélation entre chiffre d'affaires TVA et chiffre d'affaires selon l'enquête (t010). Les plafonnements sont mis en place avant les contrôles, ce qui veut dire que certains d'entre eux peuvent potentiellement, au cas par cas, être annulés s'ils sont jugés non-pertinents.

Le plafonnement automatique des chiffre d'affaires TVA se fait par NOGA2, selon la procédure suivante : la productivité de chaque unité est calculée (ratio chiffre d'affaires TVA / EPT_cadre). Par la suite, dans chaque NOGA2, le percentile à 99% de la productivité, noté $prod_{p99}^{NOGA}$, est calculé. Toute unité, de la même NOGA2, dont la productivité dépasse $prod_{p99}^{NOGA}$ voit sa productivité plafonnée à cette valeur : le chiffre d'affaires TVA plafonné de l'unité est égal à son EPT_cadre multiplié par $prod_{p99}^{NOGA}$.

La règle énoncée ci-dessus comporte une exception qui s'applique à la NOGA2 46. Dans cette NOGA2, les analyses ont montré une faible corrélation entre le chiffre d'affaires TVA et la réponse t010 du fait de la présence fréquente de chiffres d'affaires TVA très élevés qui ne sont pas associés à des t010 élevés. Ainsi, pour cette NOGA2 uniquement, le plafonnement décrit ci-dessus est appliqué en utilisant le percentile à 95% de la productivité au lieu de 99%.

Les règles automatiques décrites ci-dessus mènent à un plafonnement pour 1% des unités dans chaque NOGA2 et 5% dans la NOGA 46. Même si ces plafonnements permettent généralement d'augmenter la corrélation entre chiffre d'affaires TVA et réponse t010, ceux-ci restent assez limités et insuffisants à eux seuls. Par exemple, les NOGA2 qui comportent moins de 100 unités dans le cadre de sondage ne subissent aucun plafonnement puisque le percentile à 99% des productivités est égal à la plus forte productivité observée dans la NOGA2. De plus, comme expliqué auparavant, la grande variabilité du chiffre d'affaires TVA peut créer des situations où un faible nombre d'unités peut potentiellement dégrader la corrélation entre chiffre d'affaires TVA et t010. Le plafonnement est ici une solution utile mais parfois trop limitée pour répondre à ce problème. Il peut arriver que le plafonnement ne soit pas opportun, c'est-à-dire que la réponse à l'enquête t010 soit tout de même bien reliée au chiffre d'affaires TVA. Le plafonnement peut aussi être insuffisant ou effectué sur trop peu d'unités. De plus, le plafonnement est un outil qui vient diminuer des chiffres d'affaires TVA jugés trop élevés, mais à ce stade aucune procédure n'est mise en place pour corriger à la hausse des chiffres d'affaires TVA trop faibles. Pour ces raisons, des contrôles étendus ont été appliqués en utilisant parfois les réponses aux enquêtes précédentes.

3.3 Unités avec un grand chiffre d'affaires TVA qui répondent à l'enquête

Une fois les opérations préliminaires de la section 3.2 effectuées, une procédure de contrôle en trois étapes est mise en place. La première étape consiste à contrôler les unités du cadre de sondage dont le chiffre d'affaires TVA initial (c'est-à-dire avant l'éventuel plafonnement décrit en section 3.2) est supérieur à CHF 100 millions. Les unités de la NOGA 46 sont exclues de ces analyses : le plafonnement automatique plus important permettant déjà d'améliorer les corrélations entre chiffre d'affaires TVA et réponses à l'enquête.

L'idée des contrôles de la première étape est de déterminer si les chiffres d'affaires TVA élevés sont effectivement reliés à des réponses élevées lors des enquêtes passées. Le nombre d'unités contrôlées ici est d'environ un millier. Ces unités sont classées en 3 catégories :

- catégorie A : les unités qui n'ont jamais répondu à l'enquête dans le passé (environ 80 unités),
- catégorie B : les unités qui ont répondu dans le passé et pour lesquelles la réponse est jugée en ligne avec le chiffre d'affaires TVA (environ 660 unités),

- catégorie C : les unités qui ont répondu dans le passé et pour lesquelles la réponse n'est pas jugée en ligne avec le chiffre d'affaires TVA (environ 280 unités).

Cette section se focalise sur le classement des unités qui répondent, c'est-à-dire les unités en catégorie B ou C. Les opérations effectuées sur les unités de catégorie A, qui n'ont jamais répondu dans le passé, sont décrites en section 3.4.

Pour le plan de sondage de la WS 2021, les réponses passées considérées sont celles des WS 2020, WS 2019 et WS 2018. Si une unité à contrôler a répondu au moins à l'une de ces enquêtes, alors elle fait partie des catégories B ou C décrites ci-dessus, en fonction de sa réponse. En cas de réponse à plusieurs de ces enquêtes, c'est la réponse la plus récente à l'enquête qui est prise en compte pour fixer une "année de référence" et cette réponse lors de l'année de référence est comparée au chiffre d'affaires TVA de la même année. Si le ratio entre la réponse de référence et le chiffre d'affaires TVA de l'année de référence est dans l'intervalle $[0.7, 1.3]$, alors le chiffre d'affaires TVA est jugé valide : l'unité fait partie de la catégorie B. Pour ces unités là, le chiffre d'affaires TVA le plus récent – c'est-à-dire celui qui est contrôlé – ne subit pas de modification. Si au contraire le ratio décrit précédemment est en dehors de cet intervalle, alors l'unité fait partie de la catégorie C. Dans ce cas, une intervention a lieu sur le chiffre d'affaires TVA qui est remplacé par la réponse de référence, c'est-à-dire la réponse la plus récente à l'enquête. Voici certains détails supplémentaires sur cette procédure :

- Parmi les unités contrôlées et répondantes du passé, les cas où l'unité a répondu à l'enquête de l'année qui précède immédiatement (donc ici lors de la WS 2020) sont les plus fréquents : plus de 90% des unités qui ont répondu dans le passé ont donné une réponse lors de la WS 2020. Pour ces unités là, la classification en catégorie B ou C est simple et sans ambiguïté. Si le chiffre d'affaires TVA, *éventuellement plafonné lors de l'étape décrite en section 3.2*, est bien relié à la réponse à l'enquête (ratio entre 0.7 et 1.3) alors ce chiffre est conservé (catégorie B) et sinon il ne l'est pas et est remplacé par la réponse à l'enquête (catégorie C). Il faut noter que le plan de sondage de la WS 2021 est effectué en début d'année 2022, avec un échantillon tiré au mois de février. Cette opération se déroule avant de disposer des données complètes (réponses à l'enquête) de la WS 2020. Néanmoins, en pratique, les réponses des entreprises à fort chiffre d'affaires TVA sont souvent déjà disponibles.
- Le choix des bornes de l'intervalle $[0.7, 1.3]$ représente un compromis entre la volonté d'avoir des chiffres d'affaires TVA proches des réponses à l'enquête et celle de ne pas modifier les chiffres d'affaires TVA pour une trop grande proportion des entreprises. Parmi les entreprises contrôlées ici qui répondent dans le passé (catégorie B ou C), 28% font partie de la catégorie C avec les bornes choisies. Les analyses montrent que ce taux augmente rapidement si les bornes sont davantage resserrées. Par exemple, le taux d'entreprises en catégorie C monterait à plus de 40% avec les bornes fixées à $[0.85, 1.15]$.
- Pour les unités qui répondent, la réponse de référence est toujours calculée sur la base de la variable t_{010} (chiffre d'affaires selon l'enquête), sauf pour la NOGA 77 (activité de location et location-bail⁵) où c'est la variable t_{017} (production brute) qui est utilisée à la place de t_{010} , sur demande de la section MON.

Pour la WS 2021, environ 280 unités sont classées en catégorie C. Cela implique que la réponse à une enquête précédente est utilisée pour construire une information auxiliaire (le chiffre d'affaires TVA). Le potentiel problème dans cette pratique vient ici du fait que l'enquête n'est pas exhaustive. Le fait de disposer de ces réponses aux enquêtes précédentes est le résultat d'une expérience qui peut ainsi être vue comme "aléatoire" (probabilité de sélection dans l'échantillon net). Or, les calculs de variance et de probabilité d'inclusion effectués ici ne prennent pas ce

5. <https://www.kubb-tool.bfs.admin.ch/fr/code/77>

fait en compte et considèrent l'information auxiliaire comme simplement "disponible", sans que cette disponibilité soit conditionnelle au tirage aléatoire d'un échantillon. Il est possible de partiellement tempérer ces préoccupations pour les grandes unités qui répondent à l'enquête chaque année. En effet, les unités de plus de 50 employés (`betot_cadre`) sont toujours dans l'échantillon brut (taux de sondage de 1) et si l'on constate que certaines de ces unités répondent chaque année, il serait possible de négliger le caractère aléatoire de la disponibilité de cette réponse. Voici quelques chiffres observés :

- Sur les 281 unités en catégorie C, 255 unités ont un chiffre d'affaires TVA imputé sur la base de la réponse à l'enquête qui précède immédiatement (WS 2020). Parmi elles, toutes répondent chaque année lorsqu'elles sont interrogées (en considérant la WS 2020, WS 2019, WS 2018) avec seulement deux exceptions (réponse lors de la WS 2020 et WS 2018, mais pas lors de la WS 2019). Toutes les unités ici, du fait de leur classe de taille, font parties de strates exhaustives, ce qui est une information encourageante car il n'y a ainsi pas de caractère "aléatoire" dans le fait de tirer l'unité dans l'échantillon brut. Le fait de modifier ou non le chiffre d'affaires TVA de ces unités n'a pas d'influence sur la sélection individuelle de ces unités mais en a pour le calcul qui sert à attribuer des taux de sondage aux autres unités, voir section 4.
- 26 unités ont un chiffre d'affaires TVA imputé sur la base d'une réponse à l'enquête antérieure à la WS 2020. Le fait de ne pas disposer pour ces unités de réponse à l'enquête précédente laisse naturellement penser que, pour ces unités, le fait d'obtenir une réponse ne peut pas être considéré comme un événement "certain". Toutefois, il faut noter que ces cas sont finalement assez rares si l'on prend en compte le nombre total d'unités contrôlées ici (environ un millier) ou bien la taille du cadre de sondage (plus de 160 000 unités).

3.4 Unités avec un grand chiffre d'affaires TVA qui n'ont jamais répondu à l'enquête

Parmi les quelque 1000 unités contrôlées avec un chiffre d'affaires TVA (avant plafonnement éventuel) supérieur à CHF 100 millions, un peu moins de 80 unités ne répondent jamais à l'enquête. Ainsi, les contrôles de plausibilité du chiffre d'affaires TVA ne peuvent pas se baser sur des réponses à l'enquête et une autre procédure doit être mise en place.

Le choix d'un chiffre d'affaires TVA adapté pour des unités qui n'ont pas répondu dans le passé (et ne répondront peut être pas non plus à l'enquête actuelle) n'est pas anodin. Au niveau du plan de sondage, ces chiffres d'affaires TVA ont un impact sur les calculs de coefficients de variation et ainsi sur la taille de l'échantillon brut, voir section 4. De plus, lors de l'étape de pondération, un calage des poids se fait sur les chiffres d'affaires TVA après avoir appliqué aux poids une correction de non-réponse. Si l'on considère que les étapes antérieures au calage restent inchangées, la modification du chiffre d'affaires TVA d'une unité non-répondante affecte le calage en augmentant ou diminuant le chiffre d'affaires TVA à "compenser". Le choix de ce chiffre d'affaires TVA a ainsi un impact direct sur les résultats finaux de l'enquête, ce qui justifie une attention particulière.

Les unités avec un chiffre d'affaires TVA supérieur à CHF 100 millions et qui n'ont jamais répondu sont suffisamment rares pour justifier un contrôle individuel à la place d'une procédure automatique. Le contrôle d'une unité donnée se fait en considérant plusieurs informations, telles que la NOGA2 de l'unité, le mode de calcul du chiffre d'affaires TVA (original, réparti, ou bien imputé, voir section 2.3) et des données observées dans la NOGA2 considérée (réponses à l'enquête d'unités similaires et productivité observée par exemple). Les éléments de décision sont listés ci-dessous :

- Fiabilité générale des chiffres d'affaires TVA originaux dans la NOGA2 : dans le cas où le chiffre d'affaires TVA à contrôler est original, le premier contrôle se fait en observant les

réponses à l'enquête d'autres d'unités de la même NOGA2 qui ont aussi un chiffre d'affaires TVA à la fois original et élevé. Si l'on observe des unités dont le chiffre d'affaires TVA n'est pas du même ordre de grandeur que leur réponse, il conviendra d'être prudent, c'est-à-dire que des contrôles supplémentaires (décrits ci-dessous) sont nécessaires. Sinon, il n'y aura pas d'éléments qui remettent en cause la crédibilité du chiffre d'affaires TVA original contrôlé.

- Unités similaires / plus proches voisins dans la NOGA en nombre d'employés : dans le cas où le chiffre d'affaires TVA contrôlé est réparti ou imputé, ou bien dans le cas où il est original mais jugé pas nécessairement fiable lors de l'étape précédente, des contrôles supplémentaires sont effectués. Il est utile d'observer les réponses à l'enquête d'unités de la même NOGA2 et de taille comparable en termes de nombre d'employés, `betot_cadre`. Une variable intéressante à analyser sur les autres unités répondantes de la même NOGA2 est la productivité réelle observée ($\text{ratio } t010 / \text{EPT_cadre}$) de ces "proches voisins". Cette productivité observée peut être comparée avec la productivité de l'unité contrôlée. En cas de forte différence, une correction sur le chiffre d'affaires TVA est opportune, c'est-à-dire que la productivité observée du plus proche voisin peut être imputée à l'unité contrôlée. Il est à noter qu'ici, les analyses basées sur la productivité se font toutes sur des entreprises relativement grandes, de sorte qu'il n'y a pas de raison de remettre en question la valeur du dénominateur `EPT_cadre` dans le calcul de la productivité. Par ailleurs, le terme "proche voisin" fait ici référence à des unités de la même NOGA2 avec un nombre d'employés `betot_cadre` le plus proche possible de celui de l'unité contrôlée. Selon la situation, il peut être jugé que certaines unités contrôlées n'ont pas de voisin "proche" qui réponde aux enquêtes précédentes. Cela se produit s'il n'y a pas d'entreprise répondante de la même NOGA2 avec un `betot_cadre` dans un intervalle centré autour du `betot_cadre` de l'unité contrôlée. La largeur de l'intervalle est fixée à la moitié du `betot_cadre`, ce qui laisse peu d'unités dont le chiffre d'affaires TVA est imputé sur la base de la dernière règle décrite ci-dessous.
- Productivité maximale / productivité forte dans la NOGA2 : en l'absence de "proches voisins" dans la NOGA, il est utile d'observer, toujours dans la même NOGA2, la productivité réelle ($\text{ratio } t010 / \text{EPT_cadre}$) maximale atteinte dans la NOGA2, ainsi qu'un ordre de grandeur pour une productivité "forte", qui serait atteinte par un faible nombre d'unités de la NOGA2. Dans le cas où la productivité de l'unité contrôlée est encore nettement supérieure à ces valeurs, un chiffre d'affaires TVA plus faible, calculé sur la base de ces productivités réelles observées, est imputé. Pour la WS 2021, ceci concerne une trentaine d'unités. Selon la NOGA, l'imputation d'un chiffre d'affaires TVA "élevé" en l'absence d'information peut se faire en conservant le plafonnement automatique décrit en section 3.2, c'est-à-dire en utilisant un percentile à 99% de la productivité observée dans la NOGA. Une autre possibilité est d'utiliser un autre percentile de la productivité, par exemple le percentile à 95%. Compte tenu du faible nombre d'unités dont le chiffre d'affaires TVA est imputé sur la base de ces décisions, il est attendu que la sensibilité aux choix effectués soit faible.

Certaines règles énoncées ci-dessus pourraient paraître arbitraires et la question se pose quant à l'influence des choix effectués sur les résultats finaux. Ces préoccupations peuvent toutefois être modérées par plusieurs points. Premièrement, les unités en catégorie A sont rares. Lorsque le chiffre d'affaires TVA contrôlé est particulièrement élevé et représente un pourcentage élevé du total des chiffres d'affaires TVA dans la NOGA2, alors une attention particulière est portée sur l'unité. La section Statistique monétaire des entreprises (MON) de l'OFS a la possibilité d'effectuer des recherches pour juger de la crédibilité du chiffre contrôlé et ce n'est que dans le cas où ces recherches ne sont pas concluantes qu'il est nécessaire de prendre une décision sur la base des éléments listés ci-dessus. Deuxièmement, il est évident que les règles ci-dessus ne

résolvent pas tous les problèmes et peuvent mener à l'imputation d'un chiffre d'affaires TVA contestable, qui serait éloigné de la valeur de la variable t_{010} telle que la renseignerait l'unité si elle répondait à l'enquête. Le problème est cependant atténué si les variations d'une année à l'autre des résultats ont plus d'importance que leur niveau absolu. Dans ce cas là, un chiffre d'affaires TVA incorrectement imputé aura effectivement un impact sur le niveau absolu des résultats, mais l'impact sur les variations d'une année à l'autre sera plus modéré si la règle d'imputation reste stable.

3.5 Unités qui répondent des valeurs élevées aux enquêtes précédentes

Pour garantir une bonne corrélation entre chiffre d'affaires TVA et réponse t_{010} , il est nécessaire de contrôler les chiffres d'affaires TVA élevés (voir sections 3.3 et 3.4) mais aussi les unités dont les réponses t_{010} sont élevées. La seconde partie des contrôles concerne ainsi les unités présentes dans le cadre WS 2021 qui ont répondu des valeurs t_{010} supérieures à CHF 100 millions lors de l'enquête WS précédente et dont le chiffre d'affaires TVA est strictement inférieur à CHF 100 millions. Cette dernière condition vise simplement à ne pas contrôler à nouveaux des unités qui sont déjà contrôlées selon la procédure de la section 3.3.

Du fait de cette dernière condition, le nombre d'unités à contrôler est ici relativement limité : environ 80 unités pour la WS 2021. Le contrôle est simple et se déroule selon la même procédure que celle décrite en section 3.3. Cette procédure est bien automatisée : les unités sont classées en catégorie B (chiffre d'affaires TVA adéquat et donc pas modifié) ou bien C (chiffre d'affaires TVA corrigé à l'aide de la réponse à l'enquête) selon que le ratio entre t_{010} et le chiffre d'affaires TVA se trouve dans l'intervalle $[0.7, 1.3]$ ou non.

3.6 Fortes variations dans les chiffres d'affaires TVA

La troisième et dernière étape des contrôles sur les chiffres d'affaires TVA porte sur les unités qui enregistrent une forte variation, d'une année à l'autre, de leur chiffre d'affaires TVA.

L'idée est ici de faire contrôler par la section MON des unités qui n'auraient pas déjà été contrôlées lors des étapes précédentes, qui auraient de fortes variations à la hausse ou à la baisse de leur chiffre d'affaires TVA et qui pourraient poser des soucis importants dans les estimations dans le cas où les variations affichées du chiffre d'affaires TVA ne reflètent pas les variations réelles de la réponse t_{010} .

Il existe une dissymétrie dans l'ampleur des problèmes potentiellement causés par ce type de forte hausse ou de forte baisse du chiffre d'affaires TVA. Dans l'élaboration du plan de sondage, les chiffres d'affaires TVA élevés ne reçoivent pas le même traitement que les chiffres d'affaires bas. Les mécanismes mis en place dans la révision de la WS 2021 (voir notamment la section 5.2) font que les unités à forte productivité ont tendance à recevoir un poids faible lors de l'étape de pondération, rendant les erreurs moins problématiques en cas de réponse à l'enquête voir section 4.1, équation (3) pour une justification mathématique. Ceci justifie que les fortes hausses des chiffres d'affaires TVA posent généralement moins de soucis que les fortes baisses. La dissymétrie vient aussi du fait qu'une forte hausse du chiffre d'affaires TVA pour une unité peut mener à un chiffre supérieur à CHF 100 millions, ce qui signifie que l'unité a déjà été contrôlée lors des opérations précédentes, à l'inverse des unités dont le chiffre d'affaires TVA devient soudainement très faible.

Ainsi, une forte baisse du chiffre d'affaires TVA qui ne reflète pas les variations de la réponse t_{010} peut généralement poser plus de problèmes qu'une forte hausse sous ces conditions. Les

unités dont le chiffre d'affaires TVA baisse fortement sont susceptibles de faire partie de strates qui recevront des taux de sondage faibles, avec un minimum potentiel à 4% (voir, section 6). En cas d'erreur sur le chiffre d'affaires TVA, l'unité est susceptible d'amener un fort biais conditionnel (voir section 3.1 pour une brève présentation) du fait de son poids élevé et de sa réponse en décalage avec le chiffre d'affaires TVA.

Pour clore les contrôles sur les chiffres d'affaires TVA, la section METH produit une liste d'unités dont le chiffre varie fortement d'une année à l'autre. Cette liste est transmise à la section MON qui effectue des recherches. Compte tenu du temps et des ressources disponibles, la liste doit rester relativement petite, avec si possible environ une dizaine d'unités à vérifier. Les critères de sélection dans la liste sont les suivants :

- Une unité fait partie de la liste des fortes hausses du chiffre d'affaires TVA si son facteur de productivité (voir section 2.3 pour une définition) est supérieur à 10 et que la productivité a au moins triplé par rapport à l'année précédente. Les unités de la NOGA 46 sont exclues de la liste des fortes hausses à cause de la fiabilité plus faible des chiffre d'affaires TVA dans cette NOGA qui est traitée à l'aide des plafonnement automatiques, voir section 3.2.
- Une unité fait partie de la liste des fortes baisses du chiffre d'affaires TVA si son facteur de productivité de l'année précédente était supérieur à 10, que son facteur de productivité de l'année courante est inférieur à 10 et que la productivité d'une année à l'autre a été divisée par 2 ou plus.

Les raisons qui poussent à s'intéresser aux unités avec un facteur de productivité supérieur ou inférieur à 10 sont liées à la stratification et en particulier à la nouvelle classe de taille 9 introduite dans la révision du plan de sondage, voir section 5.2. Chacune des listes énoncées ci-dessus est divisée en deux listes, en fonction de la source du chiffre d'affaires TVA : chiffre d'affaires TVA original, réparti ou imputé, voir section 2.3 pour une définition. Ce sont ainsi au final quatre listes qui sont produites par la section METH et vérifiées par la section MON.

Pour limiter le nombre d'unités à faire contrôler par la section MON, la section METH vérifie si les unités figurant dans les listes ont le pouvoir d'influencer les estimations en cas de fort écart entre chiffre d'affaires TVA et réponse ± 0.10 . Pour cela, les outils utilisés sont les suivants :

- La proportion du chiffre d'affaires TVA de l'unité par rapport au total de la NOGA2, pour le chiffre d'affaires TVA le plus récent et celui de l'année précédente,
- l'ordre de grandeur du poids que l'unité recevrait en cas de tirage dans l'échantillon brut et de réponse à l'enquête WS 2021.

Le nombre d'entreprises sélectionnées est ainsi réduit par la section METH à quelques unités au plus en ce qui concerne les hausses de chiffres d'affaires. Concernant les baisses de chiffres d'affaires, les listes finales peuvent en revanche contenir quelques dizaines d'unités. La section METH soumet les listes à la section MON avec des commentaires individuels, pour chacune des unités soumises, détaillant en quoi le chiffre d'affaires TVA à vérifier pourrait poser problème s'il s'avérait incorrect. La section MON effectue alors des recherches, notamment à l'aide de données fiscales disponibles individuellement sur demande et indique les valeurs de chiffre d'affaires TVA à utiliser pour ces unités.

4 Hypothèses et calculs liés au plan de sondage

4.1 Variable proxy, variance et coefficient de variation

Le chiffre d'affaires TVA consolidé lors des opérations décrites en section 3 est une variable centrale pour les calculs de précision de l'enquête. Lorsqu'un plan de sondage est fixé, la variance

théorique de l'estimateur du total T_y d'une variable y repose sur l'entier du cadre de sondage. Dans le cas particulier d'un plan de sondage de Poisson, cette variance est donnée par :

$$\text{Var}(T_y) = \sum_{i \in U} \left(\frac{1}{\pi_i p_i} - 1 \right) y_i^2 \quad (1)$$

où U désigne le cadre de sondage, π_i est le taux de sondage de l'unité i , issu du plan de sondage, et p_i est la probabilité de réponse. L'estimateur T_y dont la variance est donnée par l'équation (1) est l'estimateur de Horvitz-Thompson

$$T_y = \sum_{i \in \text{échantillon net}} \frac{1}{\pi_i p_i} y_i. \quad (2)$$

Lors de l'élaboration du plan de sondage, aucune réponse y n'est disponible puisque l'enquête n'a pas commencé. Il est alors courant pour les calculs liés au plan de sondage de remplacer la réponse y par une variable "proxy", qui est disponible sur tout le cadre de sondage et qui est corrélée avec y . Avant la révision du plan de sondage de la WS 2021, la variable proxy utilisée était le nombre d'employés en équivalent plein temps, EPT_cadre. Utiliser EPT_cadre comme variable proxy revient à supposer que les réponses des unités à l'enquête sont proportionnelles au nombre d'employés. S'il est logique de penser que les unités avec plus d'employés répondent des valeurs (chiffre d'affaires t010, production brute t017, etc.) plus élevées, cette hypothèse de proportionnalité revient cependant à attribuer à chaque unité une productivité identique.

Les analyses ont montré que la corrélation entre le chiffre d'affaires TVA et plusieurs réponses (t010, t029) était supérieure à celle entre EPT_cadre et ces réponses (voir tables 10 et 11). Fort des contrôles effectués sur le chiffre d'affaires TVA, celui-ci apparaît dorénavant comme un meilleur choix. Ainsi, dans les calculs de variance de l'équation (1), la réponse y_i inconnue est remplacée par le chiffre d'affaires TVA consolidé de l'unité i .

L'équation (1) peut être adaptée dans le cas où l'estimation du total porte sur un domaine d'analyse et non pas sur toute la Suisse. Dans ce cas, la réponse des unités faisant partie du domaine est conservée telle quelle, alors que la réponse des unités qui ne font pas partie du domaine est mise à 0. Par ailleurs, il est possible de prendre en compte le fait que les estimations (une fois les réponses à l'enquête collectées) intégreront un calage des poids. En présence d'un calage, l'équation (1) continue de s'appliquer : une variance approchée (voir Tillé [2019], section 11.5) est estimée en remplaçant la réponse y_i par le résidu ε_i obtenu par une régression linéaire où les réponses y sont prédites à l'aide des variables de calage :

$$\widehat{\text{Var}}(T_y) = \sum_{i \in U} \left(\frac{1}{\pi_i p_i} - 1 \right) \varepsilon_i^2. \quad (3)$$

Lors de la pondération de la WS, un calage est effectué. Les variables de calage utilisées lors de la pondération sont déjà connues au moment d'effectuer le plan de sondage. Ces variables sont EPT_cadre et le chiffre d'affaires TVA. Le calage est effectué séparément au niveau de chaque NOGA2 et GR. Pour les calculs du plan de sondage, compte tenu du fait que le chiffre d'affaires TVA est justement la variable proxy utilisée, il n'est pas possible de refléter le bénéfice du futur calage des chiffres d'affaires TVA. En effet, cela mènerait à des résidus de régression ε_i nuls et une variance nulle dans l'équation (3) ce qui est trop optimiste puisqu'en pratique la réponse à l'enquête est différente du chiffre d'affaires TVA. Ainsi, pour le calcul des précisions réalisés dans ce rapport, les résidus ε_i sont obtenus en effectuant une régression linéaire de la réponse

dans le domaine considéré (c'est-à-dire le chiffre d'affaires TVA ou bien zéro, selon que l'unité est dans le domaine ou non) à l'aide des variables de calage sans le chiffre d'affaires TVA. Cette régression linéaire est effectuée en se basant sur toutes les unités du cadre de sondage.

Pour mesurer la précision de l'enquête, le coefficient de variation (CV) est souvent utilisé pour une statistique prenant des valeurs positives. Il est en effet plus facile à interpréter qu'une variance. Le CV, exprimé en pour cent, est donné par l'écart type de l'estimateur du total divisé par le total lui-même :

$$\widehat{CV}_y = \frac{\sqrt{\widehat{\text{Var}}(T_y)}}{\sum_{i \in U} y_i} \quad (4)$$

Le CV calculé lors d'un plan de sondage dépend des probabilités de tirage des unités π_i et des probabilités de réponse p_i . Une grandeur utile utilisée dans ce rapport (voir notamment la section 6.1) est le "CVmin". Celui-ci est égal au CV obtenu lorsque les taux de sondage π_i sont tous mis à 100%, c'est-à-dire que l'ensemble des quelque 160 000 unités du cadre de sondage sont interrogées, et en présence de non-réponse (voir section 4.2 pour le détail des hypothèses sur les taux de réponse). Le CVmin quantifie la précision maximale atteignable dans l'enquête en présence de non-réponse. Celui-ci est utilisé notamment pour l'identification des "Very Important Units" (VIU) ainsi que pour définir les objectifs de précision à atteindre, voir sections 5.1.3 et 6.1.2.

4.2 Hypothèses sur le taux de réponse des unités

Pour les calculs de précision de l'équation (3), des probabilités de réponse p_i sont attribuées aux unités. Ces probabilités de réponse pourraient éventuellement se baser sur un modèle de réponse construit à partir des enquêtes passées. Le modèle utilisé jusqu'ici (voir Assoulin [2013]) est simple et il a été décidé de le conserver pour cette révision. Les taux de réponse p_i utilisés dans l'équation (3) sont ainsi, respectivement, de 50%, 58%, 67%, 80%, 90% pour les unités en classe de taille 1, 2, 3, 4, 5. Ces chiffres se basent sur les taux de réponse des années passées et sont volontairement conservateurs dans le sens où les taux de réponse réels sont légèrement plus élevés en général.

5 Stratification et unités très importantes

Le plan de sondage de la WS est un plan de Poisson stratifié. Les strates sont définies comme étant le croisement entre NOGA2 et classe de taille modifiée `clt_x`, voir section 5.2. Dans un plan de Poisson, les unités sont sélectionnées ou non indépendamment les unes des autres. Établir un plan de sondage consiste à décider du taux de sondage dans chacune de ces strates. Cette procédure est décrite en section 6. La présente section détaille la méthode de choix d'unités particulières, les "Very Important Units" (VIU) puis conclut sur la construction des strates.

5.1 Very Important Units

5.1.1 Justifications et historique

La WS est une enquête sur les entreprises qui a la particularité d'avoir des réponses à l'enquête d'un ordre de grandeur potentiellement très différent d'une unité à l'autre. Le chiffre d'affaires selon l'enquête (`t010`) peut facilement être mille, voire un million de fois plus grand pour une unité de l'échantillon net par rapport à celui d'une autre. Or, les calculs de l'équation (3) montrent qu'un nombre limité d'unités peut potentiellement avoir un impact important sur les estimations.

Lors de la construction du plan de sondage, les calculs de précision se font à partir d'un certain nombre d'hypothèses, détaillées en section 4.2, sur les taux de réponse. Par exemple, les unités de 250 employés (variable `betot_cadre`) ou plus ont une probabilité de réponse fixée à 90% pour ces calculs. Cependant, il est important de s'interroger sur la validité de ces hypothèses pour les unités très influentes. En effet, ces hypothèses ont alors un fort impact sur les précisions estimées. Pour certaines unités, utiliser des probabilités de réponses de 1 peut améliorer significativement les précisions; mais une telle hypothèse se doit naturellement d'être réaliste.

Si certaines unités sont très influentes, le CVmin (voir section 4.1 pour une définition) est affecté négativement par la possible non-réponse de l'unité. Ceci pose la question du suivi approfondi des unités influentes, pour obtenir des réponses à l'enquête. Le terme "suivi approfondi" ne fait ici pas forcément référence à des relances plus insistantes pour obtenir des données, mais plutôt à l'utilisation par la section MON d'informations supplémentaires, telles que des données fiscales accessibles uniquement à la demande, qui peuvent servir de base à une imputation des principales réponses à l'enquête.

La notion de VIU, "Very Important Units", n'est pas nouvelle et était déjà utilisée dans la précédente révision. L'idée sous-jacente était alors principalement d'identifier des unités très influentes et de leur donner le statut de VIU. Les VIU étaient mises dans une strate exhaustive et tous les calculs de précision se faisaient en donnant une probabilité de réponse de 1 aux VIU. Cette hypothèse sur la probabilité de réponse, forcément optimiste, se justifiait par un suivi spécial des VIU pour limiter au maximum la non-réponse.

Une nouvelle procédure a été mise en place pour choisir les VIU avec notamment le souci d'effectuer des hypothèses plus réalistes sur les VIU. En effet, si l'on attribue une probabilité de réponse de 1 à des unités dans les calculs, il est important que ces unités répondent effectivement en pratique.

La nouvelle méthode de choix des VIU consiste en deux procédures. Une première procédure, déjà utilisée avec l'ancienne méthode, choisit des VIU "betot" qui sont des unités avec beaucoup d'employés. Cette procédure est détaillée en section 5.1.2. En plus des VIU betot, une nouvelle procédure d'identification d'unités influentes est mise en place. Celle-ci est basée sur les chiffres d'affaires TVA. L'influence d'une unité s'entend ici au sens de l'impact sur le CVmin de l'appartenance ou non de l'unité aux VIU. Les analyses sont réalisées au niveau de chaque NOGA2 et au niveau de chaque canton. La procédure est détaillée en section 5.1.3.

Pour traiter plus efficacement les problèmes de potentielle non-réponse des VIU, une interaction accrue est mise en place entre la section METH qui construit la liste de VIU et la section MON, qui apporte son expertise pour savoir si les hypothèses faites sur les VIU sont réalistes. Ces points sont discutés en section 5.1.4.

5.1.2 VIU betot

Des unités sont choisies comme VIU à l'aide d'un algorithme qui se base sur le nombre d'employés de l'unité (`betot_cadre`). Ces unités sont appelées "VIU betot" dans ce rapport. L'ancienne méthode de construction du plan de sondage utilisait déjà un algorithme presque identique, voir Assoulin [2014], section 2.4.

Dans le cadre de cette révision, seule une légère modification est apportée. Dans l'ancienne méthode, l'algorithme de sélection des VIU betot peut choisir des unités qui sont en classe de taille 5, avec plus de 250 employés, mais également dans les autres classes de taille. Il a été

décidé de continuer d'utiliser l'algorithme mais de ne retenir comme VIU que les unités en classe de taille 5. Les unités de moins de 250 employés qui sont choisies par l'algorithme ne sont pas immédiatement considérées comme des VIU mais plutôt comme des "candidats VIU", c'est-à-dire des unités qui peuvent potentiellement devenir des VIU si la section MON juge cela réaliste, voir section 5.1.4. L'idée de ne pas accepter automatiquement des unités en tant que VIU est liée aux hypothèses fortes qui sont faites sur les VIU, en particulier la probabilité de réponse de 1. En effet, si les unités de plus de 250 employés ont un taux de réponse très élevé, ce n'est pas forcément le cas pour les unités de moins de 250 employés.

L'algorithme de sélection des VIU betot parcourt les différentes NOGA2, avec un regroupement des NOGA 8 et 9. Il est décrit en détail dans Assoulin [2014], section 2.4. Cet algorithme mesure d'une certaine façon l'influence de chaque unité sur une moyenne et choisit l'unité comme VIU si celle-ci est jugée comme importante. Pour le choix des VIU betot, la moyenne considérée est celle de la variable `betot_cadre`. Dans l'ancienne méthodologie, un algorithme similaire de choix de "VIU umsatz" a été introduit au fil du temps, lorsque le chiffre d'affaires TVA est devenu disponible pour tout le cadre de sondage. Pour cette révision, il a été décidé de conserver l'algorithme de choix de VIU betot (avec, comme expliqué, une légère adaptation pour ne choisir automatiquement que des unités en classe de taille 5) mais pas celui de choix des VIU umsatz. Les VIU umsatz de l'ancienne méthode sont ainsi remplacés par des VIU "NOGA" et des VIU "cantons", tel que décrit dans la section suivante.

5.1.3 VIU NOGA et VIU cantons

La seconde étape de construction de la liste de VIU consiste à choisir des unités jugées influentes dans leur NOGA2 ou dans leur canton. L'idée retenue ici se base sur la notion d'intérêt/de gain lié au fait qu'une unité devienne une VIU. Comme expliqué précédemment, il a été décidé d'appliquer un choix automatique de VIU uniquement sur les unités de plus de 250 employés. Or, de telles unités sont déjà échantillonnées de manière exhaustive, avec un taux de sondage de 100%, ce qui pose la question du gain attendu si de telles unités deviennent des VIU.

Le fait de faire passer la probabilité de réponse à 1 pour une unité qui avait une probabilité de réponse plus faible a pour conséquence d'améliorer le CVmin de l'estimateur du total dans la NOGA2 de l'unité considérée. Ainsi, si un tel CVmin est pénalisé par des unités influentes, attribuer des probabilités de réponse de 1 aux unités influentes peut être avantageux (diminution de la taille d'échantillon à précision égale), à condition qu'il s'agisse d'une hypothèse réaliste.

Dans ce contexte, il est possible de calculer, pour chaque unité, le gain attendu, en termes de réduction de CVmin, lorsque la probabilité de réponse de l'unité passe à 1. Ce calcul est facilité dans la mesure où la contribution de l'unité au CVmin est nécessairement zéro si sa probabilité de réponse passe à 1, puisque le calcul des CVmin se fait en utilisant toujours un taux de sondage de 100%. En s'inspirant de l'équation (3), pour un plan de sondage exhaustif, la réduction de variance attendue si une unité i devient une VIU est donnée par :

$$\text{GainVariance}_i := \left(\frac{1}{p_i} - 1 \right) \varepsilon_i^2 \quad (5)$$

où les quantités p_i et ε_i ont été définies en section 4 et en gardant à l'esprit que le résidu ε_i dépend du domaine sur lequel une variance est estimée (ici : la NOGA à laquelle l'unité i appartient). Le gain de CVmin est obtenu en calculant simplement la racine carrée de GainVariance_i et en divisant par la somme des chiffres d'affaires TVA dans le même esprit que l'équation (4).

L'algorithme de sélection des VIU "NOGA" est appliqué séparément à chaque NOGA2 et fonctionne itérativement de la manière suivante :

1. Une régression linéaire est effectuée en se basant sur toutes les unités du cadre qui n'ont pas encore été choisies en tant que VIU. Le chiffre d'affaires TVA ⁶ est régressé à l'aide des variables de calage (autre que le chiffre d'affaires TVA lui-même, voir section 4.1).
2. À l'aide des résidus de la régression, la quantité GainVariance_i de l'équation (5) est calculée pour chaque unité de la NOGA2 considérée.
3. Les unités de la NOGA2 considérée qui sont éligibles à devenir des VIU (c'est-à-dire les unités en classe de taille 5) sont triées par ordre croissant en fonction de leur GainVariance_i .
4. Un CVmin est calculé pour la NOGA2. Si le CVmin est supérieur à un seuil fixé, alors l'unité (éligible à être une VIU) ayant le plus fort GainVariance_i est ajoutée comme VIU. Sinon, (si le CVmin est inférieur au seuil ou bien si il n'y a aucune unité restante en classe de taille 5 dans cette NOGA2) l'algorithme s'arrête et passe à la NOGA2 suivante.
5. Si après avoir ajouté une unité comme VIU le nouveau CVmin demeure supérieur au seuil fixé et si un nombre maximum de VIU pour la NOGA2 n'a pas encore été atteint, alors l'algorithme boucle sur l'étape 1 précédente pour choisir de nouvelles VIU dans la NOGA2. Sinon, l'algorithme passe à la NOGA2 suivante. Le nombre maximum de VIU par NOGA2 est fixé à 10.

L'algorithme ci-dessus contient un certain nombre de paramètres. Les seuils de CVmin à atteindre sont fixés par défaut à 2.5% lorsque la NOGA2 contient plus de 1000 unités. Le seuil est ramené à 5% si le nombre d'unités est compris entre 200 et 1000 et 7% s'il est inférieur à 200. Les seuils de CVmin à atteindre ont été choisis en relation avec les objectifs de précision (CV) au niveau NOGA2 présentés en section 6.1.2. Selon la NOGA, l'algorithme permet ou non d'atteindre les objectifs de CVmin. Les raisons possibles pour ne pas atteindre les objectifs sont :

- le nombre maximal de 10 VIU pour cette NOGA2 ne suffit pas à atteindre le seuil,
- la sélection de toutes les unités en classe de taille 5 de cette NOGA2 ne suffit pas à atteindre le seuil.

Les NOGA2 pour lesquelles l'objectif n'est pas atteint font l'objet d'une investigation. Un CVmin plus élevé que l'objectif est généralement dû à un nombre limité d'unités influentes (sur le CVmin) qui ne sont pas en classe de taille 5 et ne sont donc pas sélectionnables par l'algorithme. Ces unités influentes sont identifiées et peuvent parfois devenir des VIU, même si elles ne sont pas en classe de taille 5. Ce point est détaillé en section 5.1.4.

L'algorithme décrit ci-dessus s'applique pour choisir des VIU dites "VIU NOGA". Une fois cette procédure effectuée, un second jeu de VIU dites "VIU cantons" est choisi selon le même algorithme, en faisant cette fois-ci les calculs de variance et CV par canton et non par NOGA2. Les objectifs de CVmin à atteindre par canton sont de 5% pour les cantons qui contiennent plus de 1000 unités dans le cadre et 7% sinon. Le nombre maximum de VIU cantons pour un canton donné est lui aussi fixé à 10. Tout comme dans le cas des NOGA2, les cantons pour lesquels les objectifs ne sont pas atteints font l'objet d'un contrôle pour identifier d'éventuelles unités qui influent fortement sur le CVmin. Le fait de s'intéresser aux cantons pour choisir des VIU est nouveau. Ceci s'inscrit dans une volonté d'amélioration de la précision de l'enquête au niveau des grandes régions et cantons.

6. ou, plus précisément, le chiffre d'affaires TVA multiplié par la variable indicatrice d'appartenance de l'unité i considérée à la NOGA2 actuellement traitée,

5.1.4 Candidats VIU et application pratique de ces procédures

Comme mentionné en section 5.1.2, l'algorithme initial (c'est-à-dire utilisé avant la présente révision) de choix des VIU betot sélectionne des unités sans distinction de classe de taille. L'algorithme est toujours utilisé dans le cadre de cette révision avec une légère modification. Les unités en classe de taille 5 sont automatiquement acceptées comme VIU, alors que les unités en classe de taille inférieure ne le sont pas et sont intégrées à une liste de "candidats VIU".

Après avoir fait fonctionner l'algorithme de choix des VIU betot, un autre algorithme choisit des VIU NOGA et VIU cantons qui sont des unités en classe de taille 5. Celles-ci sont en principe acceptées automatiquement comme VIU. Comme évoqué en section 5.1.3, la section METH effectue des analyses sur les NOGA2 et cantons pour lesquels les objectifs de CVmin ne sont pas atteints. Ces analyses sont réalisées en affichant pour chacune de ces NOGA2 / chacun de ces cantons les unités (hors VIU) qui contribuent le plus au CVmin total de la NOGA2 / du canton, sur la base de l'équation (5). Il est en effet possible de calculer une contribution de chaque unité à la variance totale dans la NOGA2 / le canton. Les unités en classe de taille 5 ne figurent pas dans une telle liste d'unités puisqu'elles seraient à la place choisies comme VIU (la limite de 10 VIU par NOGA2 étant aussi suffisante en pratique). Si une unité de classe de taille inférieure à 5 a une contribution importante à la variance, alors il est en théorie intéressant – en termes de réduction du CVmin – que l'unité puisse tout de même devenir une VIU. La section METH ajoute alors une telle unité à la liste des "candidats VIU". Le fait de juger une contribution à la variance importante ou non dépend de la NOGA analysée. En pratique, les entreprises avec une contribution supérieures à 20% ont été ajoutées comme candidats VIU pour la majorité des NOGA. Pour certaines autres NOGA où l'objectif de CVmin n'était pas satisfait, ce seuil a été abaissé de sorte à pouvoir atteindre cet objectif. Finalement, la liste de candidats VIU construite est composée d'unités en classe de taille inférieure à 5 qui ne peuvent pas être automatiquement choisies comme des VIU mais qui permettraient des gains importants en termes de CVmin si elles pouvaient le devenir.

En pratique, au cours d'un premier tour de calcul, la section METH calcule une liste de VIU en classe de taille 5 et construit une première liste de candidats VIU. La liste de candidats VIU (uniquement) est soumise à la section MON qui effectue des contrôles pour savoir quelles unités dans cette liste peuvent raisonnablement devenir des VIU. Il est entendu que les VIU ont une probabilité de réponse de 1. Ainsi, la section MON se pose la question de savoir si des valeurs pour ces unités pourront être obtenues avec une très forte probabilité ou non. La section MON se base sur les réponses passées à l'enquête d'une part, et sur les informations disponibles d'autre part, en particulier la disponibilité ou l'absence de données fiscales qui serviraient de base à une imputation des réponses à l'enquête. La section MON identifie les candidats VIU qui peuvent effectivement devenir des VIU. Ceci mène à un second tour de calcul de VIU.

Lors du second tour de calcul, les algorithmes de sélection de VIU NOGA et VIU cantons sont appliqués à nouveau en prenant en compte, avant le début des calculs, la liste de VIU betot et la liste des candidats VIU acceptés par la section MON. Cette liste de candidats VIU acceptés peut mener à des choix de VIU légèrement différents de ceux du premier tour. Une fois ce calcul effectué, la section METH analyse la liste des VIU choisies en classe de taille 5. Selon le cas, certaines unités sont retirées si leur contribution à l'amélioration du CVmin est trop faible. Ceci peut se produire lorsque, pour une NOGA2 ou un canton, les objectifs de CVmin ne sont pas atteints, en raison par exemple d'un nombre limité d'unités influentes qui ne seraient pas en classe de taille 5, comme évoqué précédemment. Dans ce cas-là, les algorithmes décrits ci-dessus – programmés pour choisir des VIU jusqu'à atteindre un seuil de CVmin – ont tendance à choisir

des VIU jusqu'à atteindre le nombre maximal de VIU (fixé à 10) dans la NOGA2/le canton considéré, ou bien jusqu'à choisir la totalité des unités en classe de taille 5 de la NOGA2/du canton. À ce moment là, certaines VIU choisies offrent parfois une amélioration très marginale du CVmin. Partant du principe que VIU signifie "Very Important Units", il n'y pas lieu d'encombrer la liste des VIU d'unités qui offrent des gains faibles de CVmin. Cette procédure de suppression de VIU jugées superflues se fait sans utiliser formellement de seuil quant à l'amélioration du CVmin apportée par une unité devenant une VIU. La procédure concerne une dizaine de NOGA2/cantons et il est cependant possible de dire que, pour la WS 2021, toutes les VIU en classe de taille 5 retirées de la liste offraient des gains de CVmin inférieurs (et généralement très inférieurs) à 0.1%.

Une fois la liste de VIU en classe de taille 5 finalisée, la section METH soumet à la section MON les unités choisies. Même si les unités en classe de taille 5 répondent généralement très bien à l'enquête, la section MON effectue un dernier contrôle sur ces unités. Il s'agit de voir si certaines VIU ont déjà été tirées dans l'échantillon brut dans le passé et n'ont pas répondu à l'enquête. Il n'est en effet pas réaliste d'attribuer une probabilité de réponse de 1 à des unités non-répondantes dans le passé.

Après avoir reçu de la part de METH la liste des VIU en classe de taille 5, la section MON revient vers METH avec une liste d'unités qui ne peuvent finalement pas devenir des VIU. Un dernier tour de calcul a alors lieu pour savoir s'il est nécessaire ou non de proposer d'autres VIU pour remplacer ces unités. Au final, la procédure décrite ci-dessus choisit environ 250 unités comme VIU, ce qui est un nombre comparable avec celui obtenu avec l'ancienne méthode. Il y a environ 70 VIU en classe de taille inférieure à 5 : ce sont des candidats VIU qui ont été acceptés comme VIU par la section MON. Dans la classe de taille 5, il y a environ 150 VIU betot et 30 VIU NOGA ou VIU cantons. Sur ces 250 unités, l'objectif est d'atteindre un taux de réponse de 100%. Avant la révision, le nombre de VIU non-répondantes était généralement proche d'une ou de deux dizaines.

5.2 Stratification du plan de sondage

5.2.1 Ancienne méthode de stratification

Le plan de sondage de la WS est un plan de Poisson stratifié, avec des taux de sondage choisis dans chaque strate de sorte à satisfaire certains objectifs de précision. Jusqu'à la WS 2021, les strates étaient construites en utilisant un croisement entre les variables NOGA2 et classe de taille (c1t). Ce choix était justifié par de nombreux arguments que l'on expose ici de manière non exhaustive :

- Les objectifs de précision de l'enquête sont fixés par NOGA2. Il est ainsi utile que la stratification intègre cette variable.
- Les grandes unités ont tendance à répondre des valeurs monétaires plus élevées. Or, la précision de l'enquête dépend fortement de l'ordre de grandeur des réponses, voir par exemple l'équation (1). Il est ainsi nécessaire d'attribuer des taux de sondage élevés aux unités qui répondent des valeurs élevées, et donc d'intégrer la classe de taille de l'unité dans la stratification. Les moyens mis à disposition pour cette enquête permettent d'enquêter de manière exhaustive les unités en classe de taille 5 pour toutes les NOGA2 et les unités en classe de taille 4 pour toutes les NOGA2 sauf les NOGA 78, 85, 87.
- Historiquement, hormis les données sur le nombre d'employés de l'unité, il n'y avait pas de données disponibles au niveau de l'ensemble du cadre de sondage qui soient bien corrélées aux variables d'intérêt. En particulier, le chiffre d'affaires TVA n'était pas disponible sur l'ensemble du cadre.

Chaque strate, c'est-à-dire chaque croisement entre NOGA2 et classe de taille, reçoit un taux de sondage calculé de sorte à satisfaire des objectifs de précision. Avant de calculer un taux de sondage dans chacune des strates, une variable nommée "limit" est créée pour fixer un seuil d'exhaustivité. Le seuil est spécifique à chaque NOGA2. Dans une NOGA2 donnée, toutes les classes de taille supérieures ou égales à "limit" sont échantillonnées de manière exhaustive. Le mécanisme de calcul de la variable "limit" en fonction de la NOGA2 est décrit dans [Assoulin \[2013\]](#), section 3.1. Voici quelques points notables :

- Pour la majorité des grandes NOGA2, qui contiennent beaucoup d'unités, la variable "limit" est fixée à 4. Des exceptions s'appliquent : pour les NOGA 78, 85, 87, la variable est fixée à 5.
- Pour les NOGA2 qui contiennent peu d'unités (par exemple, les NOGA2 12, 19, 92), la variable "limit" peut prendre des valeurs inférieures à 4, selon un calcul détaillé dans [Assoulin \[2013\]](#), section 3.1.

5.2.2 Nouvelle méthode de stratification, classe de taille modifiée

Dans le cadre de la révision du plan, il apparaît qu'il peut exister dans le cadre de sondage des unités dans les classes de taille non-exhaustives avec des chiffres d'affaires TVA élevés. De telles unités peuvent générer des variances importantes dans les estimations si la strate dans laquelle elles figurent reçoit un taux de sondage π_i faible. Ceci résulte du poids de sondage élevé, $1/\pi_i$, présent dans l'équation (3). Une stratification basée uniquement sur le chiffre d'affaires TVA, croisé avec la NOGA2, serait envisageable mais aurait tout de même un certain nombre d'inconvénients. Par exemple, certaines grandes unités, sur la base de la variable `betot_cadre`, pourraient ne plus être échantillonnées de manière exhaustive.

La solution proposée ici est de conserver une stratification basée sur le croisement entre NOGA2 et classe de taille, ainsi que les règles concernant l'exhaustivité décrites en section 5.2.1. La modification la plus importante introduite dans cette révision porte ici sur la définition de la classe de taille, et permet de faire intervenir le chiffre d'affaires TVA. Ainsi, une classe de taille modifiée, `clt_x`, est introduite et est définie de la façon suivante :

- par défaut, `clt_x` est simplement égale à la classe de taille originale, `clt`,
- si une unité ne fait pas partie d'une strate déjà exhaustive (définie selon les règles édictées en section 5.2.1) et si cette unité a un facteur de productivité (voir section 2.3 pour une définition) supérieur à 10, alors une classe de taille modifiée, `clt_x = 9`, lui est attribuée.

La variable `clt_x` ainsi définie prend les valeurs 1, 2, 3, 4, 5 et 9. La classe de taille `clt_x = 9` est échantillonnée de manière exhaustive, à l'exception de la NOGA 46, voir détails ci-dessous. La nouvelle classe de taille est introduite pour cibler des unités très productives qui n'auraient pas forcément été tirées dans l'échantillon, parce qu'elles n'appartiennent normalement pas à une strate exhaustive. Par ce mécanisme, les petites unités très productives sont échantillonnées tout en continuant de sélectionner les grandes unités. Voici certains détails importants concernant la classe de taille 9 :

- Le calcul du facteur de productivité se fait sur la base du chiffre d'affaires TVA consolidé ainsi que sur la base du chiffre d'affaires TVA non-consolidé. Ainsi, si seulement l'un de ces deux facteurs de productivité dépasse 10 et que les autres conditions sont remplies, alors l'unité entre en classe de taille 9. Ce mécanisme permet de mettre en classe de taille 9 les unités avec un chiffre d'affaires TVA très élevé, mais qui a été fortement réduit lors des contrôles. Dans l'hypothèse où les contrôles auraient réduit à tort le chiffre d'affaires TVA de l'unité, l'impact sur la précision de l'enquête est moins problématique en raison

du poids modéré que l'unité recevra du fait de sa présence dans une strate exhaustive. L'inconvénient de cette démarche est de mettre dans une classe de taille exhaustive des unités qui répondent potentiellement des valeurs monétaires modestes, qui ne "nécessitent" pas forcément d'appliquer un taux de sondage de 100%.

- Le calcul de la classe de taille modifiée clt_x repose sur un seuil appliqué au facteur de productivité. Pour éviter de potentielles entrées/sorties d'une année à l'autre d'unités dont le facteur de productivité oscillerait autour de la valeur 10, une règle de persistance de la classe de taille 9 est mise en place. Si lors d'une année donnée, une unité fait partie de la classe de taille 9, alors l'année suivante l'unité restera dans cette classe de taille si son facteur de productivité dépasse 7 (au lieu de 10) et si les autres conditions sont réunies. Cette règle permet d'assurer une certaine stabilité dans les poids donnés aux unités d'une année à l'autre. L'inconvénient est ici encore une fois un nombre d'unités plus élevé dans l'échantillon du fait d'une classe de taille 9 qui pourrait avoir tendance à grandir d'année en année. Cet inconvénient est toutefois à tempérer dans la mesure où les seuils de facteurs de productivité pour appartenir à la classe de taille 9 sont fixés à des valeurs élevées, de sorte que in fine peu d'unités entrent dans cette classe de taille.
- Dans chaque NOGA2, la classe de taille 9 est échantillonnée de manière exhaustive, à l'exception de la NOGA 46. En effet, pour cette NOGA2, les chiffres d'affaires TVA non-consolidés affichent une corrélation plus faible avec les réponses à l'enquête et des plafonnements importants ont lieu, voir section 3.2. Dans la NOGA 46 uniquement, une classe de taille spéciale supplémentaire est créée – en plus de la classe de taille 9. Celle-ci est nommée "classe de taille 9.1" et contient des unités de la NOGA 46 qui entreraient normalement en classe de taille 9 et qui en plus ont un chiffre d'affaires TVA (non-consolidé) supérieur à 1 milliard de francs. Ainsi, dans la NOGA 46, la classe de taille 9 n'est pas exhaustive, mais la classe de taille 9.1 l'est.

Pour la WS 2021, le nombre d'unités qui reçoivent un taux de sondage de 100% du fait de leur appartenance à la classe de taille 9 (hors NOGA 46) ou 9.1 est d'environ 1300, voir tables 2 et 3.

6 Stratégie de calcul des taux de sondage

Une fois la stratification établie et la liste des VIU complétée, l'élaboration du plan consiste à calculer un taux de sondage dans chaque strate. Les taux de sondage dans les strates obtenues peuvent être vus comme un optimum sous contrainte, calculé à partir d'un programme d'optimisation convexe. La fonction optimisée (ici : minimisée) est la taille de l'échantillon brut, en nombre d'unités, laquelle est une fonction simple des taux de sondage dans les strates. Les contraintes portent sur les taux de sondage eux mêmes (taux de sondage minimal et/ou maximal) ainsi que sur les précisions à atteindre. Les estimateurs pour lesquels la précision est mesurée sont ceux du total par domaine. Ces objectifs de précision sont décrits en section 6.1. Les calculs relatifs à l'allocation sont présentés en section 6.2. Enfin, une fois le plan de sondage établi, un programme de tirage coordonné est appliqué pour sélectionner l'échantillon selon ce plan, voir section 6.3.

6.1 Objectifs de précision

6.1.1 Présentation des objectifs

Dans le cadre de la révision du plan de sondage, il a été décidé d'utiliser un nouvel algorithme d'allocation, c'est-à-dire un nouvel algorithme de calcul des taux de sondage dans chaque strate. L'algorithme utilisé est similaire à celui utilisé pour la révision du plan de sondage de l'enquête

sur la structure des salaires de l'OFS [Qualité & Potterat, 2020]. Auparavant, les objectifs de précision étaient fixés uniquement au niveau de chaque NOGA2. L'algorithme pour atteindre les objectifs était basé sur une allocation de Neyman, voir Assoulin [2013] section 3.2. Le nouvel algorithme d'allocation permet de fixer plusieurs objectifs de précision sur des domaines différents. Des objectifs sont ainsi fixés sur les domaines suivants :

- NOGA2,
- GR,
- canton,
- NOGA à 3 chiffres, nommée ici NOGA3, avec d'éventuels regroupements.

L'ensemble des objectifs fixés porte sur le coefficient de variation (CV), calculé dans le domaine considéré selon l'équation (4). La variance est calculée à l'aide des résidus de régression linéaire selon l'équation (3). La racine carrée de cette variance est divisée par le total du chiffre d'affaires TVA dans le domaine considéré. Les VIU ne contribuent pas à la variance et sont donc exclues de son calcul. Leur taux de sondage est en effet de 100% et leur probabilité de réponse est de 1. Les VIU sont en revanche bien prises en compte dans le calcul du total du chiffre d'affaires TVA qui permet de calculer les CV.

Les objectifs de CV prioritaires sont fixés au niveau de chaque NOGA2. Ici le mot "prioritaire" signifie que ce sont ces objectifs là – et non les objectifs au niveau des grandes régions, cantons, NOGA3 – qui conditionnent la taille de l'échantillon et les taux de sondage dans chaque strate. Plus précisément, pour le moment, les objectifs de précision supplémentaires fixés sur les autres domaines que la NOGA2 ne viennent pas augmenter la taille de l'échantillon par rapport à une situation où les objectifs seraient fixés uniquement au niveau de la NOGA2.

6.1.2 Objectifs de précision au niveau de la NOGA

La fixation d'objectifs de CV adaptés au niveau de la NOGA2 constitue une étape importante de la construction du plan de sondage décrite dans ce rapport. Ces objectifs ont en effet une influence importante sur la taille finale de l'échantillon brut. L'idée retenue ici dans le cadre de la révision du plan est double. Premièrement, il s'agit de reprendre la méthodologie développée dans Qualité & Potterat [2020]. Des objectifs relativement uniformes entre les NOGA2 sont fixés et sont basés sur le CVmin, c'est-à-dire le CV qu'il est possible d'atteindre avec une enquête exhaustive et en utilisant les hypothèses de la section 4.2 concernant les taux de réponse. Deuxièmement, la qualité de l'enquête doit être maintenue ou (à a coût égal) améliorée dans chaque NOGA2 ce qui implique que les objectifs de précision ne doivent pas être moins ambitieux qu'avec l'ancienne méthode.

Ce dernier point pose un problème de comparabilité car les objectifs de précision selon l'ancienne méthode étaient fixés en utilisant une autre variable "proxy" (voir, section 4.1 pour un rappel de définition) que le chiffre d'affaires TVA, à savoir la variable EPT_cadre. Pour calculer des précisions avec la nouvelle variable proxy et permettre une comparabilité entre l'ancienne et la nouvelle méthode la démarche suivante a été effectuée :

- Les taux de sondage de la dernière enquête (WS 2020) sont utilisés et des CV pour la WS 2020 sont calculés par NOGA2 en utilisant les taux de réponse de la section 4.2 et le chiffre d'affaires TVA comme variable proxy. Il s'agit ici de chiffres d'affaires TVA 2019 et non de réponses à la WS 2020. Les CV calculés correspondent ainsi à l'estimation du total des chiffres d'affaires TVA dans la WS 2020 avec les taux de sondage effectivement utilisés pour cette enquête.
- Les CV obtenus en 2020 sont comparés aux CVmin_20 de l'année 2020 pour chaque NOGA2. La différence (CV obtenu - CVmin_20) est calculée et est nommée "gap".

Le gap ainsi calculé dépend de la NOGA2 et donne un indice sur la précision de l'enquête, selon l'ancienne méthode, en fonction de la NOGA2. Lorsque le gap est faible, cela signifie que l'ancienne méthode obtient une précision très bonne dans le sens où celle-ci est proche de la précision optimale CVmin_20. À l'inverse, un gap élevé est signe d'une précision moins bonne pour la NOGA2 considérée. Dans la mesure où l'ancienne méthode permet d'atteindre une précision, dans chaque NOGA2, égale à (CVmin_20 + gap), il a été décidé, pour ne dégrader la qualité de l'enquête dans aucune NOGA2, de conserver la variable "gap" et de fixer pour la WS 2021 des objectifs de précision qui ne soient jamais moins ambitieux que (CVmin_21 + gap).

Pour des raisons de stabilité numérique, une tolérance de 0.1% est appliquée lorsque l'objectif de précision est fixé exactement à (CVmin_21 + gap). Ceci évite des situations où l'objectif n'est pas atteint en échantillonnant l'intégralité de la NOGA2 considérée et nécessite, pour améliorer la précision dans une NOGA2 donnée, d'augmenter la taille de l'échantillon dans d'autres NOGA2. Cette situation peut en effet se produire dans certaines NOGA2 qui étaient échantillonnées de manière exhaustive et qui ont un gap faible (mais pas nul). Les changements de structure dans ces NOGA2 dues aux variations naturelles du cadre de sondage peuvent mener, pour la WS 2021, à des CV légèrement supérieurs à (CVmin_21 + gap) même en cas d'échantillonnage exhaustif, ce qui justifie l'introduction de cette tolérance de 0.1%.

Le gap étant calculé, les objectifs de précision par NOGA2 sont les suivants :

- Si le CVmin_21 de la NOGA2 est inférieur à 2.5%, l'objectif de CV est de min(4% , CVmin_21+gap+0.1%), arrondi au 0.1% supérieur.
- Si le CVmin_21 de la NOGA2 est dans l'intervalle [2.5%, 4%), l'objectif de CV est de min(6% , CVmin_21+gap+0.1%), arrondi au 0.1% supérieur.
- Si le CVmin_21 de la NOGA2 est dans l'intervalle [4%, 7%), l'objectif de CV est de min(9% , CVmin_21+gap+0.1%), arrondi au 0.1% supérieur.
- Si le CVmin_21 de la NOGA2 est dans l'intervalle [7% et 15%), l'objectif de CV est de min(CVmin_21+3%, CVmin_21+gap+0.1%), arrondi au 0.1% supérieur.
- Si le CVmin_21 de la NOGA2 est supérieur à 15%, l'objectif de CV est de min(CVmin_21+5%, CVmin_21+gap+0.1%), arrondi au 0.1% supérieur.

Pour le plan de la WS 2021, parmi les 73 NOGA2, 50 objectifs de précision sont égaux à (CVmin_21 + gap + 0.1%), avec les éventuels arrondis décrits précédemment. Les 23 autres objectifs sont plus ambitieux que (CVmin_21 + gap + 0.1%) et sont donnés par les formules ci-dessus. Les objectifs de précision par NOGA2 sont donnés en table 4.

Il est important de noter que le gap calculé à partir de l'enquête WS 2020 ne sert pas uniquement de base aux objectifs de précision de la WS 2021. Ce même gap a été utilisé pour construire les plans de sondage de la WS 2022 et WS 2023. Les objectifs de précision par NOGA2 définis par les calculs ci-dessus peuvent être ponctuellement légèrement adaptés (à la hausse comme à la baisse) dans certaines NOGA2 de sorte à avoir un contrôle sur la taille de l'échantillon et en particulier sur l'intersection entre l'échantillon actuel et l'échantillon de l'année précédente. Ce point est traité en section 6.3.

6.1.3 Objectifs de précision sur d'autres domaines

Le nouvel algorithme d'allocation permet de fixer simultanément des objectifs sur des domaines différents. Dans le cadre de la révision, il a été décidé d'introduire des objectifs supplémentaires, pour des domaines autres que la NOGA2. Comme mentionné en section 6.1.1 les objectifs sur ces domaines supplémentaires sont pour le moment peu contraignants dans le sens où ceux-ci sont fixés de sorte à ne pas faire augmenter la taille finale de l'échantillon par rapport à une

situation où seuls des objectifs au niveau NOGA2 seraient fixés. Par ailleurs, il faut noter que pour des domaines tels que la grande région ou le canton il est difficile de fixer des objectifs de précision ambitieux en raison de la stratification actuellement utilisée, qui ne fait pas intervenir le canton ou la grande région. D'autres enquêtes telles que l'enquête sur la structure des salaires font intervenir la grande région dans la stratification [Qualité & Potterat, 2020] et ce point pourrait faire l'objet d'une future amélioration. Avec la stratification actuelle, augmenter la précision de l'enquête sur un canton donné, par exemple, nécessite de renforcer les taux de sondage dans toutes les strates qui contiennent des unités importantes faisant partie de ce canton. Le taux de sondage à l'intérieur d'une strate étant par définition homogène, ce dernier point est évidemment coûteux en termes de taille d'échantillon et mène à l'échantillonnage de beaucoup d'unités qui ne sont pas dans le canton concerné.

La mise en place d'objectifs supplémentaires, même peu contraignants, permet néanmoins d'observer les précisions obtenues sur des domaines autres que la NOGA2. Pour les grandes régions, les objectifs de précision sont fixés à 4%, à l'exception de la GR 7 (Tessin) qui a un CVmin_21 proche de cette valeur et pour laquelle un objectif de précision de 6% est fixé. Concernant les cantons, les objectifs de précision sont fixés par défaut à 5% si le CVmin_21 est inférieur à 3% et à (CVmin_21 + 5%) sinon. Des adaptations individuelles sont effectuées, notamment pour les cantons de petite taille en nombre d'unités, de sorte à ce qu'aucun de ces objectifs ne fasse augmenter la taille de l'échantillon obtenu sur la base des objectifs par NOGA2.

En plus des cantons et grandes régions, des objectifs de précision supplémentaires sont aussi fixés au niveau NOGA3. Comme certaines NOGA3 contiennent peu d'unités, un regroupement est effectué. Les objectifs sont fixés sur ces NOGA3 regroupées. L'algorithme effectue des regroupements de NOGA3 à l'intérieur de la même NOGA2 uniquement. Celui-ci fonctionne de la manière suivante : à l'intérieur de chaque NOGA2, les NOGA3 sont triées par ordre croissant de taille, en nombre d'unités. Des groupes de NOGA3 sont ensuite formés en agrégeant les plus petites NOGA3 jusqu'au premier dépassement de 500 unités. Une fois cette taille atteinte, un nouveau groupe de NOGA3 se forme avec les NOGA3 qui n'ont pas encore été regroupées. Si au terme de l'agrégation (une fois que toutes les NOGA3 de la NOGA2 considérée ont été prises en compte) le dernier groupe formé ne contient pas 500 unités, celui-ci est agrégé avec la plus petite NOGA3 regroupée créée précédemment.

6.2 Calculs, contraintes et algorithme d'allocation

Cette section décrit les outils de calcul des taux de sondage dans chaque strate une fois la stratification effectuée et les objectifs de précision définis. Les calculs reprennent largement, dans une version simplifiée, ceux effectués dans [Qualité & Potterat 2020], section 3.1.6.

Soit H le nombre de strates considérées et $\pi_1, \dots, \pi_H$ les taux de sondage dans ces strates. Le but est de déterminer les valeurs de $\pi_1, \dots, \pi_H$ de sorte à satisfaire les objectifs de précision fixés, tout en minimisant la taille visée de l'échantillon. Cette dernière se calcule facilement à partir des taux de sondage dans les strates, et vaut :

$$n = \sum_{h=1}^H \pi_h N_h \quad (6)$$

où $N_1, \dots, N_h$ est le nombre d'unités du cadre faisant partie de la strate $1, \dots, H$.

En reprenant l'équation (3), la variance de l'estimateur du total de la réponse y obtenu selon un plan de Poisson stratifié peut se décomposer en une somme selon chaque strate de la manière

suivante :

$$\widehat{\text{Var}}(T_y) = \sum_{h=1}^H \left(\sum_{i \in h} \left(\frac{1}{\pi_h p_i} - 1 \right) \varepsilon_i^2 \right). \quad (7)$$

Pour bien faire apparaître la dépendance de la variance aux taux de sondage $\pi_1, \dots, \pi_H$, l'équation (7) peut s'écrire :

$$\widehat{\text{Var}}(T_y) = \sum_{h=1}^H \left(\frac{1}{\pi_h} \sum_{i \in h} \frac{\varepsilon_i^2}{p_i} - B_h \right) \quad (8)$$

où $B_h := \sum_{i \in h} \varepsilon_i^2$ est une quantité dépendant uniquement de la strate h . Finalement, en notant également $D_h := \sum_{i \in h} \frac{\varepsilon_i^2}{p_i}$, l'équation obtenue est :

$$\widehat{\text{Var}}(T_y) = \sum_{h=1}^H \left(\frac{1}{\pi_h} D_h - B_h \right). \quad (9)$$

Comme les quantités D_h et B_h sont des constantes qui dépendent de la strate h considérée, il apparaît que la variance est une fonction simple des taux de sondage $\pi_1, \dots, \pi_H$. De plus, les objectifs de CV formulés en section 6.1, peuvent facilement être transformés en objectifs au niveau de la variance, c'est-à-dire en variance à ne pas dépasser, en multipliant l'objectif de CV par le total de la variable y (ici : le chiffre d'affaires TVA) sur le domaine considéré, et en élevant le résultat au carré. Cette variance est notée $\text{VAR}_{\text{objectif}}$ dans l'équation (11). Finalement, en notant $w_h := 1/\pi_h$ le poids de sondage dans la strate h , il apparaît que le problème posé est un problème d'optimisation (minimisation) de la taille de l'échantillon

$$n := \sum_{h=1}^H \frac{N_h}{w_h} \quad (10)$$

sous contraintes linéaires

$$\sum_{h=1}^H w_h D_h \leq \text{VAR}_{\text{objectif}} + \sum_{h=1}^H B_h. \quad (11)$$

La taille de l'échantillon à minimiser dans l'équation (10) est une fonction convexe en ses variables $w_1, \dots, w_H$. L'optimisation de fonctions multivariées convexes sous contraintes linéaires est effectuée ici à l'aide du package R "CVXR" [Fu et al., 2020].

Les développements ci-dessus s'appliquent à l'ensemble des objectifs de précision. Dans la WS, il y a 73 objectifs de précision pour les NOGA2, 7 pour les grandes régions, 26 pour les cantons, et 108 pour les NOGA3 regroupées, soit environ 200 objectifs au total. Pour chacun de ces objectifs, une régression linéaire doit être calculée pour obtenir des résidus ε_i et les vecteurs $(D_1, \dots, D_H)$, $(B_1, \dots, B_H)$ de taille H . Il y a ainsi une unique fonction convexe à minimiser, donnée par l'équation (10) et environ 200 contraintes linéaires qui s'écrivent à l'aide de l'équation (11). La fonction à minimiser dépend des poids de sondage $w_1, \dots, w_H$ dans un peu plus de $H = 400$ strates.

En plus des contraintes évoquées précédemment, des contraintes plus simples de taux de sondage minimum et maximum sont imposées. Celles-ci sont résumées ci-dessous :

- Le taux de sondage d'une strate donnée ne peut jamais être inférieur à un seuil fixé à 4%. Ce seuil est fixé à une valeur plus élevée que lors de la révision précédente. L'intérêt d'un seuil plus élevé sur le taux de sondage minimal est d'éviter d'obtenir des poids très élevés lors de la pondération. Un taux de sondage de 4% combiné avec une probabilité

de réponse faible et un calage peut déjà mener à des poids proches de 100 pour certaines unités. Ceci peut avoir un impact important sur les estimations, y compris de la variance, notamment dans le cas où les unités donnent une réponse y_i éloignée de ce qui est attendu compte tenu des variables de calage (résidus ε_i élevés en valeur absolue).

- Le taux de sondage des strates exhaustives est fixé à 100% en imposant à l'algorithme un minimum et un maximum égaux à 100%. Le mécanisme de construction des strates exhaustives est décrit en section 5.2.1. Pour chaque NOGA2, la classe de taille $clt_x = 9$ est exhaustive, sauf pour la NOGA 46 où le taux de sondage minimum est fixé à 60%. Dans cette NOGA, la classe de taille 9.1 (voir, section 5.2.2) est par contre bien exhaustive.
- Une règle de persistance du taux de sondage est mise en place pour éviter des variations à la baisse trop brutales des taux de sondage d'une année à l'autre, en particulier lors de l'entrée en vigueur de la nouvelle méthode pour le plan de sondage. Le taux de sondage d'une strate donnée ne peut pas prendre une valeur inférieure à 75% du taux de sondage appliqué l'année précédente dans la même strate. Cette règle vise à garantir une intersection suffisamment importante entre les échantillons de deux années consécutives, afin de pouvoir faire un certain nombre d'analyses sur cette intersection. Ce point est repris dans la section suivante sur les questions de rotation de l'échantillon.

La fonction `solve` du package R "CVXR" permet ainsi de calculer l'ensemble des taux de sondage $\pi_1, \dots, \pi_H$ dans les H strates. La taille de l'échantillon est donnée par l'équation (6), en gardant en tête qu'il s'agit en fait de l'espérance de la taille de l'échantillon. Celle-ci est en effet aléatoire puisque l'échantillon est tiré selon un plan de Poisson.

6.3 Algorithme de tirage coordonné

Pour tirer un échantillon selon un plan de Poisson, l'OFS dispose d'un système de coordination des échantillons. Cet algorithme est décrit dans [Qualité \[2009\]](#) et [Graf & Qualité \[2014\]](#). Celui-ci permet d'introduire des coordinations positives ou négatives entre l'échantillon à sélectionner et les échantillons des enquêtes WS des années précédentes, mais aussi avec d'autres enquêtes sur les entreprises, parmi lesquelles la statistique de l'emploi (STATEM) ou l'enquête Suisse sur la structure des salaires (ESS).

Pour permettre une rotation entre l'échantillon WS de l'année courante et les échantillons des années précédentes, l'échantillon est divisé en 5 sous-échantillons disjoints qui sont coordonnés positivement ou négativement avec les sous-échantillons des enquêtes précédentes de la WS. Introduire une coordination positive pour 4 des 5 sous-échantillons et négative pour 1 sous-échantillon permet en théorie un renouvellement de 20% de l'échantillon, ce qui assure à la fois une intersection suffisamment grande entre les échantillons de deux années consécutives et un renouvellement de l'échantillon qui est souhaitable du point de vue de la distribution de la charge d'enquête et du suivi de la structure des entreprises.

Le renouvellement de 20% de l'échantillon est cependant une valeur théorique qui s'applique dans des conditions où le cadre de sondage et les taux de sondage ne varient pas d'une année à l'autre et où les taux de sondage permettent effectivement d'appliquer ce renouvellement. Dans la pratique, le cadre de sondage évolue chaque année avec plusieurs (en proportion : quelques pour cent) entrées/sorties dans le cadre. De plus, dans les strates exhaustives (taux de sondage à 100%), la rotation est impossible puisque l'algorithme est contraint de tirer toutes les unités de la strate. La proportion d'unités qui sont à la fois tirées dans l'échantillon de l'année précédente et toujours présentes dans le nouvel échantillon a ainsi tendance à être plus basse que 80% dans les strates non-exhaustives – du fait des changements dans le cadre de sondage et/ou

dans les taux de sondage – et plus haute que 80% dans les strates exhaustives – du fait de l’absence de rotation mentionnée précédemment.

Pour la WS 2021, les taux de sondage obtenus uniquement par l’algorithme d’optimisation diffèrent parfois significativement de ceux obtenus les années précédentes avec l’ancienne méthode. Pour éviter une évolution trop rapide de l’échantillon et améliorer la comparabilité d’une année à l’autre, deux mesures ont été introduites :

- Une règle de persistance du taux de sondage est introduite, voir section 6.2 pour une description.
- Les 5 sous-échantillons tirés sont coordonnés de manière positive avec les échantillons des enquêtes passées. Le renouvellement “théorique” de 20% décrit précédemment est ainsi provisoirement abandonné pour compenser le renouvellement naturel qui provient des variations d’une année à l’autre des taux de sondage.

La dernière règle énoncée ci-dessus a vocation à être temporaire et il est souhaité de reprendre à l’avenir un mécanisme de rotation de l’échantillon.

7 Résultats de l’allocation

7.1 Cadre, nombre d’unités, stratification

L’algorithme de tirage décrit en section 6.3 calcule des taux de sondage dans chaque strate. La stratification (voir section 5.2) est obtenue en croisant la variable NOGA2 (avec un regroupement des NOGA 6, 7, 9) avec la classe de taille modifiée `clt_x`. Les tables 2 et 3 donnent le nombre d’unités de chaque strate, au niveau du cadre de sondage.

7.2 Objectifs de précision par NOGA2

Les objectifs de précision par NOGA2 sont décrits dans la table 4. Les définitions du CVmin et du gap sont données en section 6.1.2. Lorsque l’objectif de précision (colonne “CVobj”) est inférieur à la somme (CVmin₂₁ + gap), alors l’objectif de précision est plus ambitieux que celui qui serait fixé avec l’ancienne méthodologie. Comme expliqué en section 6.1.2, l’objectif de précision est fixé à minima à (CVmin₂₁ + gap) avec une marge de 0.1% et arrondi au 0.1% supérieur.

7.3 Résultats de l’allocation

Les résultats de l’allocation par NOGA2, grande région et canton sont décrits dans les tables 5 et 6. Les tables indiquent les CV obtenus lors de l’allocation (qui sont parfois plus bas que l’objectif de CV, voir les explications plus loin dans cette section), le nombre d’unités N dans le cadre et le nombre d’unités n tirées dans l’échantillon. Pour plus de la moitié des NOGA2 (41 sur 73 exactement), le CV obtenu lors de l’allocation est inférieur à la quantité (CVmin₂₁ + gap) de la table 4, c’est-à-dire au CV qui serait obtenu avec l’ancienne méthode d’allocation. De plus, pour une grande partie des 32 NOGA où le CV obtenu avec l’allocation n’est pas meilleur que (CVmin₂₁ + gap), il faut remarquer que le gap est particulièrement faible et même souvent proche de zéro. Cela signifie que ces NOGA2 étaient déjà échantillonnées de manière exhaustive ou quasi-exhaustive ce qui explique qu’il ne soit pas possible d’améliorer davantage la précision dans ces NOGA2. Il est rappelé que le gap est ici calculé sur la base du plan de sondage de la WS 2020, en calculant le CV obtenu avec ce plan et en soustrayant le CVmin de 2020, voir section 6.1.2.

TABLE 2 Nombre d'entreprises éligibles au cadre de sondage WS par NOGA2 et classe de taille modifiée clt_x, pour les NOGA 06 à 50.

NOGA2	unités	clt_x 1	clt_x 2	clt_x 3	clt_x 4	clt_x 5	clt_x 9	clt_x 9.1	VIU
0809	201	76	45	59	16		2		3
10	2 596	1 163	678	460	221	47	22		5
11	288	179	44	35	21	1	3		5
12	10	1	1	3	2	1			2
13	270	147	50	42	29		1		1
14	236	186	28	11	9				2
15	86	52	17	10	2		1		4
16	2 873	1 870	619	306	73	3	2		
17	125	37	22	29	31	2	1		3
18	892	536	174	127	48	3	4		
19	6	1	1	2					2
20	441	175	72	70	85	21	12		6
21	211	67	26	26	54	21	10		7
22	532	184	115	118	97	10	5		3
23	561	271	107	99	58	6	17		3
24	167	52	29	33	36	6	1		10
25	3 947	2 107	889	623	301	23	4		
26	1 320	461	255	246	262	71	17		8
27	546	234	103	105	76	20	2		6
28	1 383	491	245	292	290	53	10		2
29	131	65	27	20	13	1	1		4
30	139	85	16	15	13	4	1		5
31	394	201	70	75	45	2	1		
32	1 160	797	167	104	60	1	18		13
33	1 197	894	196	75	24	2	2		4
35	440	173	62	85	69	30	14		7
36	111	86	13	5	4		1		2
37	181	81	48	36	12		2		2
38	479	210	105	97	53	2	11		1
39	19	10	6	1					2
41	3 690	2 125	678	507	280	33	66		1
42	535	229	98	113	78	14	1		2
43	18 030	12 124	3 595	1 797	462	32	20		
45	6 812	5 207	968	426	113	17	79		2
46	9 949	5 890	1 726	1 023	621	105	515	65	4
47	14 508	10 536	2 575	866	357	85	87		2
49	3 671	2 121	713	460	299	55	22		1
50	108	37	20	31	11		6		3

En observant les CV obtenus dans la table 5, il est possible de remarquer que le CV obtenu (CV alloc) est parfois inférieur à l'objectif de CV (CV obj). Il est légitime de se demander comment ceci est possible avec un l'algorithme d'allocation qui cherche normalement à minimiser une taille d'échantillon. La raison est liée aux contraintes qui sont appliquées sur les taux de sondage des différentes strates. Il peut arriver, par exemple, que toutes les classes de taille d'une NOGA2 donnée soient contraintes à être exhaustives (voir, section 5.2.1 et les explications sur la variable

TABLE 3 Nombre d'entreprises éligibles au cadre de sondage WS par NOGA2 et classe de taille modifiée clt_x, pour les NOGA 51 à 96.

NOGA2	unités	clt_x 1	clt_x 2	clt_x 3	clt_x 4	clt_x 5	clt_x 9	clt_x 9.1	VIU
51	102	43	24	18	9	3	3		2
52	886	453	149	137	82	27	32		6
53	233	125	51	25	18	9	1		4
55	3 262	1 513	860	616	250	19	4		
56	14 522	10 691	2 678	879	231	34	9		
58	563	325	102	76	41	5	8		6
59	627	439	101	55	18		13		1
60	78	18	17	25	15	1	1		1
61	210	103	31	29	25	4	14		4
62	5 262	3 367	954	577	268	40	56		
63	432	259	85	47	25	4	9		3
68	5 445	4 240	622	334	109	15	123		2
69	6 810	5 454	935	311	61	11	35		3
70	4 275	3 012	556	323	209	21	142		12
71	9 207	6 198	1 694	923	291	19	82		
72	804	463	144	82	69	10	32		4
73	1 261	809	247	117	60	1	22		5
74	2 001	1 608	240	90	33		27		3
75	672	527	112	28	3	1			1
77	631	424	90	46	29		34		8
78	1 186	317	163	248	336	97	21		4
79	852	549	145	84	46	2	21		5
80	440	217	95	74	38	10	2		4
81	5 652	3 953	941	507	206	35	7		3
82	811	476	157	94	55		20		9
85	3 873	2 158	738	538	374	51	14		
87	1 076	127	82	220	588	56	2		1
88	2 392	1 279	594	356	150	10	3		
90	1 024	744	121	92	50		8		9
91	65	35	12	8	8				2
92	32	6	1	3	12				10
93	3 258	2 166	619	322	101	11	34		5
94	1 081	600	221	155	79	2	18		6
95	308	266	26	7	7				2
96	4 226	3 765	303	105	46	6	1		
Total	161 804	105 890	28 513	15 983	8 167	1 175	1 759	65	252

“limit”), ce qui peut mener à des CV meilleurs que les objectifs. Il est aussi possible que les taux de sondage minimaux imposés aux strates non-exhaustives soient déjà suffisants pour dépasser les objectifs de précision. Le taux de sondage minimum d’une strate non-exhaustive donnée est de 4% et ce seuil peut éventuellement être relevé au delà de 4% en fonction du taux de sondage appliqué l’année précédente, voir section 6.2.

Les résultats de l’allocation par grande région et canton font apparaître des objectifs pour le moment moins ambitieux et des précisions assez larges dans certains cantons. Les objectifs

TABLE 4 Objectifs de précision par NOGA2. Le CVmin calculé ici est celui de la WS 2021 : CVmin_21.

NOGA2	CVmin	gap	CVmin + gap	CVobj	NOGA2	CVmin	gap	CVmin + gap	CVobj
0809	3.7%	5.0%	8.8%	6.8%	50	1.2%	0.0%	1.2%	1.3%
10	1.5%	1.2%	2.6%	2.8%	51	4.1%	0.0%	4.1%	4.2%
11	2.8%	0.4%	3.3%	3.4%	52	4.0%	8.0%	12.0%	10.1%
12	1.1%	0.0%	1.1%	1.2%	53	3.8%	0.2%	4.0%	4.2%
13	2.9%	0.7%	3.5%	3.7%	55	1.1%	2.2%	3.3%	3.4%
14	3.0%	0.8%	3.8%	3.9%	56	0.7%	2.9%	3.7%	3.8%
15	2.3%	0.0%	2.3%	2.5%	58	2.3%	1.9%	4.2%	4.0%
16	1.3%	4.3%	5.6%	4.0%	59	5.7%	0.6%	6.3%	6.4%
17	2.8%	0.0%	2.8%	2.9%	60	4.5%	0.0%	4.5%	4.7%
18	2.4%	1.3%	3.7%	3.9%	61	3.1%	0.6%	3.7%	3.9%
19	0.3%	0.0%	0.3%	0.5%	62	2.4%	5.1%	7.5%	5.6%
20	4.1%	0.2%	4.2%	4.4%	63	4.7%	0.3%	5.0%	5.2%
21	2.4%	0.1%	2.5%	2.7%	68	2.4%	3.9%	6.3%	4.3%
22	3.6%	0.6%	4.2%	4.4%	69	1.1%	3.6%	4.7%	4.0%
23	2.8%	3.2%	6.0%	6.0%	70	4.0%	6.1%	10.1%	9.0%
24	2.9%	0.0%	2.9%	3.0%	71	1.4%	6.1%	7.5%	5.5%
25	1.0%	2.2%	3.1%	3.3%	72	4.5%	0.4%	4.9%	5.0%
26	1.9%	0.5%	2.4%	2.6%	73	2.8%	4.4%	7.2%	6.0%
27	2.2%	0.6%	2.7%	2.9%	74	2.6%	4.7%	7.3%	6.0%
28	1.4%	0.7%	2.1%	2.2%	75	1.5%	1.5%	2.9%	3.1%
29	4.9%	0.0%	4.9%	5.1%	77	8.1%	4.7%	12.8%	12.0%
30	2.0%	0.0%	2.0%	2.2%	78	1.9%	2.3%	4.2%	4.0%
31	2.2%	0.8%	3.0%	3.2%	79	3.7%	8.0%	11.7%	9.8%
32	2.5%	2.6%	5.1%	4.0%	80	1.7%	0.7%	2.4%	2.6%
33	1.0%	1.9%	2.9%	3.1%	81	1.1%	2.6%	3.6%	3.8%
35	3.6%	0.0%	3.6%	3.7%	82	3.5%	3.6%	7.2%	6.0%
36	3.4%	0.0%	3.4%	3.6%	85	1.7%	2.4%	4.2%	4.0%
37	3.2%	0.5%	3.7%	3.8%	87	1.1%	2.0%	3.1%	3.3%
38	4.9%	4.6%	9.5%	9.0%	88	1.1%	2.3%	3.5%	3.6%
39	10.6%	0.0%	10.6%	10.8%	90	2.6%	0.7%	3.3%	3.4%
41	2.1%	3.8%	5.9%	4.0%	91	10.3%	0.0%	10.3%	10.5%
42	1.5%	0.9%	2.4%	2.6%	92	1.0%	0.0%	1.0%	1.1%
43	0.4%	2.8%	3.3%	3.4%	93	1.4%	3.4%	4.8%	4.0%
45	2.2%	6.1%	8.2%	6.3%	94	2.6%	3.3%	5.9%	6.0%
46	1.9%	1.0%	3.0%	3.1%	95	2.4%	2.0%	4.5%	4.0%
47	0.8%	1.6%	2.4%	2.5%	96	1.4%	2.7%	4.2%	4.0%
49	1.2%	1.9%	3.1%	3.2%					

fixés au niveau des grandes régions et des cantons sont discutés en section 6.1.3. Comme expliqué précédemment, la stratification actuelle qui ne fait pas intervenir ces domaines ne permet pas encore de fixer des objectifs ambitieux. Cependant, l'introduction de VIU cantons (voir section 5.1.3) est nouvelle et constitue une première avancée vers des objectifs plus ambitieux.

TABLE 5 Coefficients de variation (objectif de précision et résultat de l'allocation), nombre d'unités N dans le cadre WS et nombre d'unités n obtenues dans l'échantillon brut par NOGA2.

NOGA2	CV obj	CV alloc	N	n	NOGA2	CV obj	CV alloc	N	n
0809	6.8%	5.0%	201	115	50	1.3%	1.2%	108	100
10	2.8%	2.2%	2 596	600	51	4.2%	4.2%	102	82
11	3.4%	3.4%	288	147	52	10.1%	4.8%	886	232
12	1.2%	1.1%	10	10	53	4.2%	4.1%	233	147
13	3.7%	3.7%	270	129	55	3.4%	3.4%	3 262	451
14	3.9%	3.9%	236	92	56	3.8%	2.5%	14 522	861
15	2.5%	2.5%	86	73	58	4.0%	3.8%	563	143
16	4.0%	4.0%	2 873	356	59	6.4%	6.4%	627	311
17	2.9%	2.8%	125	120	60	4.7%	4.7%	78	74
18	3.9%	3.9%	892	188	61	3.9%	3.2%	210	123
19	0.5%	0.3%	6	6	62	5.6%	5.6%	5 262	624
20	4.4%	4.1%	441	220	63	5.2%	5.2%	432	170
21	2.7%	2.5%	211	164	68	4.3%	4.3%	5 445	739
22	4.4%	4.3%	532	223	69	4.0%	4.0%	6 810	394
23	6.0%	4.6%	561	164	70	9.0%	4.9%	4 275	560
24	3.0%	2.9%	167	129	71	5.5%	5.4%	9 207	749
25	3.3%	3.3%	3 947	464	72	5.0%	4.8%	804	248
26	2.6%	2.2%	1 320	498	73	6.0%	5.5%	1 261	223
27	2.9%	2.3%	546	195	74	6.0%	6.0%	2 001	272
28	2.2%	2.1%	1 383	460	75	3.1%	3.1%	672	172
29	5.1%	5.0%	131	112	77	12.0%	8.9%	631	185
30	2.2%	2.1%	139	119	78	4.0%	3.3%	1 186	294
31	3.2%	3.2%	394	147	79	9.8%	5.1%	852	153
32	4.0%	3.9%	1 160	181	80	2.6%	2.6%	440	177
33	3.1%	3.1%	1 197	169	81	3.8%	3.6%	5 652	472
35	3.7%	3.6%	440	398	82	6.0%	4.7%	811	227
36	3.6%	3.6%	111	93	85	4.0%	4.0%	3 873	397
37	3.8%	3.8%	181	116	87	3.3%	3.3%	1 076	337
38	9.0%	6.6%	479	171	88	3.6%	3.6%	2 392	320
39	10.8%	10.6%	19	19	90	3.4%	3.4%	1 024	483
41	4.0%	4.0%	3 690	543	91	10.5%	10.5%	65	50
42	2.6%	2.5%	535	187	92	1.1%	1.0%	32	32
43	3.4%	2.3%	18 030	1 215	93	4.0%	3.4%	3 258	348
45	6.3%	5.6%	6 812	491	94	6.0%	5.0%	1 081	238
46	3.1%	2.8%	9 949	2 066	95	4.0%	4.0%	308	135
47	2.5%	2.0%	14 508	1 142	96	4.0%	4.0%	4 226	332
49	3.2%	2.9%	3 671	535	Total			161 804	22 912

TABLE 6 Coefficients de variation (objectif de précision et résultat de l'allocation), nombre d'unités N dans le cadre WS et nombre d'unités n obtenues dans l'échantillon brut, par grande région et canton.

GR	CV obj	CV alloc	N	n	Canton	CV obj	CV alloc	N	n
1	4.0%	3.1%	30 503	4 024	AG	5.0%	2.0%	10 760	1 578
2	4.0%	1.8%	32 076	4 474	AI	35.0%	32.1%	417	53
3	4.0%	1.4%	19 146	2 916	AR	25.0%	12.2%	1 030	129
4	4.0%	2.5%	28 301	4 257	BE	5.0%	2.8%	18 061	2 583
5	4.0%	2.1%	23 232	3 252	BL	10.0%	5.1%	4 541	696
6	4.0%	3.3%	19 044	2 869	BS	5.0%	2.2%	3 845	642
7	6.0%	5.2%	9 502	1 120	FR	9.0%	5.0%	5 039	661
					GE	9.0%	5.2%	9 894	1 471
					GL	15.0%	11.3%	835	126
					GR	8.0%	5.5%	4 868	543
					JU	10.0%	8.5%	1 473	198
					LU	9.0%	4.3%	8 099	1 157
					NE	10.0%	4.5%	3 157	433
					NW	15.0%	11.3%	962	136
					OW	20.0%	16.2%	844	116
					SG	5.0%	4.3%	9 665	1 458
					SH	10.0%	5.7%	1 501	265
					SO	5.0%	3.1%	4 346	599
					SZ	10.0%	7.7%	3 829	504
					TG	9.0%	5.4%	4 916	678
					TI	9.0%	5.2%	9 502	1 120
					UR	18.0%	15.7%	607	74
					VD	5.0%	3.3%	13 377	1 812
					VS	5.0%	4.6%	7 232	741
					ZG	10.0%	5.9%	4 703	882
					ZH	5.0%	2.5%	28 301	4 257

8 Conclusions et perspectives

La présente révision du plan de sondage de la statistique de la valeur ajoutée (WS) met le chiffre d'affaires TVA au centre de la nouvelle méthodologie et vise à échantillonner en priorité les unités avec un chiffre d'affaires TVA élevé. Cette stratégie permet des gains sur les coefficients de variation, notamment les coefficients de variation par NOGA2. Ces gains sont évalués ici à l'aide d'une variable proxy (chiffre d'affaires TVA) et seront estimés une fois les réponses à l'enquête obtenues, lors de la pondération. Un rapport de méthodes sur la pondération de la WS est en cours de préparation. La nouvelle méthodologie introduit la possibilité de fixer des objectifs de précision supplémentaires sur des domaines autres que la NOGA2, tels que la grande région, le canton, ou encore la NOGA3. Des objectifs de précision plus ambitieux sur ces domaines sont une perspective possible d'amélioration future, par exemple dans le cadre de la production de nouvelles statistiques. Cela nécessiterait possiblement de modifier la stratification du plan de sondage pour mieux faire intervenir les grandes régions ou cantons. Le processus de contrôle du chiffre d'affaires TVA pour assurer une bonne corrélation avec la réponse à l'enquête ± 010 (chiffre d'affaires) demeure complexe et peut faire l'objet d'améliorations et/ou d'automatisations supplémentaires. Ceci nécessite notamment une bonne compréhension des mécanismes qui font que le chiffre d'affaires TVA peut s'écarter du chiffre d'affaires relevé dans l'enquête ± 010 .

Bibliographie

- ASSOULIN, D. (2013). Methodenbericht : Wertschöpfungsstatistik, Revision 2009 : Stichprobenrahmen und Stichprobenplan. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- ASSOULIN, D. (2014). Methodenbericht : Wertschöpfungsstatistik, Revision 2009 : Statistische Datenaufbereitung und Hochrechnung. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- FAVRE-MARTINOZ, C. & RUIZ-GAZEN, A. & BEAUMONT, J. F. & HAZIZA, D. (2018). Robustness in Survey Sampling Using the Conditional Bias Approach with R Implementation. In : Perna, C., Pratesi, M., Ruiz-Gazen, A. (eds) *Studies in Theoretical and Applied Statistics. SIS 2016. Springer Proceedings in Mathematics & Statistics* **227**. Springer, Cham.
- FU, A. & NARASIMHAN, B. & BOYD, S. (2020). CVXR : An R Package for Disciplined Convex Optimization. *Journal of Statistical Software* **94**, 1–34.
- GRAF, E. & QUALITÉ, L. (2014). Sondage dans des registres de population et de ménages en Suisse : coordination d'échantillons, pondération et imputation. *Journal de la Société Française de Statistique* **155(4)**, 95–133.
- MORENO-REBOLLO, J. L. & MUÑOZ-REYES, A. & MUÑOZ-PICHARDO, J. (1999). Influence diagnostics in survey sampling : Conditional bias. *Biometrika* **86**, 923–928.
- NEDYALKOVA, D. & ASSOULIN, D. (2024). Swiss structural business statistics (STATENT) 2015-2021. Estimation and precision of the full-time equivalent. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- NEYMAN, J. (1934). On the Two Different Aspects of the Representative Method : The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society* **97(4)**, 558–625.
- 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>
- QUALITÉ, L. (2009). Unequal probability sampling and repeated surveys. *University of Neuchâtel*, Ph.D. Thesis.
- QUALITÉ, L. & POTTERAT, J. (2020). Rapport de méthodes : Révision du plan de sondage pour l'enquête Suisse sur la structure des salaires 2018. Rapport de méthodes, Office Fédéral de la Statistique OFS.
- 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
- STEKHOFEN, D.J. & BÜHLMANN, P. (2012). MissForest - non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28(1)**, 112–118.
- TILLÉ, Y. (2019). Théorie des sondages-2e éd. : Échantillonnage et estimation en populations finies. Cours et exercices corrigés. *Dunod*.

Annexes

A Liste des NOGA2

TABLE 7 Liste des sections NOGA et NOGA2 : NOGA 1 à 39

sect. NOGA	NOGA	titre
A	01	Culture et production animale, chasse et services annexes
A	02	Sylviculture et exploitation forestière
A	03	Pêche et aquaculture
B	05	Extraction de houille et de lignite
B	06	Extraction d'hydrocarbures
B	07	Extraction de minerais métalliques
B	08	Autres industries extractives
B	09	Services de soutien aux industries extractives
C	10	Industries alimentaires
C	11	Fabrication de boissons
C	12	Fabrication de produits à base de tabac
C	13	Fabrication de textiles
C	14	Industrie de l'habillement
C	15	Industrie du cuir et de la chaussure
C	16	Travail du bois et fabrication d'articles en bois et en liège, à l'exception des meubles; fabrication d'articles en vannerie et sparterie
C	17	Industrie du papier et du carton
C	18	Imprimerie et reproduction d'enregistrements
C	19	Cokéfaction et raffinage
C	20	Industrie chimique
C	21	Industrie pharmaceutique
C	22	Fabrication de produits en caoutchouc et en plastique
C	23	Fabrication d'autres produits minéraux non métalliques
C	24	Métallurgie
C	25	Fabrication de produits métalliques, à l'exception des machines et des équipements
C	26	Fabrication de produits informatiques, électroniques et optiques
C	27	Fabrication d'équipements électriques
C	28	Fabrication de machines et équipements n.c.a.
C	29	Industrie automobile
C	30	Fabrication d'autres matériels de transport
C	31	Fabrication de meubles
C	32	Autres industries manufacturières
C	33	Réparation et installation de machines et d'équipements
D	35	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné
E	36	Captage, traitement et distribution d'eau
E	37	Collecte et traitement des eaux usées
E	38	Collecte, traitement et élimination des déchets; récupération
E	39	Dépollution et autres services de gestion des déchets

TABLE 8 Liste des sections NOGA et NOGA2 : NOGA 41 à 88

sect. NOGA	NOGA	titre
F	41	Construction de bâtiments
F	42	Génie civil
F	43	Travaux de construction spécialisés
G	45	Commerce et réparation d'automobiles et de motocycles
G	46	Commerce de gros, à l'exception des automobiles et des motocycles
G	47	Commerce de détail, à l'exception des automobiles et des motocycles
H	49	Transports terrestres et transport par conduites
H	50	Transports par eau
H	51	Transports aériens
H	52	Entreposage et services auxiliaires des transports
H	53	Activités de poste et de courrier
I	55	Hébergement
I	56	Restauration
J	58	Édition
J	59	Production de films cinématographiques, de vidéo et de programmes de télévision; enregistrement sonore et édition musicale
J	60	Programmation et diffusion
J	61	Télécommunications
J	62	Programmation, conseil et autres activités informatiques
J	63	Services d'information
K	64	Activités des services financiers, hors assurance et caisses de retraite
K	65	Assurance
K	66	Activités auxiliaires de services financiers et d'assurance
L	68	Activités immobilières
M	69	Activités juridiques et comptables
M	70	Activités des sièges sociaux; conseil de gestion
M	71	Activités d'architecture et d'ingénierie; activités de contrôle et analyses techniques
M	72	Recherche-développement scientifique
M	73	Publicité et études de marché
M	74	Autres activités spécialisées, scientifiques et techniques
M	75	Activités vétérinaires
N	77	Activités de location et location-bail
N	78	Activités liées à l'emploi
N	79	Activités des agences de voyage, voyagistes, services de réservation et activités connexes
N	80	Enquêtes et sécurité
N	81	Services relatifs aux bâtiments et aménagement paysager
N	82	Activités administratives et autres activités de soutien aux entreprises
O	84	Administration publique et défense; sécurité sociale obligatoire
P	85	Enseignement
Q	86	Activités pour la santé humaine
Q	87	Hébergement médico-social et social
Q	88	Action sociale sans hébergement

TABLE 9 Liste des sections NOGA et NOGA2 : NOGA 90 à 99

sect. NOGA	NOGA	titre
R	90	Activités créatives, artistiques et de spectacle
R	91	Bibliothèques, archives, musées et autres activités culturelles
R	92	Organisation de jeux de hasard et d'argent
R	93	Activités sportives, récréatives et de loisirs
S	94	Activités des organisations associatives
S	95	Réparation d'ordinateurs et de biens personnels et domestiques
S	96	Autres services personnels
T	97	Activités des ménages en tant qu'employeurs de personnel domestique
T	98	Activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre
U	99	Activités des organisations et organismes extraterritoriaux

B Corrélations entre variables auxiliaires et réponse à l'enquête t010

TABLE 10 Corrélations de Pearson (multipliées par 100), mesurées dans chaque NOGA2, entre la réponse à l'enquête t010 de la WS 2021 et les variables auxiliaires EPT_cadre, chiffre d'affaires TVA de l'année 2020 (ums) et chiffre d'affaires TVA corrigé de l'année 2020 (ums_mod). Les méthodes de correction du chiffre d'affaires TVA sont décrites en section 3. La colonne nb. rep. indique le nombre de répondants dans la NOGA2.

NOGA2	nb. rep.	correl. EPT t010	correl. ums t010	correl. ums_mod t010	NOGA2	nb. rep.	correl. EPT t010	correl. ums t010	correl. ums_mod t010
0809	66	89.0	96.7	97.4	50	61	59.8	100.0	100.0
10	316	79.1	91.5	99.5	51	38	99.3	99.7	99.7
11	68	83.9	71.0	99.0	52	176	51.9	36.2	95.7
12	6	99.7	99.8	100.0	53	72	99.0	93.3	99.2
13	92	89.2	97.3	97.3	55	283	85.6	57.8	85.0
14	46	95.7	97.1	97.6	56	403	95.3	89.2	98.2
15	44	72.3	99.4	99.4	58	100	81.0	94.5	97.5
16	180	93.5	90.7	98.1	59	150	81.1	91.4	98.5
17	62	88.8	97.3	97.3	60	51	99.4	99.6	100.0
18	103	64.5	96.3	96.5	61	95	98.1	97.4	97.5
19	6	95.8	50.9	99.6	62	447	51.5	90.0	99.3
20	177	41.3	71.9	98.4	63	96	59.6	98.7	99.0
21	122	83.9	86.6	98.1	68	495	62.1	63.4	90.4
22	151	86.4	70.2	97.7	69	217	40.6	98.2	98.7
23	116	91.4	97.8	99.1	70	407	13.5	26.4	97.3
24	83	77.7	4.9	96.6	71	535	30.7	74.4	98.0
25	272	87.9	96.7	98.3	72	170	77.1	84.9	98.7
26	335	78.8	82.6	97.6	73	166	42.7	23.5	98.9
27	148	35.3	81.3	99.3	74	150	55.0	86.0	86.2
28	339	76.5	91.9	98.2	75	93	96.1	94.2	94.5
29	50	91.6	97.4	99.3	77	121	9.8	68.6	84.6
30	58	92.8	89.2	96.0	78	188	83.6	84.7	99.0
31	77	87.4	75.0	92.2	79	96	34.0	95.9	96.4
32	112	27.1	67.7	99.4	80	89	98.4	99.4	99.7
33	90	93.7	97.3	96.3	81	267	94.9	99.6	99.4
35	310	25.3	46.2	84.2	82	150	40.2	70.1	97.1
36	63	87.6	97.0	99.8	85	257	56.1	86.1	92.8
37	71	91.2	98.3	98.3	87	272	96.3	95.8	95.8
38	121	34.1	93.2	94.9	88	219	87.4	88.0	88.0
39	11	22.6	90.7	90.7	90	256	34.8	55.5	55.5
41	338	87.6	98.2	99.1	91	33	58.5	96.6	96.6
42	117	96.3	98.8	98.8	92	27	54.2	99.3	99.3
43	669	93.6	92.1	98.4	93	195	43.4	75.2	89.8
45	324	46.6	98.7	99.0	94	146	40.1	59.3	99.3
46	1635	19.5	94.7	41.9	95	53	98.3	95.4	99.7
47	714	98.5	99.5	99.9	96	180	96.0	98.1	98.3
49	381	90.3	98.5	99.8					

TABLE 11 Corrélations de Pearson (multipliées par 100), mesurées dans chaque NOGA2, entre la valeur ajoutée t029 mesurée pour la WS 2021 et les variables auxiliaires EPT_cadre, chiffre d'affaires TVA de l'année 2020 (ums) et chiffre d'affaires TVA corrigé de l'année 2020 (ums_mod). Les méthodes de correction du chiffre d'affaires TVA sont décrites en section 3. La colonne nb. rep. indique le nombre de répondants dans la NOGA2.

NOGA2	nb. rep.	correl. EPT t029	correl. ums t029	correl. ums_mod t029	NOGA2	nb. rep.	correl. EPT t029	correl. ums t029	correl. ums_mod t029
0809	66	92.6	80.6	81.8	50	61	67.4	99.4	99.4
10	316	80.8	93.2	95.7	51	38	99.4	98.9	98.9
11	68	90.0	61.2	94.8	52	176	85.4	29.7	56.7
12	6	99.6	99.7	100.0	53	72	99.9	93.6	97.4
13	92	89.1	94.8	94.8	55	283	80.9	53.4	76.7
14	46	94.0	98.1	99.1	56	403	96.3	84.6	92.7
15	44	66.3	89.2	89.2	58	100	86.5	95.0	96.4
16	180	95.2	87.6	96.8	59	150	56.8	92.3	78.9
17	62	93.5	82.6	82.6	60	51	99.9	99.0	99.8
18	103	92.7	64.2	66.2	61	95	99.1	98.9	99.0
19	6	98.6	79.7	87.9	62	447	77.0	85.8	80.2
20	177	48.7	44.6	71.9	63	96	86.0	91.1	84.7
21	122	68.7	73.9	86.5	68	495	71.0	65.5	77.6
22	151	88.8	67.5	92.6	69	217	67.0	96.4	96.2
23	116	94.5	91.3	93.8	70	407	53.1	46.9	75.0
24	83	82.5	29.1	82.4	71	535	55.7	57.2	75.0
25	272	89.1	89.8	91.7	72	170	44.8	51.5	70.4
26	335	84.9	81.6	92.0	73	166	52.1	24.5	82.2
27	148	88.5	48.9	40.4	74	150	56.9	59.0	54.3
28	339	82.6	91.3	92.7	75	93	94.6	90.7	91.1
29	50	93.2	99.5	99.2	77	121	16.5	81.4	89.1
30	58	80.2	74.9	74.0	78	188	88.1	85.8	93.7
31	77	88.9	73.4	89.9	79	96	-10.1	-76.6	-64.6
32	112	41.9	60.2	91.7	80	89	99.2	99.6	99.3
33	90	99.7	90.6	88.1	81	267	98.2	98.8	99.3
35	310	82.5	9.0	30.5	82	150	42.0	39.8	79.2
36	63	96.4	96.5	91.2	85	257	53.0	74.2	87.6
37	71	91.4	97.2	97.2	87	272	95.1	95.2	95.2
38	121	78.2	53.8	57.1	88	219	79.0	82.2	82.2
39	11	42.4	92.9	92.9	90	256	-28.4	21.6	21.6
41	338	95.9	85.7	87.6	91	33	-29.0	41.1	41.1
42	117	97.3	94.3	94.3	92	27	51.8	99.4	99.4
43	669	93.3	95.6	91.5	93	195	45.3	73.4	84.9
45	324	79.6	1.4	3.7	94	146	87.3	89.4	45.0
46	1635	50.8	57.5	75.6	95	53	97.6	90.4	93.8
47	714	99.6	98.2	99.0	96	180	94.8	97.5	98.1
49	381	87.9	95.2	94.7					

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. (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 Lohnindizes und deren Varianzschätzer. BFS-Nummer : 338-0026
- Graf, M. (2004). Enquête suisse sur la structure des salaires 2002. Plan d'échantillonnage et extrapolation pour le secteur privé. Numéro OFS : 338-0025
- Renaud, A. (2004). Analyse de données d'enquêtes. Quelques méthodes et illustration avec des données de l'OFS. Numéro OFS : 338-0024
- Renaud, A., Potterat, J. (2004). Estimation de la couverture du recensement de la population de l'an 2000. Echantillon pour l'estimation de la sous-couverture (P-sample) et qualité du cadre de sondage des bâtiments. Numéro OFS : 338-0023
- Graf, M. (2004). Fusion de données. Etude de faisabilité. Numéro OFS : 338-0022
- Potterat, J. (2003). Mietpreis-Strukturerhebung 2003. Entwicklung des Stichprobenplans und Ziehung der Stichprobe. BFS-Nummer : 338-0021

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Programme des publications de l'OFS

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

Les domaines statistiques

- 00 Bases statistiques et généralités
- 01 Population
- 02 Espace et environnement
- 03 Travail et rémunération
- 04 Économie nationale
- 05 Prix
- 06 Industrie et services
- 07 Agriculture et sylviculture
- 08 Énergie
- 09 Construction et logement
- 10 Tourisme
- 11 Mobilité et transports
- 12 Monnaie, banques, assurances
- 13 Sécurité sociale
- 14 Santé
- 15 Éducation et science
- 16 Culture, médias, société de l'information, sport
- 17 Politique
- 18 Administration et finances publiques
- 19 Criminalité et droit pénal
- 20 Situation économique et sociale de la population
- 21 Développement durable, disparités régionales et internationales

Les principales publications générales

L'Annuaire statistique de la Suisse



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

Le Mémento statistique de la Suisse



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

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

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

La banque de données des publications pour des informations détaillées

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

Vous souhaitez être parmi les premiers informés?



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

www.news-stat.admin.ch

STAT-TAB: la banque de données statistiques interactive



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

www.stattab.bfs.admin.ch

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



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

www.statatlas-suisse.admin.ch

Pour plus d'informations

Centre d'information statistique

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

Le plan de sondage de la statistique de la valeur ajoutée (WS) a été révisé pour l'enquête WS 2021. Des données administratives issues des décomptes de la taxe sur la valeur ajoutée (TVA) sont utilisées de manière extensive pour produire une stratification adaptée et quantifier la précision de l'enquête.

Cette révision du plan de sondage permet d'améliorer la précision de l'enquête tout en conservant une taille d'échantillon pratiquement constante. L'amélioration provient d'une sélection plus importante d'unités dont le chiffre d'affaires TVA est élevé. Cette sélection se fait via le choix de «Very Important Units» (VIU) ainsi qu'à travers une stratification revue, qui fait intervenir le chiffre d'affaires TVA.

Le calcul des taux de sondage dans chaque strate s'inspire de la méthode développée lors de la révision du plan de sondage de l'enquête Suisse sur la structure des salaires (ESS). Cette méthode permet l'intégration d'objectifs de précision multiples. Le gain d'efficacité obtenu se fait néanmoins au prix de contrôles plus importants sur les chiffres d'affaires TVA, afin de garantir une bonne corrélation entre ceux-ci et les réponses à l'enquête.

Pour la WS 2021, l'allocation mène à un échantillon brut de 22 912 unités. Les précisions attendues sont meilleures ou à minima égales à celles ciblées avec l'ancienne méthode, et ce pour chaque catégorie d'activité des entreprises.

Des objectifs de précision supplémentaires sont implémentés au niveau des grandes régions, cantons ou par catégories d'activités avec une nomenclature plus fine.

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

338-0089

ISBN

978-3-303-00761-7

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



Système d'indicateurs MONET 2030

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

**La statistique
compte pour vous.**

www.la-statistique-compte.ch