Rapport de méthodes

Statistique de l'emploi

Révision 2007: cadre de sondage et échantillonnage



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

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

Neuchâtel, 2008

Statistik der Schweiz

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

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

Statistique de la Suisse

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

0	Bases statistiques et produits généraux
1	Population
2	Espace et environnement
3	Vie active et rémunération du travail
4	Economie nationale
5	Prix
6	Industrie et services
7	Agriculture et sylviculture
8	Energie
9	Construction et logement
10	Tourisme
11	Mobilité et transports
12	Monnaie, banques, assurances
13	Protection sociale
14	Santé
15	Education et science
16	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 et disparités régionales et internationales

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Statistique de l'emploi

Révision 2007: cadre de sondage et échantillonnage

Auteur **Anne Renaud**
Office fédéral de la Statistique

Editeur **Office fédéral de la statistique**

Neuchâtel, 2008

Préambule

La section Enquêtes conjoncturelles (KE) de l'Office fédéral de la statistique (OFS) a demandé un soutien au service de Méthodes statistiques (METH) pour la révision de la statistique de l'emploi (STATEM). Le présent rapport documente la construction du cadre de sondage et l'échantillonnage effectués dans ce projet.

Un grand merci à Francis Saucy et Nathalie Nünlist (KE) pour la transmission des données et de toutes les informations nécessaires à la construction du cadre et du plan d'échantillonnage. Merci à Fabio Tomasini et Mathieu Gunzinger (BUR) pour les informations sur le Registre des entreprises et des établissements. Merci à Jann Potterat (METH) pour les importants travaux sur le suivi du cadre de sondage dans le temps. Merci également à Philippe Eichenberger et Paul-André Salamin (METH) pour les discussions au cours du projet. Merci enfin à Jean-Pierre Renfer (METH) pour la relecture du rapport.

Résumé

La statistique de l'emploi a été révisée en 2007 suite au dernier recensement des entreprises et des établissements. Cette révision a permis d'intégrer un nouveau système de cadre de sondage et de renouvellement de l'échantillon. Le plan d'échantillonnage a également été revu. Il a été construit dans le but de satisfaire les objectifs de précision de l'enquête tout en intégrant les cas spéciaux tels le «profiling» ou les multi-établissements. La stratification et l'allocation ont été revues fondamentalement. Elles sont accompagnées de remarques sur des alternatives pour une application future. Le nouvel échantillon contient 62'833 établissements, dont 41'553 normaux et 21'280 spéciaux. Il a été utilisé dès le 2ème trimestre 2007 et sera renouvelé, avec recouvrement, au 2ème trimestre 2008 sur la base d'un cadre de sondage actualisé.

Mots-clé

rapport de méthodes; cadre de sondage; plan d'échantillonnage; stratification; allocation optimale; numéro aléatoire permanent; cadre d'extrapolation; statistique de l'emploi; STATEM; BESTA

Complément d'information: Anne Renaud, tél. 032 713 62 65
Anne.Renaud@bfs.admin.ch

Réalisation: Service de méthodes statistiques, OFS

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

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

Numéro de commande: 338-0052

Prix: gratuit

Série: Statistique de la Suisse

Domaine: 0 Bases statistiques et produits généraux

Langue du texte original: Français

Graphisme/Layout: OFS

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

ISBN: 978-3-303-00407-4

Table des matières

1 Introduction **5**

1.1 Unité de tirage et population 5

1.2 Estimateurs et variables d'intérêt 6

1.3 Objectifs 6

2 Principes généraux **7**

2.1 Renouvellement annuel 7

2.2 Mutations 8

2.3 Valeurs auxiliaires de référence et extrapolation 9

3 Cadre de sondage **9**

3.1 Définition de la population 9

3.2 Types d'établissements 10

4 Plan d'échantillonnage et tirage **11**

4.1 Plan 2003 11

4.2 Révision 2007 12

5 Stratification **12**

5.1 NOGA2reg exhaustives 13

5.2 Classes de taille 13

5.3 Stratification finale 15

6 Allocation **18**

6.1 Allocation optimale 18

6.2 Répartition dans les strates 20

6.3 Allocation de base dans les strates 21

6.4 Allocation finale 21

7 Tirage **22**

8 Un peu plus sur le plan d'échantillonnage **25**

8.1 Stratification et allocation 25

8.2 Échantillons complémentaires 26

8.3 Précision des estimations 28

9 Conclusion **31**

Annexes 32

A Définitions 32

B Limite Take-All 35

C Méthode de Dalenius 36

D Extraits des programmes SAS 36

Références 40

1 Introduction

La statistique de l'emploi (STATEM) fournit des estimations trimestrielles des variables de l'emploi. Elle donne également des informations sur les places vacantes, les frontaliers, les pénuries/excédents ou encore les perspectives d'occupation.

La STATEM est révisée périodiquement à la suite de la publication des résultats des recensements des entreprises et des établissements. Ces révisions comprennent le renouvellement de l'échantillon, la mise à jour des méthodes d'estimation et l'adaptation rétrospective des séries aux données des derniers recensements.

La révision 2007 de la STATEM est réalisée après la publication en novembre 2006 des résultats du recensement des entreprises et établissements du troisième trimestre 2005. Elle fait suite également à une évaluation des méthodes d'estimations appliquées depuis la dernière révision de 2003.

Pour la révision 2007, il a été décidé de :
- revoir fondamentalement le système des cadres de sondage et des cadres d'extrapolation. Le but étant d'éviter les fluctuations provenant de la maintenance du Registre des entreprises et des établissements (REE), ainsi que l'effet des retards dans l'enquête auprès des nouveaux établissements (ERST). On vise une plus grande indépendance entre le cadre d'extrapolation et le REE, ainsi qu'un renouvellement annuel de l'échantillon ;
- revoir le plan d'échantillonnage afin de faire un tirage unique pour tous les objectifs de l'enquête et non pas un tirage pour l'échantillon national puis des compléments pour les cantons tel que cela avait été fait en 2003 ;
- revoir la méthode d'estimation, notamment les références pour le calage et la redistribution des relevés agrégés, et intégrer un traitement automatique des valeurs très influentes (robustification).
- proposer une estimation trimestrielle de la variance des estimateurs.

Le présent rapport documente les tâches liées à la définition du cadre de sondage, à l'échantillonnage et au tirage du nouvel échantillon. Seules les variables de l'emploi sont considérées. L'adaptation rétrospective des séries n'est pas abordée.

Le nouvel échantillon décrit dans ce document est à la base du relevé à partir du 2ème trimestre 2007. On notera que des estimations pour le 2ème trimestre 2007 seront également calculées avec l'ancien échantillon suivi depuis 2003. Dès le 3ème trimestre 2007, les estimations seront calculées pour le nouvel échantillon uniquement.

1.1 Unité de tirage et population

L'unité de tirage de la STATEM est l'établissement.

La population d'intérêt est formée de tous les établissements des deuxième et troisième secteurs économiques, actifs et localisés en Suisse.

Le cadre d'échantillonnage est basé sur le Registre des entreprises et des établissements (REE) ; voir détails dans la Section 3.

1.2 Estimateurs et variables d'intérêt

Les estimateurs d'intérêt sont des totaux pour l'ensemble des établissements de la population d'intérêt et les domaines suivants ; voir l'Annexe A pour les définitions :
– 49 activités économiques ;
– 7 grandes régions ;
– 7 grandes régions croisées avec les deux secteurs économiques ;
– 6 régions avec complément d'échantillon, *i.e.* les cantons GE, VD, FR, NE, SG et la ville de Zurich ;
– 6 régions avec complément d'échantillon, *i.e.* GE, VD, FR, NE, SG et la ville de Zurich, croisées avec les deux secteur économiques.

Notons que les objectifs au niveau des activités économiques concernent des NOGA2 individuelles ou regroupées. Les NOGA2 10 à 14 forment un seul domaine, tout comme 36 et 37, ainsi que 40 et 41. Ainsi, on passe de 55 à 49 activités économiques.

Les variables d'intérêt sont les effectifs suivants :
– VZM : hommes à plein temps
– VZF : femmes à plein temps
– TZ1M : hommes à temps partiel 1
– TZ1F : femmes à temps partiel 1
– TZ2M : hommes à temps partiel 2
– TZ2F : femmes à temps partiel 2
– TOTVZ = VZM + VZF : personnes à plein temps
– TOTTZ1 = TZ1M + TZ1F : personnes à temps partiel 1
– TOTTZ2 = TZ2M + TZ2F : personnes à temps partiel 2
– TOT = TOTVZ + TOTTZ1 + TOTTZ2 : emploi total
– EPT : équivalents plein temps.

On parle de plein temps pour les personnes avec un taux d'activité entre 90 et 100%. Le temps partiel 1 correspond à un taux d'activité entre 50 et 89%, et le temps partiel 2 à moins de 50%.

La variable "équivalents plein temps" EPT est construite comme combinaison linéaire des variables TOTVZ, TOTTZ1 et TOTTZ2. Les coefficients sont définis par section économique. Ils prennent des valeurs de l'ordre de 59-67% pour le temps partiel 1 (TOTTZ1) et de 29-33% pour le temps partiel 2 (TOTTZ2). Ces coefficients sont mis à jour régulièrement sur la base de l'enquête suisse sur la population active.

Pour l'établissement k de la section i, on obtient :

$$\text{EPT}_k = \text{TOTVZ}_k + \rho_{1i}\,\text{TOTTZ1}_k + \rho_{2i}\,\text{TOTTZ2}_k$$

avec ρ_{1i} et ρ_{2i} les coefficients de la section i ; voir Annexe A pour la définition des sections économiques et les coefficients utilisés.

1.3 Objectifs

Les objectifs de précision de l'enquête sont donnés sous la forme de coefficients de variations (CV) pour le total estimé des emplois en équivalents plein temps EPT. Ils sont identiques à ceux utilisés pour le plan d'échantillonnage de 2003.

Rappel : les CV mesurent la précision d'échantillonnage relative à l'estimateur. Soit le paramètre θ estimé par $\widehat{\theta}$ (exemple : le total Y dans la population est estimé par $\widehat{Y}$ basé sur l'échan-

tillon). La variance estimée de l'estimateur $\widehat{\theta}$ est notée $\widehat{var}(\widehat{\theta})$ et le CV estimé a la forme :

$$CV(\widehat{\theta}) = \frac{\sqrt{\widehat{var}(\widehat{\theta})}}{\widehat{\theta}}.$$

Le choix de la variable EPT permet de prendre en compte indirectement toutes les variables tout en mettant la priorité sur les équivalents pleins temps EPT, les temps pleins (majoritaires dans les EPT) et le total de l'emploi (très corrélé avec les EPT).

Liste des objectifs de précision pour le total des EPT :

1. CV=0.5% au niveau national ;
2. CV=2% par activité économique, *i.e.* NOGA2 avec quelques regroupements ;
3. CV=2% par grande région ;
4. CV=3% par grande région x secteur ;
5. CV=2% par région avec complément d'échantillon ;
6. CV=3% par région avec complément d'échantillon x secteur.

Les objectifs 2, 4 et 6 sont les plus contraignants. Une fois atteints, les autres objectifs sont également respectés.

2 Principes généraux

Cette section présente les principes généraux concernant le cadre de sondage et le renouvellement de l'échantillon. Les éléments essentiels de la révision 2007 sont résumés ici sans entrer dans les détails de l'application.

2.1 Renouvellement annuel

Le nouveau système de la STATEM fonctionne selon un rythme annuel de 4 trimestres.

Une fois par année, on renouvelle le cadre de sondage et l'échantillon. Le cadre de sondage renouvelé est construit à partir du REE actuel. L'échantillon renouvelé est sélectionné selon un algorithme qui permet un bon recouvrement avec l'ancien échantillon.

Le cadre de sondage et l'échantillon sont conservés pendant une année. Pendant l'année, il n'y a ni ajout de nouveaux établissements ni retrait d'établissements disparus. Le cadre de sondage est utilisé comme cadre d'extrapolation, c'est-à-dire comme information auxiliaire pour l'extrapolation. La liste et le contenu du cadre et de l'échantillon sont actualisés uniquement pour tenir compte des modifications de structures (mutations, contenu du profiling REE et des permanents STATEM ; voir définitions plus bas). Les adaptations visent le bon déroulement du relevé durant l'année entière : envoi à des établissements réels et codification des cas disparus[1].

Le renouvellement repose sur un algorithme de recouvrement maximal. Une grande partie de l'ancien échantillon reste donc dans le nouvel échantillon. Il y a quelques ajouts d'établissements apparus dans le cadre, le retrait des établissements disparus, ainsi que quelques ajouts

[1] Note : l'intégration d'une information complémentaires sur les nouveaux établissements sera évaluée prochainement.

et retraits d'établissements pour causes de modification des listes ou de l'allocation dans les strates.

Au moment du renouvellement, le relevé est effectué pour l'ancien et le nouvel échantillon afin de pouvoir traiter le possible saut dû au changement d'échantillon et de cadre d'extrapolation. Ce saut, traité par une adaptation de la série, ne devrait cependant pas être aussi conséquent que lors des derniers renouvellement de l'échantillon de la STATEM. La période entre renouvellements est en effet plus courte et on vise un bon recouvrement entre échantillons successifs.

Le premier échantillon de la STATEM révisée est sélectionné pour le relevé du 2ème trimestre 2007. Il est utilisé durant 4 trimestres plus le trimestre de renouvellement, *i.e.* jusqu'au 2ème trimestre 2008. De même pour le cadre de sondage. Un nouveau cadre et un nouvel échantillon seront définis pour le 2ème trimestre 2008 et l'année qui suit.

Le système de renouvellement annuel de l'échantillon permet de limiter les problèmes d'érosion de l'échantillon dans le temps. Il permet également de tenir compte de l'évolution de la population de référence, tout en évitant d'être trop dépendant des fluctuations dus aux travaux de maintenance du REE. Il est important de noter que ce système nécessite une coordination entre les travaux d'actualisation dans le REE et l'extraction d'un nouveau cadre de sondage.

2.2 Mutations

Les établissements sont sujets à mutations dans le temps. On observe des fermetures avec ou sans successeurs, des changements d'entreprise, ou encore des fusions avec un ou plusieurs autres établissements. Ces mutations sont détectées par comparaison de deux états du REE dans un processus spécialement développé lors de la révision de la STATEM et nommé *reporting*. Elles sont utilisées pour ajuster le cadre de sondage pendant son année d'utilisation[2].

Un établissement fermé sans successeur est identifié mais conservé dans le cadre. S'il fait partie de l'échantillon, l'envoi de questionnaire cesse.

Un établissement fermé avec successeur est remplacé par son successeur. S'il était dans l'échantillon, on intègre le successeur dans l'échantillon. Ce dernier reprend le poids d'échantillonnage de son prédécesseur.

Un changement d'entreprise peut avoir une influence sur les envois mais ne modifie pas les caractéristiques de l'établissement.

Les établissements qui fusionnent sont remplacés par le résultat de la fusion avec un nombre équivalents plein temps correspondant au total selon le cadre de référence. L'établissement fusionné fait partie de l'échantillon à partir du moment où une de ses parties faisait partie de l'échantillon. Le poids d'échantillonnage de cet établissement ne peut par contre être repris. On le détermine par la méthode du partage des poids (ou fusion des poids dans ce cas) ; voir Lavallée (2002) ou Lavallée et Deville (2002) pour les bases théoriques et Renaud (2006) pour un exemple d'application dans le cas de fusions de communes.

Lorsqu'un établissement k avec x_k équivalents plein temps et un poids d'échantillonnage d_k fusionne avec un autre établissement ℓ avec x_ℓ équivalents plein temps et un poids d'échantillonnage d_ℓ, on obtient un établissement fusionné j avec $x_j = x_k + x_\ell$ équivalents plein temps. Son poids est donné par :

$$d_j = \frac{d_k x_k + d_\ell x_\ell}{x_k + x_\ell}.$$

[2]Notons que la procédure d'ajustement n'est pas appliquée aux établissements du profiling car les informations dans le REE ne permettent actuellement pas d'identifier toutes les mutations ; voir les détails plus bas.

La formule peut être étendue à la fusion de plus de deux établissements. Le poids d'échantillonnage des établissements hors échantillon est fixé à 0 avant l'application de la méthode du partage des poids. L'établissement provenant de la fusion de deux établissements avec le même poids garde le poids inchangé. Il recevra un poids se trouvant entre ceux des deux établissements fusionnés s'ils ont des poids différents. La taille de l'établissement, en équivalent plein temps, est intégrée dans la formule pour pondérer l'influence des établissements fusionnés. Un grand établissement aura une influence plus grande dans la construction du poids qu'un petit établissement. La méthode n'introduit aucun biais dans les estimations.

2.3 Valeurs auxiliaires de référence et extrapolation

Le cadre de sondage conservé pour une année subit de légères adaptations dues aux mutations annoncées dans le reporting. Il est ainsi ajusté chaque trimestre au moment de l'envoi des questionnaires et au moment des estimations. Ce cadre sert de base non seulement à la définition de l'échantillon mais également à la définition des valeurs de référence utilisées comme informations auxiliaires pour la détermination des poids d'extrapolation.

Les poids d'extrapolation pour les estimations des totaux de la STATEM prennent en compte les poids d'échantillonnage (év. adaptés par le partage des poids), une robustification de valeurs trop influentes et un ajustement pour tenir compte de la non-réponse et des informations auxiliaires ; *i.e.* totaux des équivalents plein temps du cadre dans différents sous-groupes.

La détection des valeurs aberrantes et la robustification sont des processus nouveaux dans la STATEM. Ils permettent de détecter des erreurs manifestent dans les données puis de traiter les cas extrêmes restant sans modification des valeurs relevées si ces dernières sont considérées correctes. Seuls les poids sont potentiellement modifiés.

L'ajustement des poids pour tenir compte de la non-réponse et des totaux de référence est fait en une unique étape de calage directement à partir de l'échantillon des répondants. On ne passe pas par une correction de la non-réponse homogène par strate puis par un estimateur par le quotient comme fait entre 2003 et 2007. Ce choix repose sur un désir de stabilisation des estimations en évitant de travailler au niveau strates. Ces dernières sont en effet multiples et souvent très petites.

Les totaux de la STATEM sont estimés à partir du poids final d'extrapolation. Ils sont accompagnés d'une estimation de la variance de l'estimation. La pondération et les estimations seront documentées plus en détails dans un futur rapport.

3 Cadre de sondage

Le cadre de sondage utilisé pour le tirage de l'échantillon révisé est un sous-ensemble des établissements du Registre des entreprises et des établissements (REE) au 31 mars 2007 (fichier `week_loc_20070402`).

3.1 Définition de la population

Les établissements gardés dans le cadre ont les caractéristiques suivantes, voir codes dans l'Annexe A :
– activité économique des secteurs secondaire et tertiaire : `10 <= NOGA2 <= 93` ;

– établissement actif : `status=1` ;
– vrais établissements avec des emplois :
`betyp in ('L13','L17','L18','L20','L21','L22','L23','L24','L25','L27')`[3] ;
– localisation en Suisse, *i.e.* par ex. exclusion de quelques établissements en Allemagne et au Liechtenstein.

Le cadre reprend les variables permettant d'identifier l'établissement et son entreprise, ainsi que ses caractéristiques telles l'activité économique et la localisation. Il reprend également les variables d'emploi total (`betot`) et en équivalents plein temps (`equivpt`). Ces dernières proviennent du recensement 2005 ou de sources plus récentes.

Quelques variables du niveau entreprise sont collées aux établissements. Il s'agit notamment de l'identification des groupes d'entreprises et le nombre d'établissements dans l'entreprise.

Le cadre initial est formé de 385'098 établissements.

3.2 Types d'établissements

Plusieurs types d'établissements existent dans la STATEM. On parle d'établissements normaux (type par défaut) et d'établissements spéciaux. Les spéciaux sont formés des *multi-établissements*, des établissements du *profiling REE* et des établissements *permanents STATEM*. L'échantillon de la STATEM est formé par l'ensemble des établissements spéciaux et une partie des normaux.

Multi-établissements. Le relevé des emplois est réalisé sous la forme de valeurs agrégées et non pas détaillées au niveau des établissements. Après le relevé, une phase préliminaire distribue les emplois relevés dans les établissements individuels en fonction des équivalents pleins temps présents dans le cadre de sondage. Une fois la redistribution effectuée, les établissements sont traités comme les établissement dits "normaux".

Profiling REE. Un ensemble d'établissements du REE font partie de groupes d'entreprises ayant un contact particulier avec l'OFS. Ces groupes livrent trimestriellement les données de l'emploi pour l'ensemble de leurs entreprises et établissements. Ils envoient les données selon la structure du trimestre en cours. Les données sont reprises en bloc dans le REE chaque trimestre sans information sur les mutations interne telles que fermetures avec ou sans successeurs ou fusions. Des imputations sont appliquées dans la STATEM lorsque des données manquent. Notons également que l'on définit le profiling actuel (groupes qui fonctionnent déjà dans le système du profiling) et le profiling futur (groupes qu'il est prévu d'intégrer dans le profiling). On différencie également le profiling privé du profiling public[4].

Permanents STATEM. Quelques établissements ont un statut de "permanent STATEM". Il s'agit d'un statut proche du profiling REE mais dans le cadre spécifique de la STATEM uniquement. Ces permanents sont traités comme le profiling pour tout ce qui concerne l'échantillonnage et le cadre de sondage. Ils sont par contre traités avec les normaux dans les estimations. Il n'y a par exemple pas d'imputation généralisée en cas de non-réponse.

Les normaux représentent 94.5% des établissements du cadre et 81% des emplois ; voir Table 1. Ainsi les 21'280 établissements spéciaux, tous pris dans l'échantillon, représentent 19% des emplois total de la population d'intérêt. On notera également que les 2'270 établissements notés "parties de multi-établissements" forment 294 multi-établissements pour le relevé.

[3] Note : les réactivations `L17` et les nouvelles entrées `L18` ne restent pas dans le cadre car leur statut actuel est inconnu (`status=4`). Il sera déterminé lors de l'enquête auprès des nouveaux établissements (ERST).

[4] Note : La liste des groupes profiling évolue durant l'année.

TAB. 1 Types des établissements dans le cadre avec nombre d'établissements et emploi total betot (nombres absolus et relatifs).

Type	Nombre établ.	%établ.	Somme betot	%betot
Normal	363'818	94.5	3'011'142	81.0
Parties de multi-établissements	2'270	0.6	130'464	3.5
Profiling REE privé actuel	11'006	2.9	281'951	7.6
Profiling REE privé futur	364	0.1	17'075	0.5
Profiling REE public actuel	5'295	1.4	171'338	4.6
Profiling REE public futur	2'028	0.5	82'271	2.2
Permanent STATEM	317	0.1	23'876	0.6
Total	385'098	100.0	3'718'117	100.0

Dernière minute : Quelques modifications de dernière minute ont été effectuées dans le cadre de sondage. Ainsi l'activité économique de 14 établissements est modifiées directement dans le cadre.

4 Plan d'échantillonnage et tirage

Cette section présente brièvement le plan d'échantillonnage et la méthode de tirage de 2003, ainsi que les nouveautés de la révision 2007. La méthode appliquée pour la révision 2007 sera décrite plus en détails dans les sections 5 à 7.

4.1 Plan 2003

Le plan d'échantillonnage de la révision 2003 était composé de deux phases successives : le plan national et le plan des cantons et villes ayant demandé des résultats régionaux.

Dans le plan national, la stratification correspond au croisement de l'activité économique avec les grandes régions et les classes de taille. Les établissements spéciaux tels les permanents et les multi-établissements (appelés aussi "fictifs") sont tous dans l'échantillon. Le reste de l'échantillon repose sur l'allocation optimale des équivalents pleins temps du cadre de sondage. Des adaptations sont ensuite effectuées pour avoir une taille minimale et un taux de sondage minimal par strate. L'échantillon national est tiré selon la méthode de Bernoulli. Il n'est donc pas de taille fixe et des adaptations manuelles sont faites au niveau de l'allocation afin d'avoir assez d'éléments dans chaque strate.

La stratification au niveau des régions correspond à un partage des strates du plan national en régions suivi d'un regroupement de strates de taille. Les établissements tirés dans l'échantillon national sont exclus du tirage régional. L'allocation est définie en tenant compte du nombre d'établissements déjà sélectionnés dans l'échantillon national.

Les classes de taille sont construites selon un processus relativement complexe. Dans chaque strate primaire de l'échantillon national correspondant au croisement entre la NOGA2 et la grande région, il y a de 1 à 6 classes de tailles différentes (basées sur les équivalents pleins temps). Les classes de base sont :]0-5],]5-9],]9-19],]19-49],]49-99],]99,99'999]. Des regroupements sont effectués afin d'éviter des strates trop petites. Lors de la détermination des strates de tirage pour les compléments régionaux, de nouveaux regroupements sont effectués.

Au final, un code de taille identique peut correspondre à deux définitions différentes.

Chaque trimestre, l'échantillon STATEM était adapté pour tenir compte de la disparition d'établissements ou encore intégrer des nouveaux dans les strates de grands établissements, dans les permanents et dans les multi-établissements.

4.2 Révision 2007

La révision 2007 propose un unique tirage d'établissements. Il n'y pas de phases successives, national puis régional, comme en 2003. Le plan d'échantillonnage combine tous les objectifs de l'enquête. Stratification, allocation et tirage sont donc revus dans la révision 2007.

Les strates de tirage sont définies par les objectifs de l'enquête. Elles intègrent les types d'établissements pour définir les parties exhaustives et non exhaustives. La définition des classes de taille est revue fondamentalement, voir Section 5.

L'allocation est déterminée par étapes. Plusieurs allocations optimales sont premièrement déterminées pour satisfaire les principaux objectifs de précision. Les strates de travail sont des regroupements des strates de tirage. Ces allocations sont ensuite utilisées pour définir l'allocation finale au niveau des strates de tirage. Comme en 2003, les objectifs sont exprimés en coefficients de variation pour les équivalents plein temps. On fait également des ajustements, notamment pour atteindre une taille minimale par strate, voir Section 6.

Le tirage utilise le principe des nombres aléatoires permanents définis pour chaque établissement du cadre, voir Section 7. La méthode permet de sélectionner un échantillon de taille fixe dans chaque strate. Le tirage à taille fixe a comme avantage d'éviter un éloignement parfois important entre le nombre d'unités allouées et tirées dans les différentes strates. Il est moins adéquat dans le cas d'une rotation d'échantillon mais aucune rotation n'est prévue pour les prochaines années.

5 Stratification

Les variables à considérer pour la stratification sont au nombre de quatre, voir définitions dans l'Annexe A :
- Type d'établissement (`special`=0 pour les normaux, `special`=1 pour les multi-établissements et `special`=2 pour le profiling et les permanents).
- Activité économique NOGA2 avec 3 regroupements (`NOGA2reg`).
- Régions avec objectifs complémentaires et restes des grandes régions (`region_new`).
- Classe de taille.

L'information sur le secteur est comprise dans `NOGA2reg` et celle sur la grande région dans `region_new`. Les classes de tailles doivent être définies.

Les croisements entre `NOGA2reg`, `region_new` et les classes de taille déterminent les strates utilisées pour le tirage dans les établissements normaux. Les établissements spéciaux (profiling, permanents et multi-établissements) sont tous pris avec probabilité 1.

5.1 NOGA2reg exhaustives

Plusieurs petites NOGA2reg sont prises de manière exhaustive dans l'échantillon. Il s'agit des activités suivantes : 10-14, 16, 19, 21, 23, 27, 30, 34, 35, 61, 62 et 73 ; voir définitions dans l'Annexe A.

Codification :

```
exnoga=0;
if noga2reg in (1014,16,19,21,23,27,30,34,35,61,62,73) then exnoga=1;
```

5.2 Classes de taille

La classe de taille est un élément important de la stratification car la taille est fortement liée aux variables d'emploi relevées durant l'enquête.

On choisit, comme en 2003, les équivalents pleins temps (EPT, variable equivpt) pour définir les classes de taille. On vise par contre une définition unique des classes de tailles pour l'ensemble du plan d'échantillonnage.

Les classes de taille sont déterminées pour les établissements normaux du cadre. On commence par définir la limite des "take-all", *i.e.* limite au dessus de laquelle l'échantillon est pris de manière exhaustive, puis on évalue un partage au dessous de cette limite. La classe de taille "take-all" sera prise en compte dans l'allocation optimale. Cette méthode est souvent appliquée en cas de distribution très asymétriques. Dans la STATEM, on a typiquement un groupe de grands établissements avec un emploi très variable. Le fait de les prendre tous dans l'échantillon permet de réduire la variabilité des estimateurs. Il faut cependant noter que cette réduction de variance est réellement observée uniquement si le taux de réponse est excellent parmi ces grands établissements.

Limite take-all

L'algorithme de Hidiroglou (1986) est appliqué dans chaque NOGA2reg et dans chaque croisement secteur x region_new ; voir Appendice B pour une brève description de la méthode. Les CV objectifs sont fixés à 2% dans les NOGA2reg et 3% dans les croisements secteur x region_new. Dans les deux cas, deux itérations suffisent à stabiliser les estimations (asymétrie ou "skewness" raisonnable). Les limites déterminées pour les NOGA2reg sont inférieures à celles pour les croisements secteur x region_new. On choisit donc le niveau NOGA2reg pour fixer les limites. De plus, nous choisissons de garder deux limites possibles : 20 et 50 EPT afin d'uniformiser la définition des classes.

La limite est fixée à 20 EPT si la limite lim déterminée par l'algorithme est inférieure à 50 EPT. Elle vaut 50 sinon.

Définition :

```
if lim>=50 then limTakeall=50;
else limTakeall=20;
```

Les trois classes de taille de base deviennent :]0-20[, [20-50[et [50,∞[. Si la limite des take-all est 20, alors on prend exhaustivement les deux classes [20-50[et [50,∞[. Si la limite est 50, alors on prend exhaustivement la classe [50,∞[.

La démarche aboutit à une sélection exhaustive de tous les établissements avec 50 EPT ou plus. Dans plusieurs NOGA2reg, les établissements sont pris de manière exhaustive également s'ils ont entre 20 et 50 EPT. Les classes de tailles exhaustives contiennent en tout 12'316 établissements normaux. Ils s'ajoutent aux 21'280 établissements spéciaux pris également exhaustivement. Si l'on ajoute encore les NOGA2reg exhaustives, on note qu'environ 33'600 établissements seront pris par défauts dans l'échantillon STATEM.

Partage sous la limite de 20 EPT

Nous choisissons de construire 2 classes de tailles pour les établissements avec moins de 20 EPT. Le choix de la limite entre les deux classes est effectué à l'aide de la méthode de Dalenius appliquée sur les EPT dans chaque NOGA2reg ; voir Appendice C pour la méthode.

La méthode indique un partage entre 3 et 7 EPT selon les NOGA2reg, avec un maximum de cas à 5 et 6. Nous choisissons donc la limite 5 EPT pour toutes les NOGA2reg.

Les classes de tailles du plan d'échantillonnage sont donc définies comme suit : tailles (tai) :]0-5[, [5-20[, [20-50[, [50,∞[.

Pour les besoins du plan, nous définissons des classes de taille tai pour l'ensemble des établissements du cadre, spéciaux et normaux. La variable tai2 regroupe alors les spéciaux dans une classe unique notée tai2=99.

Définition :

```
* tailles (si equivpt);
tai=.;
if 0<equivpt<5 then tai=1;
else if 5<=equivpt<20 then tai=2;
else if 20<=equivpt<50 then tai=3;
else if 50<=equivpt then tai=4;
* regroupement des speciaux;
* special: 0 normaux, 1 multi, 2 profiling REE et permanents STATEM;
tai2=tai;
if special in (1,2) then tai2=99;
```

On note que les établissements spéciaux se trouvent dans toutes les classes de taille tai. Une bonne partie de ces derniers sont d'ailleurs très petits ; voir Table 2.

TAB. 2 Répartition des établissements normaux (special=0), multi (special=1) et profiling-permanents (special=2) dans les classes de taille tai.

	tai				
	1	2	3	4	All
special					
0	268075	74345	14449	6949	363818
1	797	758	389	326	2270
2	8292	6355	2434	1929	19010
All	277164	81458	17272	9204	385098

Les croisements entre les NOGA2reg et les classes de taille tai2 sont souvent petits ; voir Table 3. Par contre le minimum est de 59 éléments si l'on se restreint aux NOGA2reg et aux

classes de tailles non exhaustives. On note également que certaines NOGA2reg n'ont pas d'établissements spéciaux (tai2=99) alors que d'autre ont la majorité des établissements dans les spéciaux (ex. NOGA2reg=64, Postes et télécommunications).

Les croisements region_new x secteur x tai2 ont une taille supérieure à 100 dans toutes les cellules potentiellement non exhaustives (tai2 in (1,2,3)) ; voir Table 4. La grande majorité des établissements spéciaux sont dans le secteur tertiaire.

5.3 Stratification finale

Le croisement des variables special, NOGA2reg, region_new et tai2 forment 2827 strates de tirage. Ce découpage fin correspond en fait à 1664 strates prises exhaustivement (spéciaux, grandes tailles, NOGA2reg exhaustives) et 1163 strates pour lesquelles il faudra déterminer une allocation. La partie prise exhaustivement correspond à 9.4% des établissements et 52.7% des emplois du cadre ; voir Table 5.

Définition de tako0 :

```
* info sur strates exhaustives (exnoga, special et limtakeall);
tako0=0;
if (exnoga=1) or (special in (1,2)) or
   (limTakeAll=20 and tai2 in (3,4)) or
   (limTakeAll=50 and tai2 in (4)) then tako0=1;
```

Une bonne partie des strates dans lesquelles ont doit définir une allocation contiennent très peu d'établissements. Il faut cependant noter que les strates sont utilisées uniquement pour le tirage et non pas comme cellules de correction de la non-réponse et de calage durant les estimations. Le découpage fin utilisé au moment du tirage ne met donc pas à mal les estimations et leur stabilité. Pour une discussion complémentaire sur la stratification, voir la Section 8.

TAB. 3 Répartition des établissements par `NOGA2reg` et classes de taille `tai2`. Note : `exnoga`=1 si la NOGA2reg est exhaustive et 0 sinon. `limTakeAll` est la limite au dessus de laquelle les classes de tailles sont prises exhaustivement (si `exnoga`=0). Si `limTakeAll`=20, les classes de taille `tai2 in (3,4,99)` sont exhaustives. Si `limTakeAll`=50, les classes `tai2 in (4,99)` sont exhaustives.

			tai2					
			1	2	3	4	99	All
noga-2reg2	exno-ga	limTa-keAll						
10.14	1	20	171	147	61	10	4	393
15	0	50	1444	655	170	186	137	2592
16	1	50	3	3	3	3	3	15
17	0	20	351	147	72	51	2	623
18	0	20	693	81	19	11	.	804
19	1	20	159	30	12	7	.	208
20	0	20	4487	1657	246	50	11	6451
21	1	50	71	60	34	58	.	223
22	0	20	2822	925	232	133	16	4128
23	1	20	4	5	2	3	.	14
24	0	50	426	234	139	171	23	993
25	0	50	325	274	126	127	3	855
26	0	20	917	322	119	75	18	1451
27	1	50	92	84	34	71	.	281
28	0	50	4814	2117	578	275	20	7804
29	0	50	1946	1005	391	390	18	3750
30	1	20	79	33	16	8	.	136
31	0	50	573	274	114	112	15	1088
32	0	50	375	174	59	75	7	690
33	0	50	2123	766	241	294	24	3448
34	1	50	92	54	23	18	.	187
35	1	50	296	97	17	25	16	451
36.37	0	20	2961	603	173	82	5	3824
40.41	0	50	383	271	135	72	78	939
45	0	50	24396	9841	2118	687	383	37425
50	0	20	11308	3797	428	107	62	15702
51	0	50	15628	4506	1160	483	200	21977
52	0	20	36788	8624	813	253	2121	48599
55	0	50	18927	8443	1069	251	376	29066
60	0	20	5508	1490	398	160	962	8518
61	1	50	51	33	25	9	.	118
62	1	20	112	84	17	10	6	229
63	0	20	2828	980	209	138	123	4278
64	0	20	465	161	47	43	4848	5564
65	0	50	1661	1076	272	214	1222	4445
66	0	50	614	345	134	77	1040	2210
67	0	20	3651	699	154	23	501	5028
70	0	20	4414	710	98	34	108	5364
71	0	20	785	139	17	5	4	950
72	0	20	9749	1604	319	156	53	11881
73	1	20	421	119	40	38	45	663
74	0	20	52271	8239	1051	514	274	62349
75	0	20	2781	1723	317	87	3572	8480
80	0	50	7814	3138	928	224	3133	15237
85	0	50	18242	4589	1312	956	1298	26397
90	0	20	1210	527	128	24	123	2012
91	0	20	4350	1190	155	61	26	5782
92	0	20	5704	899	164	69	340	7176
93	0	20	12790	1371	60	19	60	14300
All			268075	74345	14449	6949	21280	385098

TAB. 4 Répartition des établissements selon la région, le secteur et la classe de taille `tai2`.

		tai2					
		1	2	3	4	99	All
region_new	secteur						
10	2	1931	883	220	72	38	3144
	3	8475	2474	313	102	752	12116
11	2	2315	705	168	112	30	3330
	3	13788	3582	682	374	1222	19648
12	2	3921	1507	319	174	57	5978
	3	17710	4473	732	350	2241	25506
20	2	8883	3633	935	570	131	14152
	3	32138	8547	1441	508	3396	46030
21	2	1606	618	182	91	14	2511
	3	5636	1467	202	92	1152	8549
22	2	1233	541	133	113	11	2031
	3	4430	1214	213	84	377	6318
30	2	6677	2399	684	450	114	10324
	3	29512	6621	1206	567	2589	40495
40	2	6107	2212	582	298	92	9291
	3	25683	5844	1189	481	1601	34798
41	2	1720	548	143	97	41	2549
	3	15742	4003	792	402	1987	22926
50	2	4915	2121	536	271	49	7892
	3	18315	4553	709	257	1459	25293
51	2	3449	1526	420	262	76	5733
	3	12589	3215	537	192	1187	17720
60	2	5126	2217	544	314	85	8286
	3	22508	5686	899	346	1358	30797
70	2	2120	949	268	170	45	3552
	3	11546	2807	400	200	1176	16129
All		268075	74345	14449	6949	21280	385098

TAB. 5 Répartition des établissements N et des emplois `betot` du cadre dans les strates exhaustives (`tako0=1`) et non exhaustive (`tako0=0`). Note : résultats avant l'allocation.

	N	%N	betot	%betot
tako0				
0	349066	90.6	1759639	47.3
1	36032	9.4	1958478	52.7
All	385098	100.0	3718117	100.0

6 Allocation

L'allocation est effectuée en plusieurs étapes.

1. On effectue premièrement une allocation optimale pour les domaines objectifs les plus contraignants de l'enquête :
   – activités économiques NOGA2reg (49 objectifs),
   – grande région gr x secteur (14 objectifs), et
   – région avec échantillon complémentaire x secteur (12 objectifs).
   Ces allocations sont faites au niveau de strates de travail correspondant aux différents objectifs croisés avec les classes de taille tai2. Elles n'utilisent donc pas les strates de tirages mais des agrégats de ces dernières. Pour chaque domaine objectif, on détermine une taille d'échantillon optimale et la répartition dans les classes de taille tai2 afin de satisfaire la précision désirée donnée par un CV. A la fin de cette étape, on obtient trois groupes d'allocations optimales. Un pour les strates de travail NOGA2reg x tai2, un pour les strates de travail gr x secteur x tai2 et un pour les strates de travail region_new x secteur x tai2 avec restriction aux region_new $\in$ {11, 12, 21, 22, 41, 51}.
2. On procède ensuite à une répartition des différentes tailles déterminées dans 1. dans les croisements NOGA2reg x region_new x tai2, c'est-à-dire dans les strates de tirage $h = 1, .., H$. La répartition est réalisée proportionnellement à la taille tout en tenant compte des strates exhaustives par défaut. Le résultat de cette étape est un ensemble de deux ou trois valeurs par strate h. Chaque valeur provenant d'un objectif différent. Il y a deux valeurs dans les strates de region_new sans objectifs de précision et trois valeurs dans les strates de region_new avec objectifs de précision.
3. Dans l'étape suivante, on choisit le maximum entre les deux ou trois valeurs déterminées dans chaque strate h à l'étape 2. L'idée est de respecter tous les objectifs de taille.
4. Finalement, on adapte l'allocation afin d'avoir des tailles suffisantes au niveau des strates de tirage.

La figure 1 illustre l'allocation en étapes successives. Chaque domaine objectif correspond à une partie du cube. Une allocation optimale est effectuée dans les classes de taille tai2 de chacun de ces domaines objectifs (strates de travail). La répartition dans les strates de tirage permet ensuite de combiner les différents résultats dans un plan unique. Notons que les spéciaux ne sont pas exclus de l'allocation. Ils sont tous regroupés dans la classe de taille tai2=99 (exhaustive). Dans l'application, la classe de taille tai2=99 est partagée en deux parties : special=1 et special=2.

La méthode proposée effectue une allocation optimale au niveau des domaines objectifs x tai2 à l'aide strates de travail puis une répartition proportionnelle dans les strates de tirage. Elle ne fait pas d'allocation optimale directement au niveau des strates de tirage. Cette méthode a été choisie de façon a orienter l'allocation sur les objectifs réels de l'enquête et éviter les problèmes de stabilité dans l'allocation. Les strates de travail sont en effet plus grandes que les strates de tirage et les estimateurs de variance utilisés dans l'allocation optimale sont ainsi plus stables ; voir formules ci-dessous.

6.1 Allocation optimale

Soit d le domaine objectif (ex. NOGA2reg=50). On notera par $g = 1, .., G$ les strates de travail correspondant au croisement entre d, les cinq classes de taille tai2 et la variable tako0.

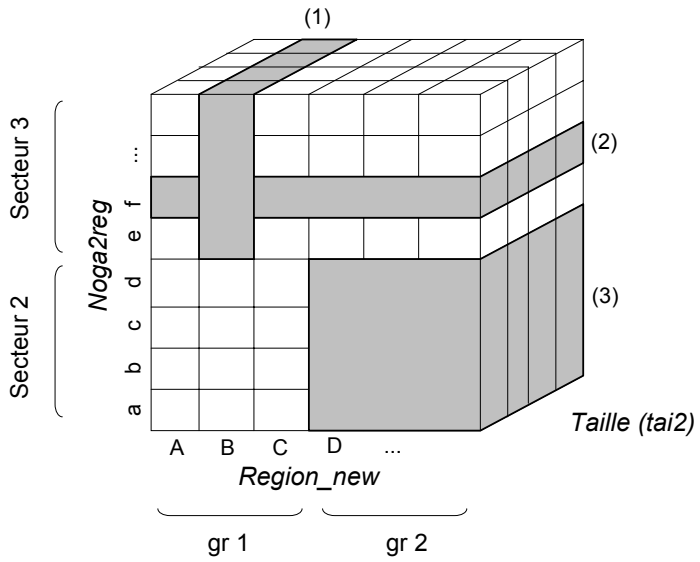


FIG. 1 Illustration du cube des strates de tirage `NOGA2reg` x `region_new` x `tai2`. Chaque domaine objectif correspond à une partie du cube. Les trois domaines suivants sont identifiés par des parallélépipèdes grisés : (1) (`secteur`=3 et `region_new`=B) x `tai2`, (2) (`NOGA2reg`=f) x `tai2` et (3) (`secteur`=2 et `gr`=2) x `tai2`. Note : regroupement des spéciaux (`special` in (1,2)) dans `tai2`=99, et variable `tako0` pas représentée ; voir texte pour description complète.

La variable `tako0` identifie les strates de tirage exhaustives par défaut, c'est-à-dire les `NOGA2reg` et les classes de taille exhaustives : `exnoga=1`, `tai2`=99 (spéciaux), `tai2`=4 (normaux avec au moins 50 EPT), et une partie de `tai2`=3 (normaux entre 20 et 50 EPT) selon la limite `limTakeAll` définie au niveau `NOGA2reg`.

Si le domaine objectif est une `NOGA2reg`, il n'y a qu'une valeur `tako0` par `tai2`. Les strates de travail $g = 1, ..G$ sont donc équivalentes à d x `tai2` (G=5 croisements). Si le domaine objectif est un croisement entre une grande région ou une région avec le secteur, il est possible d'avoir un partage de `tai2`=3 en deux parties : une partie exhaustive et une partie non exhaustive. Le nombre de strates de travail dans d sera donc égal à G=6 et non pas 5.

La strate de travail g contient N_g établissements dans la population. L'ensemble du domaine d contient $N = \sum_g N_g$ établissements.

Prenons maintenant la variable auxiliaire x connue pour tous les établissements du cadre de sondage ($x =$`equivpt`). Cette variable est utilisée pour construire l'allocation car elle est bien corrélée aux variables d'intérêt de l'enquête. Dans le domaine d, la somme des x est notée X.

Dans le cas classique, l'allocation optimale n_g dans les strates de tirage $g = 1, .., G$ pour le CV objectif *CVo* est donné par (Cochran, 1977) :

$$n_g = n_{opt} \frac{N_g S_g}{\sum_g N_g S_g},$$

avec

$$n_{opt} = \frac{(\sum_g N_g S_g)^2}{CVo^2 \, X^2 + \sum_g N_g S_g^2},$$

et

$$S_g = \frac{1}{N_g - 1} \sum_{k \in g} (x_k - \overline{x}_g)^2.$$

Dans notre application, certaines strates de travail g du domaine d sont exhaustives par défaut (`tako0`=1). Les formules sont par conséquent adaptées.

Si l'on note ta l'ensemble des g exhaustives ("take-all", `tako0`=1) et ts l'ensemble des g non exhaustifs ("take-some", `tako0`=0), on obtient (Hidiroglou, 1986) :

$$n_g = \begin{cases} N_g & \text{si } g \text{ est dans } ta \\ (n_{opt} - N^*) \frac{N_g S_g}{\sum_{g \in ts} N_g S_g} & \text{si } g \text{ est dans } ts \end{cases}$$

avec le nombre d'éléments dans les g exhaustifs :

$$N^* = \sum_{g \in ta} N_g,$$

la taille optimale

$$n_{opt} = N^* + \frac{(\sum_{g \in ts} N_g S_g)^2}{CVo^2 \, X^2 + \sum_{g \in ts} N_g S_g^2}$$

et

$$S_g = \frac{1}{N_g - 1} \sum_{k \in g} (x_k - \overline{x}_g)^2.$$

Note : Dans la STATEM, les strates de travail sont suffisamment grandes et homogènes en taille x pour éviter les cas d'allocations optimales n_g supérieures à leur taille N_g.

On remarquera également que les estimations S_g pour $g = 1.,,,G$ sont utilisées uniquement pour les strates de travail non exhaustives par défaut. Elles sont considérées comme stables car le nombre d'observations utilisées est généralement important, voir Tables 3 et 4. L'allocation optimale repose donc sur des bases solides. Pour plus de détails sur l'application, voir l'Annexe D.

6.2 Répartition dans les strates

Soit d le domaine objectif et $g = 1, .., g$ les strates de travail correspondant aux croisement entre le domaine d, les cinq classes de taille `tai2` et `tako0`. Les strates de travail g sont des agrégats de strates de tirage $h = 1, ..H_g$.

La répartition des n_g dans les strates de tirage h est donnée par :

$$n_h = \begin{cases} N_h & \text{si } h \text{ est exhaustive par défaut} \\ (n_g - N_h*) \frac{N_h}{(N_g - N_h*)} & \text{si } h \text{ n'est pas exhaustive par défaut} \end{cases},$$

avec N_h* la somme des établissements dans les strates h exhaustives par défaut de g.

Dans notre application, on note que, dans un g donné, les strates sont toutes exhaustives ou toute non exhaustives par défaut. Il n'y a pas de mélange entre strates exhaustives et non exhaustives. On peut donc utiliser la formule simplifiée :

$$n_h = \begin{cases} N_h & \text{si } h \text{ est exhaustive par défaut} \\ n_g \frac{N_h}{N_g} & \text{si } h \text{ n'est pas exhaustive par défaut} \end{cases},$$

avec N_g la somme des établissements dans g.

Pour plus de détails sur l'application, voir l'Annexe D.

6.3 Allocation de base dans les strates

Soit n_{h1}, le résultat de l'allocation dans la strate de tirage h pour les objectifs au niveau NOGA2reg (nallotNog), n_{h2} le résultat de la même démarche pour les objectifs au niveau grande région x secteur (nallotGRSEC), et n_{h3} le résultat pour les régions avec échantillon complémentaire x secteur (nallotCANTSEC).

On définit l'allocation n_h comme suit (allotMax) :

$$n_h = \max(n_{h1}, n_{h2}, n_{h3}),$$

et la valeur arrondie à l'entier supérieur (allotMaxCeil). Elle permet d'atteindre tous les objectifs de l'enquête en cas de réponse complète.

L'allocation de base (allotMaxCeil) correspond à un ensemble de 41'130 établissements avec de 1 à 1015 établissements par strate ; voir Table 6. Les objectifs au niveau NOGA2reg sont clairement dominants dans cette allocation. Sans arrondis et valeurs inférieures limites, il faut en effet déjà 40'365 établissements pour atteindre ces objectifs. L'apport des régions croisées avec les secteurs est alors seulement de 126 établissements. Environ 700 établissements sont ajoutés lorsque les tailles allouées sont arrondies.

TAB. 6 Tailles allouées dans les Nbstra strates de tirage. Minimum Min, maximum Max et somme Sum pour les différents groupes d'objectifs (nallotNog, nallotGRSEC, nallotCANTSEC), ainsi que combinaison de ces derniers (allotMax, allotMaxCeil).

Allocation	Nbstra	Min	Max	Sum
nallotNog	2827	0.01	1015	40365.0
nallotGRSEC	2827	0.00	1015	36700.2
nallotCANTSEC	1254	0.00	672	13875.7
allotMax	2827	0.01	1015	40490.8
allotMaxCeil	2827	1.00	1015	41130.0

6.4 Allocation finale

L'allocation finale est déterminée de façon à éviter des problèmes lors des estimations et à respecter le maximum de 63'000 établissements dans l'enquête.

Diverses étapes sont appliquées successivement sur l'allocation de base (allotMaxCeil) afin d'obtenir une allocation finale. Elles prennent en compte des contraintes tel qu'un nombre minimal d'observations par strate et un poids d'échantillonnage pas trop important. Ces étapes intègrent également une augmentation afin de prévenir les effets de la non-réponse. Les détails de l'application sont disponibles dans l'Annexe D.

1. Dans les region_new avec objectifs, prise de tous les éléments des strates avec moins de 20 éléments et au moins 20 éléments dans les strates plus grandes. Dans les region_new

sans échantillons complémentaires, prise de tous les éléments des strates avec moins de 10 éléments et au moins 10 éléments dans les strates plus grandes. On augmente également la taille de l'échantillon dans les strates avec un taux de sondage inférieur à 1/40, *i.e.* poids d'échantillonnage limité à 40. Le résultat est nommé `allo1020`. Les strates contiennent entre 1 et 1015 établissements pour un total de 55'471 établissements. 256 strates deviennent exhaustives en plus des 1664 strates de base. L'augmentation de l'échantillon est importante ; voir discussion dans la Section 8.

2. Augmentation des tailles pour tenir compte d'un taux de réponse estimé de 90% dans les multi-établissements, 100% dans le profiling REE, 85% pour les normaux des plus grandes tailles et 80% pour les normaux des tailles plus petites. Ces valeurs sont déterminées sur la base de l'enquête en cours. Le résultat est nommé `allo1020C`. Les strates contiennent entre 1 et 1015 établissements pour un total de 59'560 établissements. L'augmentation n'est pas excessive, vu le bon taux de réponse attendu à l'enquête.

3. Plusieurs petites strates ont un taux de sondage supérieur ou égal à 50%. Elles sont prises exhaustivement. Le résultat est nommé `allo1020C50`, avec un total de 60'865 établissements. L'augmentation est donc peu importante.

4. Les strates avec un taux de sondage inférieur à 50% mais dont le taux passe au dessus de 60% lorsque les spéciaux sont répartis dans les classes de taille des normaux sont prises exhaustivement. Le résultat est nommé `alloNew60`, avec 61'800 établissements et seulement 743 strates non exhaustives.

5. Une adaptation est faite finalement dans certaines `NOGA2reg` afin d'avoir si possible 50 éléments dans les strates avec beaucoup de variabilité (strates avec plus de 5% de la variabilité de l'ensemble des strates non exhaustives). A nouveau, on remplit dès que le taux passe 50%. Le résultat est noté `alloFinC`, avec 62'833 établissements dans 737 strates non exhaustives, 567 strates exhaustives de spéciaux et 1523 strates exhaustives de normaux.

7 Tirage

Le tirage de l'échantillon utilise des nombres aléatoires permanents (NAP) et un algorithme de taille fixe.

Les NAP permettent de faire un lien entre tirages successifs. Cette méthode a déjà été utilisée pour faire le lien entre les révisions 2000 et 2003. Elle est reprise pour la révision 2007. Ainsi, les établissements existants lors du tirage de 2003 reprennent leur NAP de l'époque et les nouveaux établissements reçoivent un NAP selon un algorithme aléatoire uniforme entre 0 et 1.

Les établissements sont alors triés en ordre croissant selon leur NAP dans chaque strate h de taille N_h et les n_h premiers éléments sont sélectionnés dans l'échantillon. La taille de l'échantillon est donc fixe - n_h dans h - et le poids d'échantillonnage de $k \in h$ est donné par $d_k = N_h/n_h$.

Le système des NAP permet de faire un tirage selon une procédure peu complexe tout en assurant un recouvrement entre échantillons successifs. Le cadre de sondage évoluant entre les tirages, le recouvrement n'est pas totalement contrôlé et on ne peut pas assurer que tel ou tel établissement se retrouvera dans le nouvel échantillon. Une bonne coordination se réalise cependant naturellement si les listes d'établissements et les plans d'échantillonnage sont

proches. Le recouvrement est parfait dans une strate si le contenu et le taux de sondage sont fixes. Il est de moins en moins grand plus le contenu varie.

Les NAP sont souvent utilisés avec un tirage de Bernoulli. Cette procédure permet de bien contrôler des renouvellements avec rotation mais a le défaut de produire des échantillons avec une taille aléatoire. Dans notre cas, nous privilégions un tirage à taille fixe car il n'est pas prévu de faire des rotations d'échantillon.

L'échantillon final contient 62'833 établissements, dont 41'553 normaux et 21'280 spéciaux ; voir Table 7. Environ 30% des établissements de l'échantillon sont dans le profiling et 66% sont des établissements normaux. Parmi tous les emplois relevés dans l'échantillon, environ 25% sont dans le profiling et 68% dans les normaux.

TAB. 7 Information sur la population et l'échantillon STATEM en fonction du type d'établissement (Type ou Sflgprof). Nombre d'unités et d'emplois selon le cadre Npop et betot avec les répartitions %Npop et %betot, ainsi que nombre d'unités et d'emplois selon le cadre dans l'échantillon sélectionné nech et betotE avec les répartitions %nech et %betotE.

Type	Sflgprof	Npop	%Npop	betot	%betot	nech	%nech	betotE	%betotE
Normal	0	363'818	94.5	3'011'142	81.0	41'553	66.1	1'491'184	67.8
Prof. privé	1	11'370	3.0	299'026	8.0	11'370	18.1	299'026	13.6
Prof. public	2	7'323	1.9	253'609	6.8	7'323	11.7	253'609	11.5
Permanent	3	317	0.1	23'876	0.6	317	0.5	23'876	1.1
Multi	4	2'270	0.6	130'464	3.5	2'270	3.6	130'464	5.9
Total		385'098	100.0	3'718'117	100.0	62'833	100.0	2'198'159	100.0

On observe une structure différente entre les régions. La proportion d'établissements profiling varie par exemple entre 25% (Ostschweiz et Zentralschweiz) et 35% (Tessin) selon les grandes régions et entre 12% (NE) et 37% (ZH ville) dans les régions avec échantillons complémentaires ; voir Table 8.

Le nouvel échantillon est prévu pour les enquêtes du 2ème trimestre 2007 au 2ème trimestre 2008. On observe que 42'091 établissements sélectionnés faisaient partie de l'échantillon de l'enquête du 1er trimestre 2007. Il y a 20'742 nouveaux établissements, dont 19'494 normaux (à contacter), 3 parties de multi-établissements et 1'245 dans le profiling. Un groupe de 25'611 établissements sortent de l'échantillon[5]. Environ 2/3 des établissements de l'échantillon du 1er trimestre 2007 se retrouvent dans le nouvel échantillon. Le recouvrement sera plus important lors du renouvellement annuel mis en place dans la révision 2007 (env. 85% entre les deux échantillons du 2ème trimestre 2008). Les cadres et les plans seront en effet beaucoup plus proches.

[5]Note : La comparaison des listes et réalisée à partir des identificateurs d'établissements sans tenir compte des mutations.

TAB. 8 Distribution de l'échantillon par type dans les grandes régions et régions avec échantillons complémentaires. Rappel des codes : `SflgProf`=0 : normaux, `SflgProf`=1 : profiling privé, `SflgProf`=2 : profiling public, `SflgProf`=3 : permanents STATEM et `SflgProf`=4 : parties de multi-établissements.

		SflgProf					
		0	1	2	3	4	All
gr							
1	Région lémanique	8719	1918	1873	70	479	13059
2	Espace Mittelland	9304	2850	1799	73	359	14385
3	Nordwestschweiz	4617	1505	728	35	435	7320
4	Zurich	7148	1827	1569	65	260	10869
5	Ostschweiz	6305	1744	536	39	452	9076
6	Zentralschweiz	3354	949	236	21	237	4797
7	Tessin	2106	577	582	14	48	3327
All		41553	11370	7323	317	2270	62833

		SflgProf					
		0	1	2	3	4	All
region_new	autokz						
11	GE	3273	435	477	22	318	4525
12	VD	3834	996	1165	42	95	6132
21	FR	2239	381	731	7	47	3405
22	NE	2046	264	16	15	93	2434
41	ZH (ville)	3135	685	1214	22	107	5163
51	SG	3407	666	290	20	287	4670
All		17934	3427	3893	128	947	26329

8 Un peu plus sur le plan d'échantillonnage

Le plan d'échantillonnage développé pour la révision 2007 de la STATEM est basé sur l'expérience des plans précédents, avec une réévaluation des différentes étapes. Dans cette section, on présente quelques réflexions sur les choix retenus et des alternatives pour le futur.

8.1 Stratification et allocation

Dans la construction du plan d'échantillonnage, on note que l'allocation de base (`allotMaxCeil`= 41'130) augmente fortement lorsque l'on ajoute un critère de taille minimale par strate et une limite de poids pour prévenir des problèmes d'estimation (`allo1020`=55'471). Cette augmentation provient principalement de la taille minimale, avec ajout de l'ordre de 10'500 établissements, mais est également liée au poids maximal pour 3'800 établissements. L'augmentation provenant de la taille minimale est clairement liée au nombre de strates définies pour le tirage. Plus il y a de petites strates, plus on va augmenter l'échantillon. L'exigence de contrôle des tailles pour tous les objectifs de l'enquête (`NOGA2reg`, grandes régions et `region_new` croisées avec le secteur) a pour conséquence une stratification très détaillée avec beaucoup de petites strates. L'augmentation est donc importante.

Deux approches sont envisageables pour éviter la forte augmentation observée dans la STATEM : définir des strates moins détaillées et donc globalement plus grandes (ex. `NOGA2reg` x classe de taille) ou conserver la stratification de base sans ajouter l'exigence de taille minimale ou de poids maximal.

Dans l'approche des strates moins détaillées, nous perdons le contrôle sur la taille de l'échantillon et la précision attendue au niveau de certains objectifs. Si la variable `region_new` n'est par exemple pas considérée dans la stratification, une estimation à ce niveau devrait utiliser une estimation par domaine sur la base d'une taille aléatoire d'échantillon. Cette approche n'a pas été poursuivie dans le cas de la révision de la STATEM car les tailles exactes étaient désirées à tous les niveaux. Elle pourrait cependant être réévaluée dans le futur en combinaison avec une allocation qui permette d'atteindre un sous-échantillon de répondants jugé raisonnable - taille et structure - pour les objectifs non contrôlés. Une précision désirée, et non plus une taille d'échantillon brut assurée, pourrait donc être prévue pour les résultats régionaux.

Dans l'approche des strates détaillées mais sans critère de taille minimale, nous perdons en stabilité pour toutes les opérations basées sur les strates. Le nombre de répondants par strate serait en effet souvent petit ou même nul. Pour des estimations de qualité, les estimateurs classiques basés sur les strates devraient donc être remplacés par des estimateurs plus complexes basés sur des cellules ou marges différentes des strates. De tels estimateurs sont disponibles pour les estimations de totaux par exemple ; voir l'approche calage de Särndal (2007) ou Särndal et Lundström (2007). Les estimations de variances sont cependant plus difficiles.

Dans l'état de nos connaissances, les deux approches offrent des potentiels de réduction de la taille de l'échantillon. Elles demandent cependant de revoir les contraintes de l'enquête (par ex. regroupements de NOGA2, taille variable dans les cantons ou même les grandes régions). Nous noterons encore que les regroupements de classes de taille ne sont pas recommandés car la taille est fortement corrélée aux variables d'intérêt de la STATEM.

Dans la révision 2007, les estimateurs des totaux de la STATEM ont été revus de façon à s'éloigner des problèmes liés aux multiples petites strates. La détermination des poids finaux utilise l'approche calage de Särndal (2007) avec des cellules correspondant à des agrégats de strates. Seul le poids d'échantillonnage est déterminé au niveau strate. Les strates ne sont

pas utilisées comme cellules de correction de non-réponse ou de calage sur des valeurs auxiliaires de référence. Par contre, les estimations de variance utilisent encore le niveau strate. Le choix de taille minimale appliqué dans la révision 2007 est donc directement bénéfique aux estimations de variance.

8.2 Échantillons complémentaires

Lors de la révision de l'échantillon STATEM de 2003, l'échantillonnage avait été effectué en deux étapes : plan et tirage national puis plan et tirage pour les compléments cantonaux et la ville de Zurich. La méthode était relativement complexe, notamment dans les plans et tirages successifs, la détermination des poids de sondage et les estimations basées sur les strates (quelles strates utiliser ?). Elle avait par contre l'avantage de donner explicitement les tailles des compléments régionaux.

Dans la révision 2007, nous avons un système général qui intègre tous les objectifs. Le tirage, la détermination des poids de sondage et les estimations basées sur les strates sont plus simples. Par contre, nous n'avons pas d'information directe de l'effet des compléments régionaux (GE, VD, FR, NE, ZH ville) sur la taille totale de l'échantillon.

Pour avoir une information sur l'influence des objectifs régionaux, nous comparons le plan complet décrit dans les Sections 5 et 6 avec un plan d'échantillonnage "Suisse" qui considère les objectifs nationaux uniquement, c'est-à-dire `NOGA2reg` et grandes régions x secteur.

La stratification et l'allocation du plan "Suisse" reprennent les méthodes appliquées dans le plan complet tout en tenant compte de la réduction des objectifs.
- Les strates de tirage sont définies par les croisements suivants : type d'établissements (`special`) x activité économique (`NOGA2reg`) x grandes régions (`gr`) x classes de taille (`tai2`). Elles comportent donc en général plus d'éléments que celles du plan complet.
- Deux variantes sont choisies pour la taille minimale par strate : 10 et 20 établissements.
- Le poids d'échantillonnage maximal est fixé à 40.
- Les tailles allouées sont augmentées pour tenir compte de la non-réponse espérée selon les mêmes taux que le plan complet.
- Les strates avec un taux de sondage $\geq$50% sont prises exhaustivement.
- Les corrections complémentaires de l'allocation ne sont pas intégrées dans la comparaison du plan complet avec le plan "Suisse".

L'allocation de base du plan "Suisse" est très proche de celle du plan complet ; 40'701 resp. 41'130 établissements ; voir `allotMaxCeil` dans la Table 9. Par contre, on observe une vraie différence au moment de l'application de la taille minimale et du poids maximal (`allot1020`, resp. `allo10` et `allo20`). Elle est clairement due à la différence dans le nombre de petites strates. Le plan complet comporte 2827 strates, dont 1163 strates non exhaustives par défaut, et le plan simplifié comporte 1655 strates dont 630 non exhaustives par défaut. La différence reste ensuite pour l'allocation corrigée pour la non-réponse (`"C"`). L'effet de remplissage des strates avec taux de sondage $\geq$50% augmente encore la taille à 60'865 dans le plan complet. Il est petit, voire négligeable, dans le plan simplifié (`"C50"`).

L'effet global de l'intégration des régions `regions_new` est de l'ordre de 3'500 à 9'000 établissements pour une limite de 20, respectivement 10. Une fois encore, on voit l'effet important du choix de la taille minimale abordé dans la Section 8.1. Ces valeurs sont cependant des agrégats qui combinent plusieurs effets.

La répartition de l'allocation "Suisse" dans les grandes régions et `region_new` (proportionnel à

TAB. 9 Comparaison des plans complet et "Suisse". Allocation de base pour les différents objectifs : `nallotNog`, `nallotGRSEC`, `nallotCANTSEC`, `allotMax` et `allotMaxCeil`. Allocation avec limite par strate et maximum de poids : `allot1020`, `allo10` et `allo20`. Allocation avec augmentation pour non-réponse : même noms avec ajout de `"C"`. Allocation avec remplissage si taux $\geq$50% : même nom avec `"C50"`.

	Complet	"Suisse"	Différence
`nallotNog`	40365	40365	
`nallotGRSEC`	36700	36700	
`nallotCANTSEC`	13876		
`allotMax`	40491	40383	108
`allotMaxCeil`	41130	40701	429
`allot1020`	55471		
`allo10`		48597	6874
`allo20`		52699	2772
`allot1020C`	59560		
`allo10C`		51418	8142
`allo20C`		56571	2989
`allot1020C50`	60865		
`allo10C50`		51722	9143
`allo20C50`		57464	3401

la taille) permet d'étudier les différences au niveau régional, voir Table 10.

Les grandes régions 3, 6 et 7 n'ont pas de compléments régionaux et la limite par strate est fixée à 10 dans le plan complet. Le plan "Suisse" arrive donc à la même allocation lorsque la limite est également 10 (`alloKt10`). Ainsi, la différence entre plan complet et plan "Suisse" se trouve entre 6'076 (limite=20) et 9'143 (limite=10) établissements si l'on se restreint aux grandes régions touchées par les compléments (Total sans 3, 6, 7). Ces valeurs cachent cependant encore des différences de traitements au niveau des `region_new`.

Dans la deuxième partie de la Table 10, on se restreint aux `region_new` dans les 4 grandes régions touchées par les compléments. On observe alors une différence de 7'830 établissements entre le plan complet et le plan "Suisse" avec limite 10 pour les régions sans complément et limite 20 avec complément (valeurs soulignées dans la table)[6]. L'intégration des objectifs régionaux dans la plan correspond donc à une augmentation de env. 7'800 établissements dans l'échantillon.

Il est intéressant de noter que le plan complet alloue plus d'établissements que le plan "Suisse" dans toutes les régions et non seulement dans les régions avec complément. Par exemple, le fait que Genève et Vaud demandent un complément provoque une augmentation aussi dans le canton du Valais. Les demandes de GE et VD provoquent donc une augmentation de 2'519 établissements dans la grande région. Si cette valeur est répartie dans les deux cantons, on obtient de augmentation de 1200-1300 établissements dans les deux cantons, et non pas 900-1000. Cela est dû principalement à la stratification mise en place pour satisfaire les objectifs régionaux ; voir Section 8.1.

[6] Rappel : Le plan complet fixe la limite 10 dans les régions sans compléments et 20 dans les régions avec complément.

TAB. 10 Comparaison des allocations complètes et "Suisse". Nombre d'établissements Npop dans la région, allocation complète `alloComp`, allocation "Suisse" avec limite=10 `alloKt10` et limite=20 `alloKt20`, ainsi que différences : `Diff10=alloComp - alloKt10` et `Diff20=alloComp - alloKt20`. Le total combiné reprend les différences avec limite 10 dans les régions sans complément et 20 dans les régions avec complément.

Grandes régions

gr		Npop	alloComp	alloKt10	alloKt20	Diff10	Diff20
1	Région lémanique	69722	12720	9613	10391	3107	2329
2	Espace Mitteland	79591	13882	10952	11687	2930	2195
3	Nordwest CH	50819	6898	6898	7796	0	-898
4	Zurich	69564	10510	9236	9970	1274	540
5	Ostschweiz	56638	8907	7075	7895	1832	1012
6	Zentralschweiz	39083	4653	4653	5519	0	-866
7	Tessin	19681	3295	3295	4206	0	-911
	Total	385'098	60'865	51'722	57'464	9143	3401
	Total sans 3,6,7	275'515	46'019	36'876	39'943	9143	6076

Régions dans grandes régions avec compléments

region_new		Npop	alloComp	alloKt10	alloKt20	Diff10	Diff20
10	VS	15'260	2'361	1'799	1'989	562	372
11	**GE**	22'978	4'402	3'152	3'362	1'250	1'040
12	**VD**	31'484	5'957	4'662	5'040	1'295	917
20	BE,SO,JU	60'182	8'114	7'941	8'491	173	-377
21	**FR**	11'060	3'334	1'940	2'032	1'394	1'302
22	**NE**	8'349	2'434	1'071	1'165	1'363	1'269
40	ZH sans ville	44'089	5'386	5'197	5'746	189	-360
41	**Zurich ville**	25'475	5'124	4'039	4'224	1'085	900
50	GL,SH,AI,AR,GR,TG	33'185	4'293	3'947	4'412	346	-119
51	**SG**	23'453	4'614	3'128	3'483	1'486	1'131
	Total	275'515	46'019	36'876	39'943	9'143	6'076
	Total combiné					7'830	

8.3 Précision des estimations

L'échantillon de base de `allotMaxCeil`=41'130 établissements du plan complet est construit de façon à satisfaire les objectifs de l'enquête en cas de réponse complète de tous les établissements. L'échantillon final de `allotFinC`=62'833 établissements prend en compte un ajustement pour contrer l'effet de la non-réponse dans les strates non-exhaustives. Il comporte de plus des augmentations importantes dans les petites strates et des augmentations complémentaires telles que remplissage dès que le taux de sondage est supérieur ou égal à 50%.

Si les taux de réponse utilisés dans le plan sont effectivement observés, l'échantillon des répondants contiendra `alloFinCNR`=54'088 établissements. Ce nombre est bien supérieur à `allotMaxCeil` mais la structure est différente. En effet, `alloFinCNR` prend en compte de la non-réponse dans toutes les strates hors profiling. L'échantillon est donc réduit dans les strates exhaustives, notamment celles des grands établissements connues pour être souvent très variables.

Pour tester l'allocation et ses limites, on détermine les CV théoriques des totaux des variables

equivpt et betot disponibles dans le cadre comme suit :

$$CV(\widehat{Y}) = \frac{\sqrt{\sum_h N_h^2 (1 - \frac{m_h}{N_h}) \frac{S_h^2}{m_h}}}{Y}$$

avec N_h la taille de la strate h, m_h le nombre de répondants dans h, Y le total de y selon le cadre et S_h^2 la variance de y dans la strate h selon le cadre. La formule est appliquée pour les différents domaines objectifs de l'enquête avec alloFinCNR=54'088 répondants.

Les CV sont très proches pour les deux variables equivpt et betot et ils satisfont la plupart des objectifs de l'enquête. Le CV national de 0.35% est bien au dessous de l'objectif de 0.5%. Les objectifs de 2% par grandes régions et region_new, ainsi que 3% par grandes régions x secteurs sont tous atteints. Par contre, le CV par region_new x secteurs est inférieur à l'objectif de 3% dans le secteur secondaire de Genève (4.6%).

Les CV par NOGA2reg ne sont pas tous atteints, voir Table 11. Des problèmes apparaissent clairement dans plusieurs NOGA2reg exhaustives (exnoga=1) pour cause de non-réponse. Certaines NOGA2reg de taille non négligeable ont également des CV supérieurs à 2%. Comme exemple, la classe NOGA2reg=66 a 2'210 établissements, dont 1'991 dans l'échantillon brut, 1'771 prévus dans le net, et un CV de l'ordre de 4% pour le total des equivpt et des betot. Ainsi, même en reprenant equivpt utilisé dans le plan et en étant prudent dans l'allocation, les objectifs ne sont pas toujours respecté au niveau de certains domaines d'intérêt, notamment en cas de non-réponse parmi les grands établissements.

Deux éléments nous laissent cependant espérer une amélioration au moins partielle des CV dans les résultats finaux. Premièrement, nous prendrons en compte la variable auxiliaire equivpt dans l'estimateur des totaux de l'emploi. Ainsi, l'estimateur de calage aura une variance estimée en principe inférieure à celle estimée avec la méthode simple du plan stratifié. Deuxièmement, des imputations sont prévues pour éviter une partie des effets dévastateurs de la non-réponse de très grands établissements.

Les plus grands établissements de chaque NOGA2reg ont été identifiés par le code IDVB=1 ("very big") et des efforts particuliers seront déployés durant l'enquête pour obtenir les données de ces établissements afin d'éviter un maximum de non-réponses. De plus, des imputations sont prévues pour parer à certaines non-réponses restantes. On peut donc espérer avoir un taux de réponse supérieur aux 85% utilisés dans cette simulation. Notons finalement que le profiling peut également avoir des valeurs manquantes. Dans ce cas, des imputations sont mises en place et il n'y pas d'ajout à la variance estimée.

TAB. 11 Estimation des CV avec l'échantillon final de la STATEM au niveau des NOGA2reg. Nombre d'établissements dans la population N, échantillon brut alloFinC, échantillon net espéré alloFinCNR, et CV de equivpt et betot repris du cadre.

Obs	noga2reg	exnoga	N	allo FinC	allo FinCNR	CVEPTFin	CVbet Fin
1	15	0	2592	900	733	1.90	1.84
2	16	1	15	15	12	4.81	5.16
3	17	0	623	458	353	2.04	1.98
4	18	0	804	400	310	3.47	3.38
5	19	1	208	208	166	1.87	2.25
6	20	0	6451	886	705	1.94	1.98
7	21	1	223	223	174	4.44	4.31
8	22	0	4128	877	700	2.03	2.11
9	23	1	14	14	12	0.36	0.16
10	24	0	993	689	541	4.95	4.89
11	25	0	855	567	439	2.00	1.97
12	26	0	1451	614	483	2.42	2.40
13	27	1	281	281	221	3.93	3.93
14	28	0	7804	1011	804	1.67	1.69
15	29	0	3750	1045	835	2.43	2.39
16	30	1	136	136	105	3.79	3.59
17	31	0	1088	630	492	2.95	2.93
18	32	0	690	494	381	5.70	5.56
19	33	0	3448	1040	825	2.04	2.01
20	34	1	187	187	149	11.14	10.91
21	35	1	451	451	370	12.94	12.77
22	45	0	37425	2469	2046	1.11	1.11
23	50	0	15702	1280	1042	1.60	1.69
24	51	0	21977	1703	1398	1.44	1.47
25	52	0	48599	4674	4191	0.99	1.08
26	55	0	29066	1797	1495	1.33	1.61
27	60	0	8518	2063	1842	1.30	1.40
28	61	1	118	118	90	10.61	9.79
29	62	1	229	229	179	47.37	48.47
30	63	0	4278	922	746	2.37	2.47
31	64	0	5564	5564	5407	1.28	1.68
32	65	0	4445	2185	1969	1.24	1.23
33	66	0	2210	1991	1771	4.04	3.90
34	67	0	5028	1246	1091	1.84	2.32
35	70	0	5364	852	690	1.74	2.12
36	71	0	950	598	469	2.39	2.57
37	72	0	11881	1147	933	1.74	1.77
38	73	1	663	663	552	4.33	4.09
39	74	0	62349	3792	3132	1.05	1.15
40	75	0	8480	4586	4339	0.64	0.70
41	80	0	15237	4449	4126	1.57	1.61
42	85	0	26397	3362	2933	1.25	1.27
43	90	0	2012	973	780	1.69	1.68
44	91	0	5782	918	731	1.62	1.86
45	92	0	7176	1206	1007	1.53	1.90
46	93	0	14300	956	764	2.05	2.37
47	1014	1	393	393	316	1.89	1.87
48	3637	0	3824	920	731	1.86	1.90
49	4041	0	939	651	508	6.22	6.11

9 Conclusion

Le nouveau plan d'échantillonnage de la STATEM repose sur une approche globale des objectifs de l'enquête. Il redéfinit les strates et la méthode d'allocation. Ce nouveau plan intègre également les cas spéciaux tels que le profiling du REE, les multi-établissements et les permanents STATEM.

Le plan d'échantillonnage est intégré dans un modèle général qui propose un cadre et un échantillon fixes pendant une année, c'est-à-dire pendant 4 trimestres d'enquête. Les établissements sont suivis durant une année afin d'intégrer les mutations telles les disparitions ou les fusions par exemple. La liste de base n'est par contre pas modifiée. Notons que la présence du profiling a créé quelques difficultés dans la méthode. Un traitement spécial a en effet dû être mis en place car le suivi dans le temps n'est actuellement pas possible de la même manière que pour les établissements normaux.

L'échantillon de 62'833 établissements, dont 41'553 normaux et 21'280 spéciaux, est construit de façon à atteindre tous les objectifs de l'enquête. L'échantillon et le cadre documentés dans ce rapport sont utilisés pour les enquêtes entre le 2ème trimestre 2007 et le 2ème trimestre 2008. Il sera renouvelé pour le 2ème trimestre 2008 avec un recouvrement maximal.

La construction du plan d'échantillonnage a permis de montrer l'influence directe de la stratification sur la taille de l'échantillon ; principalement pour raison de taille minimale par strate. Pour le futur, il serait intéressant de revoir les objectifs, notamment au niveau des régions et des activités économiques, de façon a tenter de réduire l'échantillon tout en gardant des objectifs de précision acceptables. La mise en place de la nouvelle classification NOGA08 est par exemple une bonne occasion de revoir les regroupements d'activités économiques. De même, des estimations par domaines pourraient être évalués pour le niveau des régions.

A Définitions

Activités économiques NOGA2 :

NOGA2	
10	Extraction de houille, lignite et tourbe
11	Extraction d'hydrocarbures et services annexes
12	Extraction de minerais d'uranium et thorium
13	Extraction de minerais métalliques
14	Extraction de la pierre et de la terre, autres industries extractives
15	Industries alimentaires et industrie des boissons
16	Industrie du tabac
17	Industrie textile
18	Industrie de l'habillement
19	Industrie du cuir et d'articles en cuir
20	Travail bois et fabrication articles en bois, liège, vannerie et sparterie (sans fabrication meubles)
21	Industrie du papier, du carton et de leurs dérivés
22	Edition, impression, reproduction d'enregistrements sonores, vidéo et informatiques
23	Cokéfaction ; raffinage ; traitement nucléaire
24	Industrie chimique
25	Fabr. de caoutchouc et matières plastiques
26	Fabr. d'autres prod. minéraux non métalliques
27	Métallurgie
28	Travail des métaux
29	Fabrication de machines et d'équipements
30	Fabr. de machines de bureau, ordinateurs
31	Fabr. de machines et appareils électriques
32	Fabr. d'équip. de radio, TV, communication
33	Instruments médicaux/précision, horlogerie
34	Industrie automobile
35	Fabrication d'autres moyens de transport
36	Fabrication meubles, bijoux, instruments musique, articles sport, jeux, jouets et autres produits
37	Récupération
40	Production et distribution d'électricité, de combustibles gazeux et de chaleur
41	Captage, traitement et distribution d'eau
45	Construction
50	Commerce et réparation de véhicules
51	Commerce de gros, intermédiaires du com.
52	Commerce de détail, rép. articles domestiques
55	Hôtellerie et restauration
60	Transports terrestres et par conduites
61	Transports par eau
62	Transports aériens
63	Auxiliaires des transp., agences de voyage
64	Postes et télécommunications
65	Intermédiation financière
66	Assurances (sans sécurité sociale obligatoire)
67	Activités auxiliaires liées à l'intermédiation financière et aux assurances
70	Activités immobilières
71	Location de machines et équipements
72	Activités informatiques
73	Recherche et développement
74	Autres services fournis aux entreprises
75	Adminsitration publique, défense ; sécurité sociale obligatoire
80	Enseignement
85	Santé et activités sociales
90	Assainissement, voirie et gestion des déchets
91	Activités associatives
92	Activités récréatives, culturelles et sportives
93	Services personnels

Secteurs économiques :

```
if 10 <= noga2 <= 45  then secteur=2;
if 50 <= noga2 <= 93  then secteur=3;
```

Sections économiques :

```
if 10 <= noga2 <= 41  then section='CDE';
if       noga2  = 45  then section='F';
if 50 <= noga2 <= 52  then section='G';
if       noga2  = 55  then section='H';
if 60 <= noga2 <= 64  then section='I';
if 65 <= noga2 <= 67  then section='J';
if 70 <= noga2 <= 74  then section='K';
if       noga2  = 75  then section='L';
if       noga2  = 80  then section='M';
if       noga2  = 85  then section='N';
if 90 <= noga2 <= 93  then section='O';
```

Regroupement de classes NOGA2 pour la STATEM :

```
noga2reg=noga2;
if 10<= noga2 <= 14 then noga2reg='10.14';
if 36<= noga2 <= 37 then noga2reg='36.37';
if 40<= noga2 <= 41 then noga2reg='40.41';
```

Note : dans le cadre utilisé pour le tirage la codification était '1014', '3637', et '4041'.

Construction des équivalents plein temps :

```
* valable du 3ème trimestre 2006 au 2ème trimestre 2007;
if section='CDE' then do; coeffTZ1=0.6231136; coeffTZ2=0.3083883; end;
if section='F' then do; coeffTZ1=0.6104126; coeffTZ2=0.2954923; end;
if section='G' then do; coeffTZ1=0.6173345; coeffTZ2=0.2906290; end;
if section='H' then do; coeffTZ1=0.5922656; coeffTZ2=0.2973226; end;
if section='I' then do; coeffTZ1=0.6148650; coeffTZ2=0.3097486; end;
if section='J' then do; coeffTZ1=0.6020523; coeffTZ2=0.2976034; end;
if section='K' then do; coeffTZ1=0.6019188; coeffTZ2=0.2815877; end;
if section='L' then do; coeffTZ1=0.6470481; coeffTZ2=0.2930233; end;
if section='M' then do; coeffTZ1=0.6777581; coeffTZ2=0.3124356; end;
if section='N' then do; coeffTZ1=0.6518832; coeffTZ2=0.3305299; end;
if section='O' then do; coeffTZ1=0.6329756; coeffTZ2=0.2922567;end;
* vollz: temps plein, teilz1: temps partiel 1, teilz2: temps partiel 2;
equivpt = vollz + coeffTZ1*teilz1 + coeffTZ2*teilz2;
```

Grandes régions :

```
if autokz in ('VD','VS','GE')                       then gr =1;
if autokz in ('BE','FR','SO','NE','JU')             then gr =2;
if autokz in ('BS','BL','AG')                       then gr =3;
if autokz in ('ZH')                                 then gr =4;
if autokz in ('GL','SH','AI','AR','SG','GR','TG')   then gr =5;
if autokz in ('LU','UR','SZ','OW','NW','ZG')        then gr =6;
if autokz in ('TI')                                 then gr =7;
```

Régions avec échantillons complémentaires STATEM et restes des grandes régions :

```
region_new=10*gr;
if autokz = 'GE'                 then region_new =11;
if autokz = 'VD'                 then region_new =12;
if autokz = 'FR'                 then region_new =21;
if autokz = 'NE'                 then region_new =22;
if autokz = 'ZH' and gdenr = 261 then region_new =41;
if autokz = 'SG'                 then region_new =51;
```

Status établissement selon le REE :

status	UNIT_STATUS_CD
1	actif
2	inactif
3	radié
4	inconnu (betyp = L17 ou L18)
5	fictif
6	transféré

Code betyp selon le REE :

betyp	F_S	betyp	F_S
L00	Fictive, c/o adresse	L22	Public : District
L01	Agricole	L23	Public: Commune
L02	Unité agr. sans pers	L24	Autre Org. Corpor.
L05	Provisoire fictive	L25	Eglises
L12	Société de capitaux	L26	Pub: Etabl. seul. BZ
L13	Unité active privé	L27	Entreprise publique
L14	Unité sans personnel	L28	Public: Adr. s. pers
L15	Inactif, dépôt	L30	Etat étrangers
L17	Réactivation	L50	En liquidation
L18	Nouvelle entrée	L51	Inactiv
L19	Nouvelle, par fusion	L91	Radiation
L20	Public : confédérat.	L99	Radiation (double)
L21	Public: Canton		

Types d'établissements STATEM :

```
special       Définition
    0         normal
    1         multi-établissement
    2         profiling REE ou permanents STATEM

Sflgprof      Définition
    0         normal
    1         profiling REE privé
    2         profiling REE public
    3         permanents STATEM
    4         multi-établissement

stemppro      Définition
   "."        hors profiling BUR
    0         profiling BUR normal
    1         futur profiling BUR
```

B Limite Take-All

Description très rapide de la méthode. Référence : Hidiroglou (1986).

L'algorithme permet de déterminer la limite λ optimale entre la strate avec relevé exhaustif et les strates avec relevé non-exhaustif. Cette algorithme itératif détermine à chaque pas une valeur majorant la valeur optimale λ^* ; et cela de façon toujours plus précise (approximation par le haut).

Soit y la variable d'intérêt et les valeurs $y_{(1)}, .., y_{(i)}, .., y_{(N)}$ ordonnées de façon à ce que $y_{(i)} \leq y_{(i+1)}$, pour $i = 1, .., N$.

Note : y est remplacée par une variable auxiliaire x bien corrélée si elle n'est pas disponible.

Base :

$$\lambda^* \leq \mu_{[N]} + \sqrt{CV^2 \cdot Y^2/N + S^2_{[N]}} \tag{1}$$

avec $\mu_{[N]}$ la moyenne de y dans l'ensemble des N données, CV le coefficient de variation objectif pour le total Y, et $S^2_{[N]}$ la variance de y dans l'ensemble des N données.

Pas $j+1$ de l'itération :

Soit t_j le nombre d'éléments dans la strate take-all à l'itération j. On détermine la nouvelle limite comme suit :

$$\lambda^*_{j+1} = \mu_{[N-t_j]} + \sqrt{\frac{N-t_j-1}{(N-t_j)^2} CV^2 \cdot Y^2 + S^2_{[N-t_j]}} \tag{2}$$

avec :

$$\mu_{[N-t_j]} = \frac{1}{N-t_j} \sum_{i=1}^{N-t_j} y_{(i)} \tag{3}$$

$$S^2_{[N-t_j]} = \frac{1}{N-t_j-1} \sum_{i=1}^{N-t_j} (y_{(i)} - \mu_{[N-t_j]})^2 \tag{4}$$

Les t_{j+1} éléments avec $y_{(i)} \geq \lambda^*_{j+1}$ font partie de la strate des take-all.

L'itération est stoppée dés que l'asymétrie ou "skewness" des données sous la limite des take-all est petite (par exemple valeur 3).

C Méthode de Dalenius

Description très rapide de la méthode. Références : Särndal and al. (1992) ou Cochran (1977).

Soient y_0 et y_L les valeurs minimales et maximales de la variable y. Le problème du choix des limites entre les strates, lorsque le nombre de strates L est connu, est de trouver les limites entre les strates $y_1, ..., y_{L-1}$ telles que la variance du total soit minimale.

La méthode de Dalenius effectue les étapes suivantes :

1. partage de l'intervalle $[y_0, y_L]$ en $K >> L$ intervalles de longueurs constantes ;
2. calcul, pour chaque intervalle $k = 1, .., K$, du nombre d'occurrences $f_k(y)$ de y (histogramme), de la racine du nombre d'occurrence $\sqrt{f_k(y)}$, et du cumul $\Phi_k = \sum_{j=1}^{k} \sqrt{f_k(y)}$;
3. partage de l'intervalle $[\Phi_1 ; \Phi_K]$ en L intervalles égaux (nombre désiré de strates) ; les limites entre les intervalles sont notées $\Phi_{[j]}$, $j = 1, ..L-1$;
4. détermination des limites entre les strates y_j , $j = 1, ..., L-1$ correspondant aux valeurs $\Phi_{[j]}$, $j = 1, ..., L-1$.

La meilleure allocation dans les strates obtenues avec la méthode décrite ci-dessus est l'allocation de Neyman basée sur la variable d'intérêt y. Les valeurs de y étant inconnues, les valeurs d'une variables auxiliaire x, connue et bien corrélée avec y, les remplacent dans la procédure.

D Extraits des programmes SAS

Cette annexe propose des extraits des programmes SAS utilisés dans la construction du plan d'échantillonnage de la STATEM.

Allocation de base (extraits) :

```
*************************************;
* Restriction variables pour allocation;
*************************************;
data infoAlloc;
 set libFin07.cadreStraAll;
 keep betot burnr entid equivpt exnoga gr limTakeall noga2reg region_new
      secteur special straTir tai tai2 tako0
      teilz1 teilz2 tzf1 tzf2 tzm1 tzm2 vollz vzf vzm;
run;

*******************************************************;
* Macro allocation par var1 x var2                     ;
* 1) détermination nopt par var1 x var2                ;
* 2) allocation nalloh de nopt dans var1 x var2 x tai2  ;
* 3) distribution nallt de nalloh dans strates de tirage;
*******************************************************;
```

```
*----------------------------------------;
%macro alloc (dataIn=, cv=, var1=, var2=, dataOut=);
 * objectif CV;
 data _NULL; call symput("cv",&cv); run;
 * informations par strates niveau var1 x tai2;
  proc sort data=&dataIn;
   by &var1 &var2 tako0 special tai2;run;
  proc univariate data=&dataIn noprint;
   by &var1 &var2 tako0 special tai2;
   var equivpt;
   output out=info1 N=Nh sum=Xh std=Sh;
  run;
 * création variables pour allocation;
 data info11;
  set info1;
  NhI = Nh * tako0; /*Nh si take all et 0 sinon*/
  NhShI = Nh * Sh * (1-tako0); /*NhSh si hors take all et 0 sinon*/
  NhSh2I = Nh * Sh**2 * (1-tako0);/*idem pour NhSh^2*/
 run;
 * Détermination nopt et nallo par var1;
 proc univariate data=info11 noprint;
  var Nh NhI NhShI NhSh2I Xh;
  by &var1 &var2;
  output out=inter sum=Ntot Nstar NhShIsum NhSh2Isum X;
 run;
 data infoN;
  set inter;
  nopt=Nstar + NhShIsum**2/( (&cv * X)**2 + NhSh2Isum );
  nallo=nopt-Nstar;
  t=0; if nopt>Ntot then t=1;/*test*/
 run;
 title'Test taille dans var1';
 proc means data=infoN n min max sum maxdec=1; var t Ntot Nstar nopt nallo;run;
 title'';
 * Allocation dans les var1 x var2 x tai2;
 data infovar1;
  merge info11 infoN (keep=&var1 &var2 Nstar nallo NhShIsum);
  by &var1 &var2;
  if tako0=1 then nalloh=Nh;
  else nalloh = nallo * (Nh*Sh) / NhShIsum;
  t=0; if nalloh>Nh then t=1;/*test*/
 run;
 title'Test taille dans var1 x tai2';
 proc means data=infovar1 n min max sum; var nalloh t;run;
 title'';
 * Tailles des strates de tirage;
 proc sort data=&dataIn;
   by straTir noga2reg exnoga special tai2 region_new gr secteur tako0;run;
 proc univariate data=&dataIn noprint;
  by straTir noga2reg exnoga special tai2 region_new gr secteur tako0;
  var equivpt;
  output out=info2 N=Nt;
 run;
 * Répartition dans les strates de tirage;
 proc sort data=info2; by &var1 &var2 tako0 special tai2;run;
 proc sort data=infovar1; by &var1 &var2 tako0 special tai2;run;
```

```
 data &dataOut;
  merge info2 infovar1 (keep=&var1 &var2 tako0 special tai2 Nh nalloh);
  by &var1 &var2 tako0 special tai2;
  * répartition et taux;
  if tako0=1 then nallot=Nt;
  else nallot=nalloh*Nt/Nh;
  * poids sondage;
  wt=Nt/nallot;
 run;
 * test;
 proc means data=&dataOut n min max sum maxdec=1 fw=8;
  var nallot wt;
 run;
 proc means data=&dataOut n min max sum maxdec=1 fw=8;
  class tako0;
  var nallot wt;
 run;
%mend;
*---------------------------------------;

* Allocation optimale pour objectif NOGA2REG ;
* CV=2%                                     ;
*********************************************;
%alloc(dataIn=infoAlloc, cv=0.02,
       var1=noga2reg, var2=exnoga,
       dataOut=infoNogaTir);

* Allocation optimale pour objectif GR x secteur;
* CV=3%                                        ;
***********************************************;
%alloc(dataIn=infoAlloc, cv=0.03,
       var1=gr, var2=secteur,
       dataOut=infoGRSECTir);

* Allocation optimale pour objectif canton x secteur;
* CV=3%                                             ;
*****************************************************;
data ex;
 set infoAlloc;
 if region_new in (11,12,21,22,41,51);
run;
%alloc(dataIn=ex, cv=0.03,
       var1=region_new, var2=secteur,
       dataOut=infoCANTSECTir);

*******************************;
* Mise ensemble et détermination;
* allocation maximale, etc.     ;
*******************************;

proc sort data=infoNogaTir;
 by straTir gr region_new secteur noga2reg exnoga tako0 special tai2;run;
proc sort data=infoGRSECTir;
 by straTir gr region_new secteur noga2reg exnoga tako0 special tai2;run;
proc sort data=infoCANTSECTir;
 by straTir gr region_new secteur noga2reg exnoga tako0 special tai2;run;
```

```
data libFin07.comballo;
 merge infoNogaTir (in=a keep= Nt straTir gr region_new secteur
                               noga2reg exnoga tako0 special
                               tai2 nallot
                         rename=(nallot=nallotNog))
       infoGRSECTir(in=b keep= straTir gr region_new secteur
                               noga2reg exnoga tako0 special
                               tai2 nallot
                            rename=(nallot=nallotGRSEC))
       infoCANTSECTir(in=c keep= straTir gr region_new secteur
                                  noga2reg exnoga tako0 special
                                  tai2 nallot
                            rename=(nallot=nallotCANTSEC));
  by  straTir gr region_new secteur noga2reg exnoga tako0 special tai2;
  * cas spécial des cantons;
  if c=0 then nallotCANTSEC=0;
  * maximum et ceil;
   allotMax=max(nallotNog, nallotGRSEC, nallotCANTSEC);
   allotMaxCeil=ceil(allotMax);
run;
```

Correction final allocation (extraits) :

```
   * reprise allocation de base;
   allo1020=allotMaxCeil;
   * adaptation dans strates non exhaustives;
   if tako0=0 then do;
     if region_new in (11,12,21,22,41,51) then do;
       if Nt<20 then allo1020=Nt;
       else if Nt>=20 and allo1020<20 then allo1020=20;
      end;
     else do;
       if Nt<10 then allo1020=Nt;
       else if Nt>=10 and allo1020<10 then allo1020=10;
      end;
     if Nt/allo1020>40 then allo1020=ceil(Nt/40);
    end;

  * taux de réponse attendu;
 if special=1 then taux=0.9;/*multi*/
 else if special=2 then taux=1;/*profiling*/
 else if special=0 and tako0=1 then taux=0.85;
 else if special=0 and tako0=0 then taux=0.8;
 * augmentation ou c'est possible;
 allo1020C=max( min(floor(allo1020/taux),Nt) ,1);

 * exhaust si taux >=50%;
 taux1020C=allo1020C/Nt*100;
 allo1020C50=allo1020C;
 if taux1020C>=50 then allo1020C50=Nt;

 * exhaust si taux établissements total (normaux et spéciaux);
 * dans tai est >60%;
 * Requis: répartition des spéciaux dans les classes de taille;
 Ntot=Nt+nSpec; /*total normaux + total spéciaux*/
 AllocTot=allo1020C50 + nSpec;
```

```
ftot=AllocTot/Ntot*100;
fbase=allo1020C50/Nt*100;
alloNew60=allo1020C50;
if fbase<50 and ftot>=60 then alloNew60=Nt;

* adaptation finale;
alloFinC=alloNew60;
* Requis: variance total EPT par strate de tirage selon alloNew60NR: varEPTCNRNew60;
*  avec alloNew60NR= max(floor(alloNew60*taux),1),;
*  et   somme varEPTCNRNew60 par NOGA2reg: sumvarNew60;
pvar=varEPTCNRNew60/sumvarNew60*100;
* augmentation;
if noga2reg in (18, 67, 70, 71, 3637, 17, 90, 91, 92, 93) and
   pvar>5 and takoNew60=0 then alloFinC=min(50,Nt);
* remplissage dès 50%;
if alloFinC/Nt>=0.5 then AlloFinC=Nt;
```

Références

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Hidiroglou, M. A. (1993). The construction of a self-representing stratum of large units in survey design. *The American Statistician*, *40*(1), 27-31.

Lavallée, P. (2002). *Le sondage indirect ou la méthode généralisée du partage des poids*. Editions de l'Université de Bruxelles et Ellipses.

Lavallée, P., Deville, J.-C. (2002). Theoretical Foundations of the Generalised Weight Share Method. In *Proceedings of the International Conference on Recent Advances in Survey Sampling, 10-13 juillet 2002.*

Renaud, A. (2006). *Statistique suisse des bénéficiaires de l'aide sociale. Pondération des communes 2004* (Office fédéral de la statistique. Rapport de méthodes 338-0039).

Särndal, C. E. (2007). La méthode de calage dans la théorie et la pratique des enquêtes. *Techniques d'enquêtes*, *33*(2), 113-135.

Särndal, C. E. and Lundström, S. (2007). *Estimation in survey with nonresponse*. John Wiley & Sons.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. Springer-Verlag, New-York.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Publikationsprogramm BFS

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

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

Programme des publications de l'OFS

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

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

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

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

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

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

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

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

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

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

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

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

La statistique de l'emploi a été révisée en 2007 suite au dernier recensement des entreprises et des établissements. Cette révision a permis d'intégrer un nouveau système de cadre de sondage et de renouvellement de l'échantillon. Le plan d'échantillonnage a également été revu. Il a été construit dans le but de satisfaire les objectifs de précision de l'enquête tout en intégrant les cas spéciaux tels le «profiling» ou les multi-établissements. La stratification et l'allocation ont été revues fondamentalement. Elles sont accompagnées de remarques sur des alternatives pour une application future. Le nouvel échantillon contient 62'833 établissements, dont 41'553 normaux et 21'280 spéciaux. Il a été utilisé dès le 2ème trimestre 2007 et sera renouvelé, avec recouvrement, au 2ème trimestre 2008 sur la base d'un cadre de sondage actualisé.

N° de commande: 338-0052
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 978-3-303-00407-4