Rapport de méthodes

Enquête suisse sur la structure des salaires 2002

0

Statistische Grundlagen und Übersichten
Bases statistiques et produits généraux
Basi statistiche e presentazioni generali

Plan d'échantillonnage et extrapolation pour le secteur privé

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

OFS BFS UST

Neuchâtel, 2004

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 Verkehr und Nachrichtenwesen
12 Geld, Banken, Versicherungen
13 Soziale Sicherheit
14 Gesundheit
15 Bildung und Wissenschaft
16 Kultur, Medien, Zeitverwendung
17 Politik
18 Öffentliche Verwaltung und Finanzen
19 Rechtspflege
20 Einkommen und Lebensqualität der Bevölkerung
21 Nachhaltige Entwicklung und regionale Disparitäten

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 Transports et communications
12 Monnaie, banques, assurances
13 Protection sociale
14 Santé
15 Education et science
16 Culture, médias, emploi du temps
17 Politique
18 Administration et finances publiques
19 Droit et justice
20 Revenus et qualité de vie de la population
21 Développement durable et disparités régionales

Statistik der Schweiz

Methodenbericht

Statistique de la Suisse

Rapport de méthodes

Enquête suisse sur la structure des salaires 2002

Plan d'échantillonnage et extrapolation pour le secteur privé

Auteur **Monique Graf**
Office fédéral de la statistique

Editeur **Office fédéral de la statistique**

OFS BFS UST

Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
Swiss Federal Statistical Office

Neuchâtel, 2004

Préambule

L'enquête suisse sur la structure des salaires 2002 est la première pour laquelle les estimations au niveau des grandes régions (NUTS 2) sont prises en compte au stade de la planification. La régionalisation induite a posteriori en 2000 par l'introduction d'une enquête exhaustive dans deux cantons (Genève et Tessin) est donc redéfinie de telle sorte que chaque grande région corresponde à une combinaison de strates. En outre la stratification tient compte des cantons ayant demandé une enquête spécifique. Ce rapport décrit en détails la méthode utilisée. L'estimation de la précision fera l'objet d'une étude séparée.

Un grand merci à Anne Renaud (METH) qui a relu ce rapport et à Jacques Méry (LOHN) pour les nombreuses discussions. Mes remerciements vont aussi à Sabine Herbel (CSC) qui a préparé les fichiers de données.

Résumé

L'enquête sur la structure des salaires est une enquête bisannuelle auprès des entreprises dont le but principal est d'estimer des salaires médians pour un certain nombre de domaines (classe d'activité, niveau de qualification, sexe ...). Le présent rapport décrit les aspects méthodologiques du plan d'échantillonnage et le calcul des poids d'extrapolation pour l'enquête suisse sur la structure des salaires 2002.

Le plan d'échantillonnage est décrit en détail. Il prévoit un échantillon stratifié en grappes à deux niveaux, comme dans les enquêtes précédentes. La différence avec le passé apparaît au niveau de la définition des strates. La stratégie est de garantir une précision analogue pour l'estimation du salaire standardisé moyen au niveau grande région, puis de faire les adaptations nécessaires pour les cantons désirant une enquête spécifique.

La pondération reprend pour l'essentiel les poids d'extrapolation de Horvitz-Thompson. La non-réponse étant supposée ignorable dans les strates, les poids sont donc basés sur l'effectif par strate des entreprises répondantes. Les poids jugés trop grands sont alors tronqués et le déficit est distribué sur les autres classes de taille de même région et classe d'activité économique.

Mots-clé

Rapport de méthodes; LSE; ESS; plan d'échantillonnage; variance; salaire.

Complément d'information: Monique Graf, tél. 032 713 66 15
monique.graf@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: 338-0025

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 2004
La reproduction est autorisée, sauf à des fins commerciales,
si la source est mentionnée.

ISBN: 3-303-00296-7

Table des matières

1 Plan d'échantillonnage pour le secteur privé **5**
1.1 Introduction 5
1.2 La stratification 5
1.3 Le plan suisse 7
1.4 Plan pour les grandes régions 13
1.5 Adaptation aux cantons 15
1.6 Résultats 16

2 Extrapolation **19**
2.1 Poids initiaux 19
2.2 Variations dans le cadre de sondage 19
2.3 Non réponse totale dans une strate 20
2.4 Pondération finale 22

Annexes **22**

Références **25**

1 Plan d'échantillonnage pour le secteur privé

1.1 Introduction

La statistique 2002 sur la structure des salaires se caractérise par une régionalisation. La stratification tient compte non seulement des 7 grands régions, mais aussi de certains cantons ayant demandé une enquête dont la précision soit analogue à celle de l'enquête suisse 2000. Comme la décision définitive de ces cantons au sujet d'une enquête cantonale dépendait de l'option choisie au niveau suisse (enquête suisse, comme les années précédentes, ou enquête plus lourde, garantissant la précision par grande région), nous avons prévu 3 plans pour parer à toute éventualité :

- un plan A à 15'000 entreprises, analogue à celui de 2000
- un plan B à 45'000 entreprises, option grandes régions
- un plan C à 56'000 entreprises, pour le cas où il serait nécessaire de fournir en plus des estimations par grandes régions, également des estimations pour les cantons de VD, NE, FR, SG et ZG, ayant fait une demande provisoire. A cette liste se rajoutent les cantons de ZH et TI qui sont en eux-mêmes des grandes régions, et GE qui avait déjà demandé une enquête complémentaire en 2000.

Finalement, c'est le plan C à 56'000 entreprises qui a été réalisé. Ce plan a été construit par étapes (voir plus bas).

1.2 La stratification

La stratification est basée sur 3 classifications : la classe d'activité, la classe de taille de l'entreprise et une classification géographique. Les strates du plan d'échantillonnage sont des combinaisons de ces trois classifications.

1.2.1 Classe d'activité

Elle est donnée comme les années précédentes par la NOGA2, avec certains regroupements. Plus précisément, la classe d'activité est définie par la variable UNOGA du REE, qui est l'activité de la majorité des employés de l'entreprise. La variable définissant la subdivision en classes d'activité est notée NOG_2. Il n'y a donc pas de changement à ce niveau par rapport aux années précédentes [4].

1.2.2 Classe de taille

Pour diminuer la prolifération des strates, nous avons regroupé les 5 classes de tailles des années précédentes en 3 seulement, variable TA3. Ce sont des classes de taille fixe, basée sur la variable UBETOT du REE (Table 1)[1]. Remarquons qu'en 2000, les entreprises de taille UBETOT=2 étaient incluses dans la classe 1 ; elles sont exclues en 2002.

[1] Cette variable s'appelle ANZBE dans le fichier des résultats.

TAB. 1 – Classes de taille

Nouvelle subdivision	3-19	20-49	50+
TA3	1	2	3
Classes 2000 (GRAN)	1+2	3	4+5

1.2.3 Classification géographique

La classification géographique est donnée par les sous-régions GRS des grandes régions statistiques GR (Table 2). On donne aussi en regard la subdivision géographique de l'enquête 2000 (où RS signifie "reste Suisse"). On voit que la classification définie par GRS est une subdivision des régions 2000, variable NUKT.

TAB. 2 – Classes géographiques

Grande région	GR	GRS	Cantons	Régions 2000	NUKT
Région lémanique	1	11	GE	GE	2
		12	VD	RS	1
		13	VS	RS	1
Espace Mittelland	2	21	FR	RS	1
		21	NE	RS	1
		22	BE+SO+JU	RS	1
Suisse du Nord-Ouest	3	30	AG+BL+BS	RS	1
Zurich	4	40	ZH	RS	1
Suisse Orientale	5	51	SG	RS	1
		52	SH+TG+AI+AR+GL+GR	RS	1
Suisse centrale	6	61	ZG	RS	1
		62	UR+SZ+NW+OW+LU	RS	1
Tessin	7	71	TI	TI	3

Remarques :

1. FR et NE ont le même code GRS, car dans ces cantons on a procédé à une enquête exhaustive.
2. Le canton UKTO de l'entreprise est défini comme le canton de la majorité des employés de classe d'activité UNOGA. Ce canton n'est pas nécessairement le même que le canton du plus grand établissement de l'entreprise de classe UNOGA. Le tableau croisé de ces deux définitions du canton (Annexe B) montre qu'il n'y a de différences que pour 47

des 157'345 entreprises du secteur privé inscrites au REE au moment du tirage. Nous conservons donc UKTO comme canton de référence de l'entreprise.

3. La région (GR ou GRS) de l'entreprise est la région d'appartenance de son canton de référence UKTO.

1.3 Le plan suisse

Nous appellerons plan "standard" le plan d'échantillonnage basé sur la régionalisation de 2000, car c'est la base statistique sur laquelle nous pouvons faire des calculs de précision (Table 2). Rappelons que ces calculs basés sur l'enquête précédente servent à obtenir les nouveaux effectifs à tirer dans les strates : l'effectif augmente avec la variabilité du salaire moyen par entreprise dans la strate. Ces calculs n'ont aucune influence sur le choix des entreprises interrogées, qui sont tirées au hasard parmi les entreprises de la strate. Il a été décidé a priori d'interroger toutes les entreprises de 50 employés et plus. Dans le registre en juin 2002, il y avait 6855 entreprises de taille TA3=3, sur un total de 156177. Le problème du plan standard est donc le suivant :

Problème du plan standard : *Pour une taille totale donnée, trouver l'allocation dans les strates, combinaisons de* NUKT, NOG_2 *et de* TA3 *= 1 ou 2, qui estime le salaire moyen global avec la meilleure précision possible.*

Remarques :

1. La précision est définie par le coefficient de variation (CV) du salaire moyen. Une analyse des résultats de l'enquête 2000 montre que ce CV est du même ordre de grandeur que le CV synthétique de la médiane [4]. L'avantage de prendre le CV de la moyenne comme base de calcul est qu'on peut utiliser des méthodes connues et bien décrites. C'est aussi la méthode utilisée dans le passé.
2. En pratique on désire une bonne précision à un niveau plus détaillé. Le plan standard aura tendance à favoriser (en terme d'effectif à tirer) les strates dans lesquelles le salaire moyen est le plus différent de la moyenne globale. Cette propriété est assez raisonnable.

Le principe de la méthode consiste à faire dans une première étape une allocation de Neyman basée sur la variance inter-entreprises, voir par exemple [6, p. 106]. Dans une deuxième étape, les effectifs <10 sont ajustés de telle sorte que la taille minimum soit de 10 entreprises, ou que la strate soit exhaustive (au cas où l'effectif de population serait <10). Cette taille est la taille nette souhaitée de l'échantillon. Enfin, dans une troisième étape, la taille nette est amplifiée par l'inverse du taux de réponse prévu, ce qui donne la taille brute, c'est-à-dire le nombre d'entreprises à interroger dans la strate.

Le taux de réponse prévu est basé sur les résultats des enquêtes précédentes. On le trouve à la Table 3 avec le taux de sondage intra-entreprises, caractéristique du plan également dépendante de la taille. A cause de cette non-réponse, et même avec un taux de sondage de

100%, il n'est pas toujours possible dans les strates à petits effectifs de garantir la même précision qu'ailleurs, car la taille brute est limitée par l'effectif total de la strate.

TAB. 3 – Paramètres du plan d'échantillonnage dépendant de la taille

classe de taille	taux de sondage (%) 2ème niveau	taux de réponse (%) 1er niveau
TA3	f_{2h}	α_h
1	100	65
2	50	75
3	33	80

Dans le plan suisse, nous optimisons la précision au niveau global. Nous calculons d'abord l'estimation de la variance inter-entreprise du salaire MBLS dans chaque strate de la LSE 2000, puis faisons une moyenne pondérée de ces variances sur les trois sous-régions NUKT. Sa racine donne alors un écart-type pondéré qui est proportionnel à l'effectif à tirer dans la strate. Pour les strates ayant une classe TA3=1, cette variance est la variance totale, puisque les entreprises concernées nous livrent en principe tous les salaires. Pour TA3=2, on pourrait encore tenir compte de la variance intra-entreprise, mais cet effet a toujours été négligé au stade de la planification.

1.3.1 Formalisation du problème

Dans cette formalisation du problème, nous allons dans un premier temps tenir compte des deux composantes de la variance (inter et intra), puis, au niveau de l'application, nous effectuerons la simplification usuelle consistant à négliger la part intra. Nous reprenons les mêmes notations que dans [4, p. 19]. Nous les rappelons ici pour mémoire (Table 4). Notons encore y_j le salaire mensuel brut standardisé (MBLS) du salarié j de l'entreprise hi de la strate h. Les différentes transformations du salaire sont données à la Table 5. Ces transformations apparaissent dans le calcul de la variance dans la strate.

Le salaire pondéré moyen $\overline{y}$ estimé au niveau global doit être vu comme un quotient de variables aléatoires.

$$\overline{y} = \frac{\sum_h N_h \overline{v}_h \overline{y}_h}{\sum_h N_h \overline{v}_h} \tag{1}$$

Sa valeur théorique au niveau global est donnée par :

$$\overline{Y}_U = \frac{\sum_h \sum_{i \in U_h} \sum_{j \in U_{hi}} b_j y_j}{\sum_h \sum_{i \in U_h} \sum_{j \in U_{hi}} b_j}$$

où $\sum_h \sum_{i \in U_h} \sum_{j \in U_{hi}} b_j = \sum_h \sum_{i \in U_h} M_{hi} \overline{b}_{U,hi} = \sum_h N_h \overline{v}_{U,h}$ représente la taille de la population des emplois en équivalents plein temps.

TAB. 4 – Notations

N	nombre d'entreprises
N_h	nombre d'entreprises dans la strate h
U_h	univers des entreprises de la strate h
s_h	échantillon d'entreprises de la strate h
n_h	taille de s_h
$f_h = n_h/N_h$	taux de sondage dans la strate h
U_{hi}	univers des emplois de l'entreprise i de la strate h
M_{hi}	nombre d'emplois dans U_{hi}
s_{hi}	échantillon d'emplois dans l'entreprise hi
m_{hi}	taille de s_{hi}
$f_{hi} = m_{hi}/M_{hi}$	taux de sondage dans l'entreprise hi
b_j	taux d'occupation standardisé de l'emploi j de s_{hi}
w_{hi}	poids d'extrapolation des emplois de s_{hi}
$\overline{b}_{hi} = \sum_{s_{hi}} b_j/m_{hi}$	taux d'occupation standardisé moyen dans s_{hi}
$v_{hi} = w_{hi}\overline{b}_{hi}$	poids d'extrapolation pondéré par le taux d'occupation moyen
$\overline{v}_h = \sum_{s_h} v_{hi}/n_h$	poids d'extrapolation pondéré moyen dans s_h
$\overline{b}_{U,hi} = \sum_{U_{hi}} b_j/M_{hi}$	taux d'occupation standardisé moyen dans U_{hi}
$M_{hi}\overline{b}_{U,hi} = \sum_{U_{hi}} b_j$	taille réelle de l'entreprise U_{hi} en équivalents plein temps
$N_h\overline{v}_{U,h} = \sum_{U_h} M_{hi}\overline{b}_{U,hi}$	taille réelle de la strate U_h en équivalents plein temps

La variance approchée du salaire pondéré moyen $\overline{y}$ est établie par étapes :

1. En notant **AV** la variance approchée par linéarisation, et **V** la variance exacte, on obtient :

$$\begin{aligned}\mathbf{AV}\left(\overline{y}\right) &= \frac{1}{\left(\sum_h N_h\overline{v}_{U,h}\right)^2}\mathbf{V}\left(\sum_h N_h\overline{v}_h\left(\overline{y}_h-\overline{y}\right)\right)\\ &= \frac{1}{\left(\sum_h N_h\overline{v}_{U,h}\right)^2}\sum_h\left(\frac{N_h}{n_h}\right)^2\mathbf{V}\left(\sum_{i\in s_h} v_{hi}\left(\overline{y}_{hi}-\overline{y}\right)\right) \qquad (2)\end{aligned}$$

2. On estime la variance $\mathbf{V}\left(\sum_{i\in s_h} v_{hi}\left(\overline{y}_{hi}-\overline{y}\right)\right)$ dans une strate h. Posons $\breve{\overline{e}}_{hi} = v_{hi}\left(\overline{y}_{hi}-\overline{y}\right)$, voir Table 5 :

$$\widehat{\mathbf{V}}\left(\sum_{i\in s_h} v_{hi}\left(\overline{y}_{hi}-\overline{y}\right)\right) = \widehat{\mathbf{V}}\left(\sum_{i\in s_h}\breve{\overline{e}}_{hi}\right) \qquad (3)$$

Tenant compte du plan en grappes dans la strate h, nous définissons les parts inter et intra entreprise de la variance.

TAB. 5 – Transformations du salaire

y_j	salaire standardisé de l'employé j de l'entreprise hi
$\overline{y}_{hi} = \left\{ \begin{array}{l} \sum_j b_j y_j / \sum b_j = \\ \frac{1}{m_{hi}} \sum_j \left(b_j / \overline{b}_{hi}\right) y_j \end{array} \right.$	salaire moyen dans l'entreprise hi pondéré par le taux d'occupation
$\sum_i v_{hi}$	nombre d'emplois estimés dans la strate h en tenant compte du taux d'occupation (nb. estimé d'équivalents plein temps)
$\overline{y}_h = \left\{ \begin{array}{l} \sum_i v_{hi} \overline{y}_{hi} / \sum_i v_{hi} = \\ \frac{1}{n_h} \sum_i \left(v_{hi} / \overline{v}_h\right) \overline{y}_{hi} \end{array} \right.$	salaire moyen dans la strate h pondéré par le poids d'extrapolation tenant compte du taux d'occupation
$\breve{e}_j = \left\{ \begin{array}{l} w_{hi} b_j \left(y_j - \overline{y}\right) = \\ v_{hi} \left(b_j / \overline{b}_{hi}\right) \left(y_j - \overline{y}\right) \end{array} \right.$	écart du salaire j au salaire moyen global multiplié par le poids d'extr. de l'entreprise hi et le taux d'occupation de l'employé j de hi
$\overline{\breve{e}}_{hi} = \left\{ \begin{array}{l} \frac{1}{m_{hi}} \sum_j \breve{e}_j = \\ v_{hi} \left(\overline{y}_{hi} - \overline{y}\right) \end{array} \right.$	écart du salaire moyen de l'entreprise hi au salaire moyen global, multiplié par le poids d'extrapolation tenant compte du taux d'occupation
$\overline{\breve{e}}_h = \left\{ \begin{array}{l} \frac{1}{n_h} \sum_i \overline{\breve{e}}_{hi} = \\ \overline{v}_h \left(\overline{y}_h - \overline{y}\right) \end{array} \right.$	écart du salaire moyen de la strate h au salaire moyen global, multiplié par le poids moyen d'extrapolation tenant compte du taux d'occupation

Part inter : Soit A_h la variance empirique des moyennes $\overline{\breve{e}}_{hi}$ dans s_h (Table 5). A_h est donnée par (4)

$$A_h = \frac{1}{n_h - 1} \sum_{i \in s_h} \left(\overline{\breve{e}}_{hi} - \overline{\breve{e}}_h\right)^2 \tag{4}$$

Part intra : Soit B_{hi} la variance empirique dans l'entreprise hi des écarts extrapolés $\breve{e}_j$ dans s_{hi} (5), et B_h la part intra-entreprise de la variance de la strate (6). Elles valent :

$$B_{hi} = \frac{1}{m_{hi} - 1} \sum_{j \in s_{hi}} \left(\breve{e}_j - \overline{\breve{e}}_{hi}\right)^2 \tag{5}$$

$$B_h = \sum_{i \in s_h} m_{hi} \left(1 - f_{hi}\right) B_{hi} \tag{6}$$

La variance (3) est donnée par

$$\widehat{\mathbf{V}} \left(\sum_{i \in s_h} \overline{\breve{e}}_{hi} \right) = \left\{ n_h \left(1 - f_h\right) A_h + f_h B_h \right\} \tag{7}$$

3. La variance approchée (2) est estimée par (8). En substituant (7), on obtient (9), puis

(10) en remplaçant f_h par n_h/N_h.

$$\widehat{\mathbf{V}}_{\text{str,2st}}(\overline{y}) = \frac{1}{\left(\sum_h N_h \overline{v}_h\right)^2} \sum_h \left(\frac{N_h}{n_h}\right)^2 \widehat{\mathbf{V}} \left(\sum_{i \in s_h} \overline{\breve{e}}_{hi}\right) \tag{8}$$

$$= \frac{1}{\left(\sum_h N_h \overline{v}_h\right)^2} \sum_h \left(\frac{N_h}{n_h}\right)^2 (n_h(1-f_h)A_h + f_h B_h) \tag{9}$$

$$= \frac{1}{\left(\sum_h N_h \overline{v}_h\right)^2} \sum_h \left(-N_h A_h + \frac{N_h^2 \left(A_h + B_h/N_h\right)}{n_h}\right) \tag{10}$$

L'expression (10) est sous la bonne forme pour l'allocation de Neyman, car l'espérance des variables aléatoires $\overline{v}_h$, A_h et B_h ne dépend pas de n_h. En première approximation, $\widehat{\mathbf{V}}_{\text{str,2st}}(\overline{y})$ ne dépend donc de n_h que par le dénominateur dans le dernier terme.

1.3.2 Allocation

Pour une taille n d'échantillon fixée, l'allocation de Neyman consiste à chercher les effectifs n_h dans les strates qui minimisent $\widehat{\mathbf{V}}_{\text{str,2st}}(\overline{y})$. Cela revient à minimiser le critère C_0 (11) comme fonction de n_h $(h = 1, ..., H)$ et λ, où λ est un multiplicateur de Lagrange :

$$C_0 = \sum_h \frac{N_h^2 \left(A_h + B_h/N_h\right)}{n_h} + \lambda \left(\sum_h n_h - n\right) \tag{11}$$

On trouve la valeur optimale de n_h minimisant C_0 (12) :

$$n_h = n \frac{N_h \sqrt{A_h + B_h/N_h}}{\sum_{h'} N_{h'} \sqrt{A_{h'} + B_{h'}/N_{h'}}} \tag{12}$$

Cette allocation est celle qui minimise la variance de la moyenne générale $\overline{y}$.

Le plan d'échantillonnage standard au niveau suisse, déduit de cette allocation, consiste à tirer au hasard $\sum_{h \in \text{NOG_2×TA3}} n_h$ entreprises dans NOG_2×TA3, en "oubliant" les 3 sous-régions de 2000.

Ce plan d'échantillonnage ne concerne que les classes TA3=1 ou 2 (la classe 3 est exhaustive). Dans les enquêtes précédentes, le terme B_h/N_h associé à la variance intra a été négligé. Il est d'ailleurs théoriquement nul si TA3=1, car on demande tous les salaires des entreprises tirées de cette classe de taille. Le fait de négliger B_h/N_h pour TA3=2 revient à tirer un peu plus de petites entreprises par rapport aux entreprises de taille moyenne que la formule (12) ne l'aurait prescrit. Nous avons repris cette méthode pour 2002[2]. Avec cette simplification,

[2]Il y a déjà beaucoup de changements au niveau de la structure de l'enquête (introduction des grandes régions et de certains cantons). Nous n'avons pas voulu changer aussi la méthode d'allocation.

l'effectif échantillonné pour la combinaison NOG_2×TA3 est donc donné par

$$n_{\text{NOG_2×TA3}} = \sum_{h \in \text{NOG_2×TA3}} n_h = n \frac{\sum_{h \in \text{NOG_2×TA3}} N_h \sqrt{A_h}}{\sum_{h' \in U_{12}} N_{h'} \sqrt{A_{h'}}} \tag{13}$$

où U_{12} est l'ensemble de toutes les strates associées aux classes de taille 1 et 2.

1.3.3 Prise en compte de la non-réponse

Nous savons, d'après les enquêtes précédentes, que le taux de réponse des entreprises dépend de la taille, voir Table 3. La formule (13) est donc modifiée comme suit :

$$n_{\text{NOG_2×TA3}} = n \frac{\sum_{h \in \text{NOG_2×TA3}} \left(N_h \sqrt{A_h}\right) / \alpha_h}{\sum_{h' \in U_{12}} \left(N_{h'} \sqrt{A_{h'}}\right) / \alpha_{h'}} \tag{14}$$

où α_h est fixé à 0.65 (resp. 0.75) pour les strates appartenant à la classe de taille TA3 =1 (resp. 2).

1.3.4 Aménagements par rapport à l'allocation de Neyman

L'allocation (14) doit encore subir quelques aménagements pour être pratiquement utilisable.

1. La taille brute est limitée par l'effectif total des entreprises de la strate.
2. Il n'est pas souhaitable d'avoir de trop petits effectifs par strate, car en cas de forte non-réponse, on risque de se trouver en présence de strates totalement inobservées. La borne inférieure est fixée à 10 entreprises à interroger au minimum par strate.
3. Comme nous avons défini nos strates par croisement de classes d'activité et de taille, il peut arriver que, dans de rares cas, l'effectif total des entreprises de la strate soit inférieur à 10. Dans ces cas, on prend toute la strate.
4. Les effectifs des strates non exhaustives sont enfin multipliés par un facteur calculé de telle sorte que la taille totale souhaitée de l'échantillon soit conservée[3].

Le plan CH standard est calculé pour une taille cible de 45'000 entreprises. Comme il y a 6855 entreprises de classe de taille TA3=3, on pose $n = 45000 - 6855 = 38145$ dans la formule (14). Le plan obtenu avec les aménagements ci-dessus a pour taille finale 38029 entreprises.

[3] Une tolérance de 1000 sur le total général est introduite pour simplifier l'algorithme.

1.4 Plan pour les grandes régions

1.4.1 Hypothèse de base

On part d'une hypothèse d'homogénéité des variances inter pour les strates analogues des grandes régions. Nous faisons cette hypothèse pour l'enquête 2002, car les résultats de 2000 ne fournissent pas d'estimations fiables au niveau des grandes régions. Elle ne sera pas nécessaire en 2004.

On commence par calculer le plan standard CH d'un effectif total de 45'000 entreprises, décrit ci-dessus. Cela définit des taux de sondage provisoires identiques pour les strates analogues des 7 grandes régions.

Indiçons par k les nouvelles strates, définies par les combinaisons NOG_2×TA3×GR. Soit N_k l'effectif total des entreprises dans la nouvelle strate k. (Les anciennes strates de l'enquête 2000, combinaisons de NUKT×NOG_2×GRAN - voir Tables 1 et 2 -, sont toujours indicées par h.).

Représentons chaque entreprise i de la strate k par son salaire moyen pondéré $\overline{y}_{U,ki}, k = 1, 2, ..., N_k$, où l'indice U signifie que la moyenne est calculée sur tous les emplois. Soit σ_k^2 la variance dans la strate k des écarts pondérés des salaires à la moyenne générale, $\overline{b}_{U,ki}\left(\overline{y}_{U,ki} - \overline{y}_U\right)$:

$$\sigma_k^2 = \mathbf{V}\left(\overline{b}_{U,ki}\left(\overline{y}_{U,ki} - \overline{y}_U\right)\right)$$

L'hypothèse que les variances inter sont homogènes s'exprime par

$$\sigma^2_{\text{NOG_2×TA3×GR}} = \sigma^2_{\text{NOG_2×TA3}}$$

que l'on peut interpréter en disant que la variance des écarts pondérés des entreprises de la strate k à la moyenne générale dépend de la classe d'activité et de la classe de taille, mais pas de la grande région.

1.4.2 Allocation de Neyman

Posons dans (14)[4]

$$\begin{aligned} N_{\text{NOG_2×TA3}} &= \sum_{h \in \text{NOG_2×TA3}} N_h \\ \sqrt{A'_{\text{NOG_2×TA3}}} &= \frac{\sum_{h \in \text{NOG_2×TA3}} N_h \sqrt{A_h}}{N_{\text{NOG_2×TA3}}} \end{aligned}$$

$\sqrt{A'_{\text{NOG_2×TA3}}}$ est donc un écart-type inter moyen estimé. Cela permet d'exprimer (14) à l'aide de combinaisons des ancienne strates h qui sont aussi des combinaisons des nouvelles strates k :

[4] C'est possible, car chaque ancienne strate h appartient à une seule combinaison NOG_2×TA3.

$$
\begin{aligned}
n_{\text{NOG_2×TA3}} &= n\frac{N_{\text{NOG_2×TA3}}\sqrt{A'_{\text{NOG_2×TA3}}}/\alpha_{\text{TA3}}}{\sum\limits_{\text{NOG_2×TA3}} N_{\text{NOG_2×TA3}}\sqrt{A'_{\text{NOG_2×TA3}}}/\alpha_{\text{TA3}}} \\
&\approx N_{\text{NOG_2×TA3}}\sqrt{A'_{\text{NOG_2×TA3}}}/\alpha_{\text{TA3}} \qquad (15)
\end{aligned}
$$

(où "$\approx$" signifie "proportionnel à").

L'hypothèse d'homogénéité des variances inter appliquée aux écarts-type inter moyens estimés s'exprime par

$$
\sqrt{A'_k} \cong \sqrt{A'_{\text{NOG_2×TA3}}} \text{ pour tout } k = \text{NOG_2} \times \text{TA3} \times \text{GR}
$$

(où "$\cong$" signifie "approximativement égal").

L'allocation de Neyman dans les nouvelles strates revient à minimiser le critère C_1 :

$$
C_1 = \sum_k \frac{(N_k/\alpha_k)^2 A'_k}{n_k} + \lambda\left(\sum_k n_k - n\right)
$$

1.4.3 Recherche d'une précision homogène au niveau grande région

On sait que cette allocation a tendance donner une précision beaucoup plus élevée sur la moyenne générale et dans les strates de grand effectif que dans les petites strates. Pour éviter la sous-représentation des petites unités géographiques, on peut pondérer la première somme de C_1. On obtient alors le critère C_2 :

$$
C_2 = \sum_k t_k \frac{(N_k/\alpha_k)^2 A'_k}{n_k} + \lambda\left(\sum_k n_k - n\right)
$$

où t_k est le poids de la strate k, spécifié plus bas. L'allocation optimale minimisant C_2 est donnée par :

$$
n_k = n\frac{\sqrt{t_k}N_k\sqrt{A'_k}/\alpha_k}{\sum\limits_{k'}\sqrt{t_{k'}}N_{k'}\sqrt{A'_{k'}}/\alpha_{k'}}
$$

Comme nous ne voulons pas modifier l'allocation relative à l'intérieur des grandes régions, le poids ne dépendra que de GR.

Soit N_{GR} le nombre total d'entreprises de classe de taille TA3=1+2 de la région GR. La valeur du poids choisie est $t_k = N_{\text{GR}}^{-1}, k \in$GR. Ainsi, le poids de la région GR dans la première somme de C_2 est inversement proportionnel à N_{GR}, de sorte que le critère C_2 optimise la somme des variances moyennes par grande région (sous la contrainte que la taille totale de l'échantillon soit égale à la valeur donnée n). Cette méthode suit le même raisonnement que la "Power allocation" de Bankier (1988), voir [6, p. 471].

L'allocation pour l'enquête 2002 est donc donnée par

$$n_{\text{NOG_2×TA3×GR}} = n\frac{N_{\text{GR}}^{-\frac{1}{2}} N_{\text{NOG_2×TA3×GR}}\sqrt{A'_{\text{NOG_2×TA3}}}/\alpha_{\text{TA3}}}{\sum_{\text{GR}} N_{\text{GR}}^{-1/2} \sum_{k'\in\text{GR}} N_{k'}\sqrt{A'_{k'}}/\alpha_{k'}} \quad (16)$$

$$\approx \frac{1}{\sqrt{N_{\text{GR}}}} N_{\text{NOG_2×TA3×GR}}\sqrt{A'_{\text{NOG_2×TA3}}}/\alpha_{\text{TA3}} \quad (17)$$

On peut simplifier la formule (17), en écrivant $n_{\text{NOG_2×TA3×GR}}$ en fonction de $n_{\text{NOG_2×TA3}}$, voir (15) :

$$n_{\text{NOG_2×TA3×GR}} \approx \frac{1}{\sqrt{N_{\text{GR}}}} N_{\text{NOG_2×TA3×GR}} \frac{n_{\text{NOG_2×TA3}}}{N_{\text{NOG_2×TA3}}}$$

$$= \frac{1}{\sqrt{N_{\text{GR}}}} N_{\text{NOG_2×TA3×GR}} f_{\text{NOG_2×TA3}} \quad (18)$$

où

- $f_{\text{NOG_2×TA3}}$ est le taux de sondage dans la catégorie NOG_2×TA3, déduit du plan standard ;
- $N_{\text{NOG_2×TA3×GR}}$ est le nombre d'entreprises dans la strate NOG_2×TA3×GR ;
- N_{GR} est le nombre total d'entreprises dans les classes de taille 1+2 de la grande région GR.

Le taux de sondage dans la strate NOG_2×TA3×GR est donné par

$$f_{\text{NOG_2×TA3×GR}} = \frac{n_{\text{NOG_2×TA3×GR}}}{N_{\text{NOG_2×TA3×GR}}}$$

$$= \frac{f_{\text{NOG_2×TA3}}/\sqrt{N_{\text{GR}}}}{\sum_{\text{NOG_2×TA3×GR}} f_{\text{NOG_2×TA3}}/\sqrt{N_{\text{GR}}}} \quad (19)$$

Les mêmes aménagements que ceux exposés sous 1.3.4 sont appliqués aux effectifs calculés à l'aide de (18).

1.5 Adaptation aux cantons

Certains cantons (GE, VD, SG, NE, FR et ZG) ont demandé de pouvoir bénéficier d'une enquête au niveau cantonal, dont la précision soit analogue à celle obtenue pour les grandes régions. Nous avons alors subdivisé les grandes régions concernées par ces cantons. La classification correspondante est donnée par GRS (Table 2). Les nouvelles strates sont donc les combinaisons NOG_2×TA3×GRS. On appellera "strate standard" les strates qui ne concernent pas ces cantons. L'adaptation du plan d'échantillonnage à cette situation comporte les points suivants, consignés dans une variable AUGM :

- Les taux de sondage $f_{\text{NOG_2×TA3×GR}}$ définis par (19) sont appliqués aux strates standard. Il n'y a pas d'augmentation de l'échantillon dans ces strates, donc AUGM= 0.

TAB. 6 – Facteur par région proportionnel à la racine du nb. d'entreprises BURNH

GR	BURNH	FACT
1	26432	0.835
2	33274	0.744
3	18418	1.000
4	25942	0.842
5	22959	0.895
6	14825	1.114
7	7472	1.570
		====
		7.000

- Pour les trois plus grands cantons (GE, VD et SG), ce taux est multiplié par 1.88, de façon à leur attribuer le même taux de sondage qu'au Tessin : en multipliant le facteur régional des régions 1 et 5 (Table 6) par 1.88, on obtient approximativement le taux de la région 7 (TI). Remarquons que ce facteur est légèrement supérieur à $\sqrt{N_{\text{GR}}/N_{\text{GRS}}}$. (Ce quotient vaut resp. 1.84, 1.49 et 1.55 pour GE, VD et SG). Nous posons alors AUGM=1. On va donc vers la sûreté.
- On procède à une enquête exhaustive dans les trois plus petits (NE, FR et ZG)[5]. Dans ce cas, AUGM=2.

La fig. 1 montre la répartition des taux de sondage dans les strates non exhaustives, pour le plan par grandes régions et les strates avec AUGM=0 ou 1 dans le plan adapté aux cantons. On voit que l'histogramme du plan par régions est bimodal. Cette bimodalité se retrouve logiquement au niveau des strates standard, alors que pour AUGM=1, le deuxième mode a disparu dans les strates exhaustives.

En conclusion, notre but a été d'établir un plan d'échantillonnage qui garantisse la comparabilité avec les enquêtes passées, tout en s'adaptant aux nouvelles demandes. Le plan d'échantillonnage 2002 est le premier qui soit construit pour l'estimation séparée au niveau NUTS 2 (grande région). Il se base sur une hypothèse d'homogénéité de la variation des salaires dans les entreprises de même classe de taille et de même classe d'activité dans les différentes grandes régions. Cette hypothèse était implicite dans les enquêtes CH publiées précédemment. Elle pourra être levée en 2004.

1.6 Résultats

Le plan par régions (TA3=1 ou 2) contient 542 strates non vides, parmi lesquelles 158 sont non exhaustives. Pour le plan adapté aux cantons, la répartition des 897 strates, exhaustives

[5] Ceci explique pourquoi NE et FR ont le même nom GRS.

ou non, selon la valeur de AUGM (voir § 1.5) est donnée à la Table 7. Les Tables 8 et 9 donnent la répartition pour les classifications par grande région et taille. La répartition des strates exhaustives ou non par classes d'activité est donnée à la Table 10. Chaque classe d'activité est subdivisée en 14 strates au maximum - certaines peuvent être vides - dans le plan par régions et 24 dans le plan adapté aux cantons. On a, dans les deux cas, un tirage exhaustif de 21 classes d'activité sur les 40 classes définies pour le secteur privé.

La comparaison entre les univers de 2000 et de 2002 montre que, par combinaison NOG_2×TA_3, le nombre d'entreprises est analogue, exception faite des classes NOG_2 = 52, 55, 72 et 85, pour lesquelles la classe de taille 1 est surreprésentée en 2002. Il y a par exemple, en 2002, 11725 entreprises de taille 1 dans la classe NOG_2 = 85, alors qu'elles n'étaient que 5763 en 2000, ceci, malgré le fait que nous avons exclu de l'univers les entreprises de taille UBETOT=2.

TAB. 7 – Nombre de strates dans le plan adapté aux cantons

Strates	AUGM			
exhaustives	0	1	2	total
non	157	43	0	200
oui	375	181	141	697
total	532	224	141	897

TAB. 8 – Nombre de strates exhaustives ou non par grande région

	plan régions		plan cantons	
GR	non	oui	non	oui
1	25	54	53	166
2	25	53	25	125
3	23	55	23	55
4	25	52	25	52
5	25	53	39	115
6	20	56	20	123
7	15	61	15	61

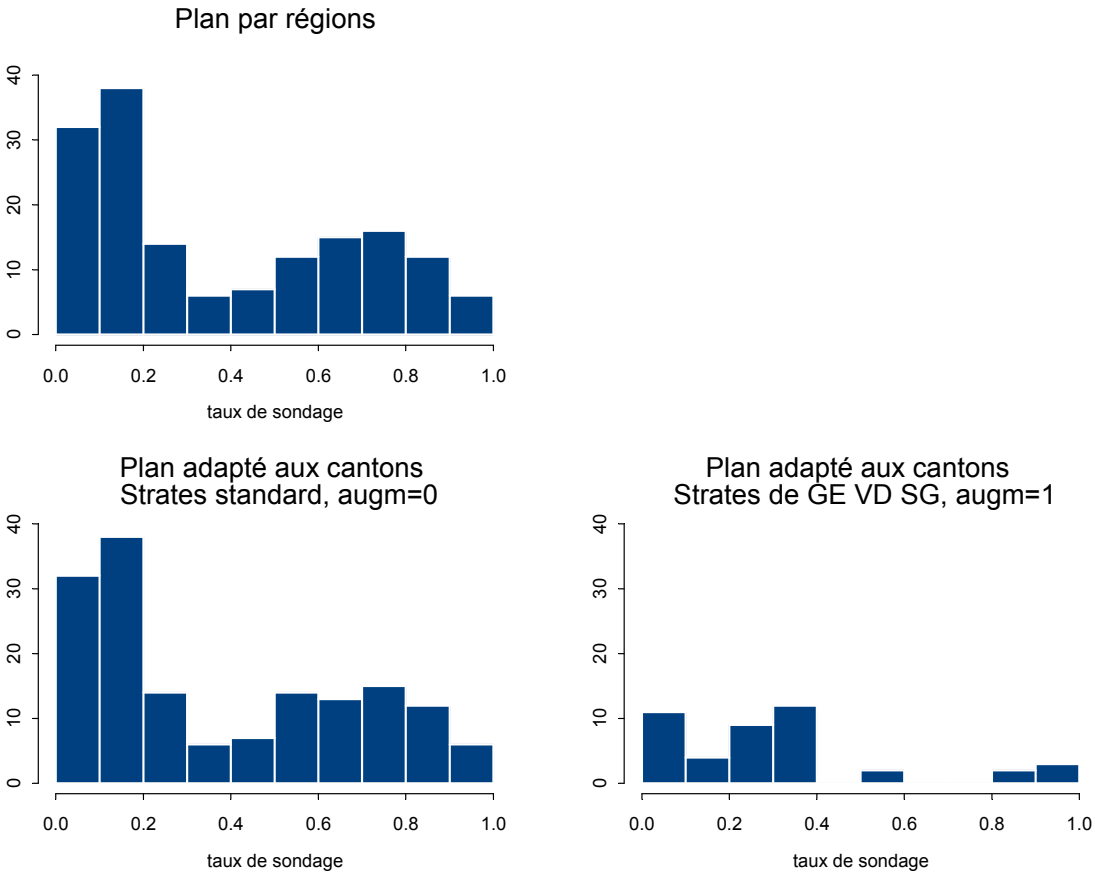


FIG. 1 – Répartition des taux de sondage dans les strates non exhaustives.

TAB. 9 – Nombre de strates exhaustives ou non par classe de taille

	plan régions		plan cantons	
TA3	non	oui	non	oui
1	123	150	161	302
2	35	234	39	395

2 Extrapolation

2.1 Poids initiaux

- Poids inter : Le poids inter entreprises INITGEW1 est le rapport du nombre total d'entreprises selon le REE sur le nombre d'entreprises répondantes de la strate. C'est donc l'inverse du produit du taux de sondage et du taux de réponse dans la strate. Les taux de réponse par classe de taille sont donnés à la Table 11. On constate qu'ils sont meilleurs que les taux utilisés au niveau de la planification (voir Table 3).
- Poids intra : La variable UBETOT sur laquelle est basée la classification donnée par TA3 est la taille de l'entreprise tirée du REE au moment du tirage : c'est le nombre total d'emplois corrigé du nombre de propriétaires et d'apprentis. La taille de l'entreprise selon cette définition est redemandée sur le questionnaire et nommée UNTGROE. Le facteur d'extrapolation intra-entreprise INITGEW2 est basé sur UNTGROE : INITGEW2 = UNTGROE / ANZLOHN, où ANZLOHN est le nombre de salaires fournis par l'entreprise.
- Poids global initial : Définissons encore le poids d'extrapolation global initial d'un salaire : INGEW=INITGEW1·INITGEW2. Ce poids subira quelques modifications, dues en particulier aux strates avec non réponse totale.

2.2 Variations dans le cadre de sondage

Appelons GROE3 la même répartition en classes que TA3[6], mais basée sur UNTGROE. La Table 12 montre le nombre d'entreprises pour chaque combinaison de TA3 et GROE3. Cette table est calculée en utilisant le poids d'extrapolation INITGEW1. On constate que la table est approximativement symétrique, c'est-à-dire qu'il y a autant d'entreprises dans la combinaison (TA3,GROE3) $= (i, j)$ que (j, i). Le nombre d'entreprises hors diagonale représente environ 5.6% du total, dont 4.1% pour les passages de 1 à 2 ou de 2 à 1. Pour l'analyse, nous gardons la répartition en strates donnée par TA3 (pas de post-strates).

La variation est négligeable pour les deux autres classifications.

[6] Classes : 3-19, 20-49, 50+

TAB. 10 – Nombre de strates exhaustives ou non par classe d'activité

	plan régions		plan cantons			plan régions		plan cantons	
NOG_2	non	oui	non	oui	NOG_2	non	oui	non	oui
10	0	14	0	21	51	7	7	10	14
15	7	7	9	15	52	13	1	16	8
16	0	2	0	2	55	11	3	14	10
17	0	13	0	22	60	7	7	10	14
18	0	14	0	21	61	0	12	0	16
19	0	13	0	19	62	0	13	0	17
20	7	7	10	14	63	5	9	5	19
21	0	14	0	23	64	0	14	0	24
22	6	8	6	18	65	0	14	0	23
23	0	14	0	24	66	0	14	0	24
25	0	14	0	23	67	0	14	0	22
26	4	10	4	20	70	0	14	0	24
27	13	1	16	8	72	7	7	10	14
29	11	3	10	14	73	0	13	0	21
30	0	14	0	24	80	0	14	0	24
33	6	8	6	18	85	14	0	19	5
36	5	9	5	19	90	0	14	0	23
40	0	14	0	21	91	0	14	0	24
45	14	0	20	4	92	7	7	10	14
50	7	7	10	14	93	7	7	10	13

2.3 Non réponse totale dans une strate

Certaines combinaisons NOG_2×TA3×GRS n'existent pas, c'est-à-dire ne contiennent aucune entreprise. Ce sont des strates vides. D'autres ont de très petits effectifs. Lorsqu'on est en présence d'une non réponse totale dans une strate, le poids INITGEW1 est non défini. Comme il y a de nombreuses strates de petit effectif total, ce phénomème se produit plusieurs fois[7]. La technique couramment utilisée dans ce cas est le regroupement de strates "voisines". Comme nous nous intéressons en priorité aux combinaisons GRS×NOG_2, ou GR×NOG_2, nous définissons le voisinage d'une strate avec non réponse totale (strate non vide et non observée) par les strates de même GRS et de même NOG_2. Le programme vide02.sas amplifie les poids des strates voisines des strates non observées, de façon à tenir compte du nombre d'entreprises et du nombre d'emplois de ces strates. Appelons "Vois.tot" l'ensemble des strates non vides, voisines d'une ou plusieurs strates non observées (y compris cette ou ces strates), et "Vois.obs" le sous-ensemble des strates observées de "Vois.tot". Nous allons

[7]Il y a, en 2002, 36 strates non observées d'un effectif de 1 à 3 entreprises.

TAB. 11 – Taux de réponse en 2002

	TA3	1	2	3
taux de réponse		0.72	0.80	0.85

TAB. 12 – Classes de taille

	GROE3			
TA3	1	2	3	total
1	133697	2961	283	136940
2	3392	7790	1169	12351
3	144	752	5948	6843
total	137233	11502	7399	156134

amplifier les poids de "Vois.obs", de telle sorte que leur somme coïncide avec celle des poids initiaux de "Vois.tot". Posons :

- BURNH_S = nombre d'entreprises dans le REE pour la combinaison NOG_2×TA3×GRS
- BURMH_S = nombre d'emplois dans la strate : somme de UBETOT pour la combinaison NOG_2×TA3×GRS
- W1 = facteur correctif de INITGEW1 : $\sum_{h\in \text{Vois.tot}}$ BURNH_S / $\sum_{h\in \text{Vois.obs}}$ BURNH_S. Ce facteur varie en 2002 entre 1.01 et 2.00.
- W2 = facteur correctif de INGEW : $\sum_{h\in \text{Vois.tot}}$ BURMH_S / $\sum_{h\in \text{Vois.obs}}$ BURMH_S. Ce facteur varie en 2002 entre 1.01 et 5.55.

Remarque : Il arrive qu'une strate non observée soit isolée. Dans ce cas, "Vois.obs" est vide et la correction ci-dessus ne peut pas se faire. On étend alors la notion de voisinage à toutes les strates de GR×NOG_2. Une même strate peut alors appartenir à plusieurs voisinages de strates vides : Un voisinage du premier type - GRS×NOG_2 - pour les strates non observées de cette catégorie, et un voisinage du second type - GR×NOG_2 - pour des strates non observées de la même grande région GR et classe d'activité NOG_2, mais pas de la même subdivision GRS de cette grande région. Les facteurs correctifs finaux sont donc

- TOTW1 = facteur correctif de INITGEW1 : somme des W1 pour les deux types de voisinage d'une même strate observée. Le poids corrigé s'appelle FINGEW1.
- TOTW2 = facteur correctif de INGEW : somme des W2 pour les deux types de voisinage d'une même strate observée.

La plupart du temps, seul le premier type de voisinage est nécessaire. On dénombre en 2002 seulement 4 strates non observées isolées, totalisant 102 emplois répartis sur deux

entreprises dans la région 1, et 19 dans deux entreprises de la région 6. On peut donc dire que ces modifications ne biaisent pas les résultats sur le plan géographique.

La correction pour la non réponse totale concerne en tout, en 2002, 42 entreprises totalisant 1999 emplois. C'est donc un phénomène marginal.

2.4 Pondération finale

La pondération finale suit le même schéma qu'en 2000, aux corrections ci-dessus près.

- Le poids total GEW=TOTW2·INGEW est tronqué à 60. Le 99e percentile correspond à un poids de 15, et le poids initial maximum est de 132. Peu de poids sont donc touchés par cette correction. Par classe NOG_2, le 99e percentile aussi est toujours inférieur à 60, et le poids maximum ne dépasse cette valeur que dans les classes 45, 52, 55 et 85.
- Le poids intra INITGEW2 est tronqué à 20. Cette valeur est largement supérieure au 99e percentile qui vaut 6.5. La modification ne concerne que 122 salaires.
- Les poids GEW sont calés sur les marges données par les effectifs totaux par NOG_2 et par GR, en utilisant comme en 2000 la macro SAS CALMAR, avec l'option "méthode linéaire" (ajustement d'un modèle linéaire, spécifié par les variables dans les marges, avec contraintes). Les ajustements sont minimes, puisqu'ils varient entre un facteur de 0.999 à 1.05. Le poids global final prend le nom GEWICHT. Ce poids varie entre 1 et 62, avec un 99e percentile à 15.
- Le poids final intra entreprise est donné par FINGEW2=GEWICHT/FINGEW1.

Remarques :

1. Les poids finaux FINGEW1 et FINGEW2 ne sont pas utilisés pour les estimations. Leur intérêt réside dans la décomposition de GEWICHT en part intra et inter.
2. Le poids GEWICHT est appliqué aux estimations du salaire mensuel net non standardisé MBLOU[8]. Il peut être utilisé tel quel pour les estimations cantonales de GE, VD, FR, NE, SG et ZG. Le biais géographique introduit par les corrections des poids est pratiquement nul, et largement compensé par la meilleure stabilité des estimations résultantes.
3. Pour les estimations liées au salaire mensuel brut, standardisé MBLS[9], le poids tient compte du taux d'occupation IBGRS.

   Ainsi le poids utilisé est GEWIBGRS = GEWICHT·IBGRS. Ce poids est aussi applicable aux estimations cantonales de GE, VD, FR, NE, SG et ZG.

[8] Les salaires nets indiquent les montants effectivement versés par mois sans conversion en équivalents plein temps, cf [3] par exemple.

[9] Les montants relevés sont convertis en salaires mensuels standardisés, c'est-à-dire qu'ils sont recalculés sur la base d'un équivalent plein temps de 4 1/3 semaines à 40 heures de travail. Cette conversion permet d'établir des comparaisons entre les personnes occupées à plein temps et celles travaillant à temps partiel, cf. [3] par exemple.

ANNEXES

A Liste des programmes SAS

A1 Calcul des plans LSE 2002

Procédure	description	Output
initia.sas	création du fichier input de 2000 en vue du plan 2002	pla.priv00
ech02.sas	macro adaptation de ICmoy.sas : allocation de Neyman	
effectif2.sas	effectif CH par NOG_2 et TA3	pla.NCH
CHplan.sas	appel à ech02	pla.tail12
CHplan02b	création du plan A (15'000 entr.)	pla.ech15
	et du plan standard (45'000 entr.)	pla.ech45
effectif02.sas	effectifs de l'univers 2002 par grs	pla.Ngrs
grsplan02a.sas	plans B et C pour GR et GRS à partir du plan standard	
	copie de pla.ech15 (plan A)	pla.echCH02
	plan B	pla.echgr02
	plan C	pla.echgrs02
plan00.sas	Comparaison des plans 2000 et 2002 avec 15000 entr.	

A2 Pondération

Procédure	Description	Output
06b_Erstellung_ Input_Calmar_289.sas	séparation NE=28 et FR=29 et calcul des poids initiaux	privgtot02_289
_07_vorbereit_1.sas	calcul des 99e percentiles et stockage du privé	priv_289
vide02.sas	amplifie les poids des strates voisines des strates non observées	priv_289_neu
cal02_V3.sas	calage	priv02

B Entreprises du BUR avec UBETOT > 2

Kto2 = UTKTO = canton de la majorité des employés de l'entreprise.
Kto3 = canton du plus grand établissement de l'entreprise qui soit aussi de classe UNOGA.

Kto3 / Kto2	AG	AI	AR	BE	BL	BS	FR	GE	GL	GR	JU	LU	NE	NW	Total
AG	10579	0	0	0	0	0	0	0	0	0	0	1	1	0	**10584**
AI	0	352	0	0	0	0	0	0	0	0	0	0	0	0	**352**
AR	0	0	1099	0	0	0	0	0	0	0	0	0	0	0	**1099**
BE	0	0	0	20493	0	0	0	0	0	0	0	0	0	0	**20495**
BL	0	0	0	0	5210	1	0	0	0	0	0	0	0	0	**5212**
BS	0	0	0	0	4	3811	0	0	0	0	0	0	0	0	**3817**
FR	0	0	0	0	0	0	4638	0	0	0	0	0	0	0	**4638**
GE	0	0	0	2	1	0	0	8302	0	0	0	0	0	0	**8310**
GL	0	0	0	0	0	0	0	0	902	0	0	0	0	0	**902**
GR	0	0	0	0	0	0	0	0	0	5231	0	0	0	0	**5232**
JU	0	0	0	0	0	0	0	0	0	0	1523	0	0	0	**1523**
LU	0	0	0	0	0	0	0	0	0	0	0	7055	0	0	**7059**
NE	0	0	0	0	0	0	0	0	0	0	0	0	3468	0	**3469**
NW	0	0	0	0	0	0	0	0	0	0	0	0	0	848	**848**
OW	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**788**
SG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10037**
SH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1597**
SO	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**4875**
SZ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3002**
TG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4796**
TI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**7821**
UR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**677**
VD	0	0	0	0	0	0	0	2	0	0	0	0	0	0	**12515**
VS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6837**
ZG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3221**
ZH	1	0	0	2	0	0	0	0	0	0	0	0	0	0	**27639**
Total	**10581**	**352**	**1099**	**20497**	**5215**	**3812**	**4638**	**8304**	**902**	**5231**	**1523**	**7056**	**3469**	**848**	**157345**

Kto3 / Kto2	OW	SG	SH	SO	SZ	TG	TI	UR	VD	VS	ZG	ZH	Total
AG	0	0	0	0	0	0	0	0	0	0	1	2	**10584**
AI	0	0	0	0	0	0	0	0	0	0	0	0	**352**
AR	0	0	0	0	0	0	0	0	0	0	0	0	**1099**
BE	0	0	0	0	0	0	0	0	1	0	0	1	**20495**
BL	0	0	0	0	0	0	0	0	1	0	0	0	**5212**
BS	0	0	0	0	0	0	0	0	0	0	0	2	**3817**
FR	0	0	0	0	0	0	0	0	0	0	0	0	**4638**
GE	0	0	0	0	0	0	1	0	4	0	0	0	**8310**
GL	0	0	0	0	0	0	0	0	0	0	0	0	**902**
GR	0	0	0	0	0	0	0	0	0	0	0	1	**5232**
JU	0	0	0	0	0	0	0	0	0	0	0	0	**1523**
LU	0	2	0	0	0	0	0	1	0	0	0	1	**7059**
NE	0	0	0	0	0	0	0	0	1	0	0	0	**3469**
NW	0	0	0	0	0	0	0	0	0	0	0	0	**848**
OW	788	0	0	0	0	0	0	0	0	0	0	0	**788**
SG	0	10035	0	0	0	0	0	0	1	0	0	(b) 1	**10037**
SH	0	0	1597	0	0	0	0	0	0	0	0	0	**1597**
SO	0	0	0	4874	0	0	0	0	0	0	0	0	**4875**
SZ	0	1	0	0	3001	0	0	0	0	0	0	0	**3002**
TG	0	0	0	0	0	4795	0	0	0	0	0	1	**4796**
TI	0	0	0	0	0	0	7820	0	0	0	1	0	**7821**
UR	0	0	0	0	0	0	0	677	0	0	0	0	**677**
VD	0	0	0	0	0	0	0	0	12512	0	0	1	**12515**
VS	0	0	0	0	0	0	0	0	0	6837	0	0	**6837**
ZG	0	0	0	0	0	0	0	0	0	0	3221	0	**3221**
ZH	0	(a) 1	1	0	0	4	1	0	1	1	2	27625	**27639**
Total	**788**	**10039**	**1598**	**4874**	**3001**	**4799**	**7822**	**678**	**12521**	**6838**	**3225**	**27635**	**157345**

Exemple (a) Kto2=ZH : entreprise dont la majorité des employés travaillent dans le canton de ZH
Kto3=SG : mais dont le plus grand établissement de classe UNOGA se situe dans le canton de SG
(b) Kto2=SG : entreprise dont la majorité des employés travaillent dans le canton de SG
Kto3=ZH : mais dont le plus grand établissement de classe UNOGA se situe dans le canton de ZH

Références

[1] (2000). L'enquête suisse sur la structure des salaires 1998. Résultats commentés et tableaux. *Statistique de la Suisse*, OFS.

[2] (2002). L'enquête suisse sur la structure des salaires 2000. Résultats commentés et tableaux. *Statistique de la Suisse*, OFS.

[3] (2003). Enquête suisse sur la structure des salaires 2002 (ESS 2002). *Actualités OFS,* **3** Vie active et rémunération du travail.

[4] Graf, M. (2002). Enquête suisse sur la structure des salaires 2000. Plan d'échantillonnage, pondération et méthode d'estimation pour le secteur privé. *Rapport de méthode*, OFS. *http ://www.statistik.admin.ch/stat_ ch/ber00/meth/fmeth010_ 01.htm*

[5] Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse ; établissement du plan d'échantillonnage. *Rapport de méthodes*, OFS.

[6] Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer Series in Statistics.

Methodenberichte des Dienstes Statistische Methoden des BFS
Rapports de méthodes du Service de méthodes statistiques de l'OFS
Methodological reports of the Statistical Methods Unit of SFSO

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). Erhebung über Forschung und Entwicklung in der schweizerischen Privatwirtschaft 2000. Bereinigung der Stichprobe, Ersatz fehlender Werte und Schätzverfahren. 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 élèves (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>>News>>Neuerscheinungen.

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>>Nouvelles publications.

Methodenberichte des Dienstes Statistische Methoden
Rapports de méthodes du Service de méthodes statistiques
Methodological reports of 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.

L'enquête sur la structure des salaires est une enquête bisannuelle auprès des entreprises dont le but principal est d'estimer des salaires médians pour un certain nombre de domaines (classe d'activité, niveau de qualification, sexe ...). Le présent rapport décrit les aspects méthodologiques du plan d'échantillonnage et le calcul des poids d'extrapolation pour l'enquête suisse sur la structure des salaires 2002.

Le plan d'échantillonnage est décrit en détail. Il prévoit un échantillon stratifié en grappes à deux niveaux, comme dans les enquêtes précédentes. La différence avec le passé apparaît au niveau de la définition des strates. La stratégie est de garantir une précision analogue pour l'estimation du salaire standardisé moyen au niveau grande région, puis de faire les adaptations nécessaires pour les cantons désirant une enquête spécifique.

La pondération reprend pour l'essentiel les poids d'extrapolation de Horvitz-Thompson. La non-réponse étant supposée ignorable dans les strates, les poids sont donc basés sur l'effectif par strate des entreprises répondantes. Les poids jugés trop grands sont alors tronqués et le déficit est distribué sur les autres classes de taille de même région et classe d'activité économique.

N° de commande: 338-0025

Commandes:
032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch

Prix: gratuit

ISBN 3-303-00296-7