Rapport de méthodes

Statistique de l'emploi

Révision 2007: méthodes d'estimation



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: méthodes d'estimation

Auteurs **Anne Renaud, Christian Panchard et Jann Potterat**
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 les travaux d'estimation effectués dans ce projet. Il fait suite au rapport de méthodes «Statistique de l'emploi. Révision 2007: cadre de sondage et échantillonnage» (numéro de commande: 338-0052).

Un grand merci à Francis Saucy et Nathalie Nünlist (KE) pour les informations et la collaboration durant le projet. Merci à Mathieu Gunzinger de la section Registre des entreprises et des établissements (BUR) pour sa contribution dans le reporting et les informations sur le registre. Merci également à Paul-André Salamin (METH) pour les discussions méthodologiques au cours du projet et la relecture attentive 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. Les estimateurs des totaux ont également été revus. Ils utilisent une procédure de détection de valeurs aberrantes et un calage direct sur des valeurs de référence connues au niveau du cadre sans passer par des modèles de non-réponse. Une estimation de la variance est également proposée afin de donner des informations sur la précision des estimateurs. Le rapport présente finalement une évaluation des nouvelles méthodes sur les trois premiers trimestres d'application, c'est-à-dire les 2ème, 3ème et 4ème trimestres 2007.

Mots-clé

rapport de méthodes; estimateurs; estimation; variance; calage; robustification; 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: 338-0055

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: 978-3-303-00413-5

Table des matières

1 Introduction **5**

1.1 Bases et définitions 5

1.2 Cadre de sondage et plan d'échantillonnage 7

2 Estimations **10**

2.1 Estimations depuis 2003 10

2.2 Révision 2007 11

3 Cadre d'extrapolation **11**

4 Préparation des données d'enquête **13**

4.1 Multi-établissements 13

4.2 Profiling 13

4.3 Très grands établissements 14

5 Pondération et estimation des totaux **15**

5.1 Robustification 16

5.2 Calage 17

6 Estimation de la variance et du CV **19**

6.1 Estimateur GREG sans non-réponse 19

6.2 Intégration de la non-réponse dans l'estimation 20

6.3 Quelques particularités de la STATEM 21

7 Résultats **22**

7.1 Préparation du cadre d'extrapolation 22

7.2 Préparation des données d'enquête 23

7.3 Pondération 23

7.4 Estimations 26

8 Un peu plus sur les estimations **27**

9 Conclusion **29**

Annexes **30**

A Définitions **30**

B Construction du cadre d'extrapolation **34**

C Estimations **37**

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 entre dans le cadre des travaux préparatoire du programme général de statistiques des entreprises (GUS, "Gesamtprogramm Unternehmensstatistik"). GUS vise à développer un système cohérent et intégré d'enquêtes auprès des entreprises. Ses objectifs sont de produire des informations et résultats mieux adaptés aux besoins des clients, de moins solliciter les fournisseurs de données et de répondre à des exigences nouvelles, notamment au niveau international. On vise donc à mieux coordonner les enquêtes, exploiter les synergies, utiliser davantage les données administratives et de registres, utiliser de nouvelles technologies et méthodes statistiques et davantage standardiser les processus de production.

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'estimation appliquées depuis la dernière révision de 2003. 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 méthodes d'estimation des totaux d'intérêt et de leur variabilité. Il fait suite au rapport Renaud (2008) qui documente les bases de la révision, de la construction du cadre et du plan d'échantillonnage.

1.1 Bases et définitions

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 ; voir Annexe A pour les définitions.

Le cadre d'échantillonnage est un extrait du Registre des entreprises et des établissements (REE).

Les paramètres d'intérêt sont des totaux pour l'ensemble des établissements de la population d'intérêt et les domaines suivants :
– 49 activités économiques (NOGA2 avec les regroupements 10-14, 36-37 et 40-41) ;
– 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 croisées avec les deux secteur 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 une combinaison linéaire des variables TOTVZ, TOTTZ1 et TOTTZ2 ; voir l'Annexe A pour la définition.

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 :

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.

1.2 Cadre de sondage et plan d'échantillonnage

Cette section résume les principes de base du cadre de sondage et de l'échantillonnage de la STATEM révisée. Pour plus de détails, voir Renaud (2008).

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 une partie des normaux et l'ensemble des établissements spéciaux du cadre.

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 et le cadre de sondage est utilisé comme cadre d'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). Les adaptations visent le bon déroulement du relevé durant l'année entière (envoi à des établissements réels, codification des cas disparus). Les autres modifications ne sont pas prise en compte en cours d'année afin d'assurer une plus grande indépendance entre la STATEM et le REE, notamment éviter d'être influencé par les travaux de maintenance du registre.

La stratification est définie par le croisement entre les quatre variables suivantes ; voir l'Annexe A pour les définitions : Type d'établissement (`special`), NOGA2 avec regroupements (`NOGA2reg`), régions avec objectifs complémentaires et restes des grandes régions (`region_new`), et classe de taille basée sur les équivalents pleins temps du cadre `equivpt` (`tai2`). Cette stratification très détaillée est utilisée pour le tirage de l'échantillon. On verra plus bas que des regroupements sont effectués pour les estimations.

L'allocation part des objectifs de précision exprimés en coefficients de variation, voir Section 1.1. Les résultats provenant des différents objectifs sont alors combinés au niveau des strates de tirage avant de faire des ajustements pour tenir compte de la non-réponse et d'une taille minimale par strate.

Le tirage utilise l'idée des nombres aléatoires permanents définis pour chaque établissement du cadre. On sélectionne un échantillon de taille fixe dans chaque strate de tirage avec un algorithme assurant un recouvrement entre échantillons successifs. Le tirage de 2007 est coordonné avec celui de 2003. Le renouvellement du 2ème trimestre 2008 sera coordonné avec celui du 2ème trimestre 2007. Et ainsi de suite, avec un rythme annuel.

Au 2ème trimestre 2007, les établissements normaux correspondent à 94.5% des établissements et 81% des emplois du cadre ; voir Table 1. L'échantillon final contient 62'833 établissements, dont 41'553 normaux et 21'280 spéciaux. Environ 30% des établissements de l'échantillon sont dans le profiling et 66% sont normaux. Parmi tous les emplois relevés dans l'échantillon, environ 25% seront relevés dans le profiling et 68% dans les normaux.

TAB. 1 Information sur la population et l'échantillon STATEM en fonction du type d'établissement. État au 2ème trimestre 2007. 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. Les emplois betot et betotE sont donnés en milliers.

Type	sflgprof	Npop	%Npop	betot	%betot	nech	%nech	betotE	%betotE
Normal	0	363'818	94.5	3'011	81.0	41'553	66.1	1'491	67.8
Prof. privé	1	11'370	3.0	299	8.0	11'370	18.1	299	13.6
Prof. public	2	7'323	1.9	254	6.8	7'323	11.7	254	11.5
Permanent	3	317	0.1	24	0.6	317	0.5	24	1.1
Multi	4	2'270	0.6	130	3.5	2'270	3.6	130	5.9
Total		385'098	100.0	3'718	100.0	62'833	100.0	2'198	100.0

Une partie de l'échantillon est identifiée comme le groupe des très grands établissements, voir `IDVB`=1 dans la Table 2. Ces 4200 établissements sont spécialement suivis durant l'enquête afin de limiter la non-réponse des grands établissements.

Les très grands établissements sont identifiés dans le cadre de sondage par activité économique `NOGA2reg`. La limite `limVB` entre "pas très grand" et "très grand" ("very big") est définie sur la base du 99ème percentile des équivalent plein temps selon le cadre `equivpt`. Elle est restreinte à l'intervalle allant de 50 à 100 équivalents plein temps [1]. On définit également l'identificateur `IDVB` des très grands ou "very big"[2]. Les variables `limVB` et `IDVB` sont fournies dans le cadre.

[1] Définition de `limVB` par `NOGA2reg` : `if P99<50 then limVB=50 ; else limVB=round(min(P99,100),0.1) ;`

[2] Identificateur des "very big" : `IDVB=0 ; if equivpt>=limVB then IDVB=1 ;`

TAB. 2 Répartition de l'échantillon dans les NOGA2reg, les tailles tai2 et les très grands "very big" IDVB. Information également sur les NOGA2reg exhaustives (exnoga=1) et de la limite des très grands limVB. Notons que seuls les très grands permanents et multi sont comptabilisés dans tai2=99 et IDVB=1. Le profiling est traité différemment.

			tai2							
			1	2	3	4		99		
			IDVB	IDVB	IDVB	IDVB		IDVB		
			0	0	0	0	1	0	1	All
NOGA-2reg	exno-ga	limVB								
10.14	1	94.2	171	147	61	7	3	5	.	394
15	0	100	236	222	119	78	108	124	17	904
16	1	100	3	3	3	1	2	.	3	15
17	0	100	186	147	72	30	21	3	.	459
18	0	54.7	289	81	19	2	9	.	.	400
19	1	100	159	30	12	4	3	.	.	208
20	0	50	269	310	246	.	50	11	2	888
21	1	100	71	60	34	20	38	.	.	223
22	0	100	235	261	232	75	58	16	4	881
23	1	100	4	5	2	.	3	.	.	14
24	0	100	201	155	139	78	93	22	4	692
25	0	100	166	159	112	76	51	4	.	568
26	0	100	231	171	119	52	23	18	3	617
27	1	100	92	84	34	28	43	.	.	281
28	0	100	279	245	192	169	106	18	6	1015
29	0	100	255	219	163	190	200	14	7	1048
30	1	100	79	33	16	4	4	.	.	136
31	0	100	244	145	114	49	63	16	.	631
32	0	100	202	151	59	35	40	7	.	494
33	0	100	251	288	183	130	164	13	15	1044
34	1	100	92	54	23	9	9	.	.	187
35	1	100	296	97	17	8	17	12	4	451
36.37	0	88.7	273	387	173	47	35	3	4	922
40.41	0	100	181	185	135	40	32	78	11	662
45	0	76.3	778	379	242	365	322	369	39	2494
50	0	50	408	275	428	.	107	60	2	1280
51	0	89	522	264	234	285	198	194	16	1713
52	0	50	1156	331	813	.	253	2123	3	4679
55	0	50	613	329	228	.	251	374	7	1802
60	0	100	289	254	398	111	49	966	8	2075
61	1	100	51	33	25	5	4	.	.	118
62	1	100	112	84	17	6	4	5	4	232
63	0	100	248	204	209	76	62	120	6	925
64	0	100	465	161	47	24	19	4848	.	5564
65	0	100	234	265	250	115	99	1216	34	2213
66	0	100	433	307	134	39	38	1043	11	2005
67	0	50	328	240	154	.	23	496	7	1248
70	0	50	389	223	98	.	34	105	8	857
71	0	50	433	139	17	.	5	4	.	598
72	0	60.9	378	241	319	42	114	53	1	1148
73	1	100	421	119	40	16	22	40	7	665
74	0	50	1636	317	1051	.	514	263	19	3800
75	0	100	228	382	317	61	26	3585	16	4615
80	0	100	423	402	267	148	76	3145	24	4485
85	0	100	592	263	253	612	344	1279	65	3408
90	0	72.9	255	443	128	12	12	126	1	977
91	0	57.6	351	325	155	7	54	27	1	920
92	0	68.6	352	281	164	21	48	340	9	1215
93	0	50	532	285	60	.	19	60	1	957
All			16092	10185	8327	3077	3872	21205	369	63127

Définition :

tai2=1 : normal avec equivpt$<$5

tai2=2 : normal avec 5$\leq$equivpt$<$20

tai2=3 : normal avec 20$\leq$equivpt$<$50

tai2=4 : normal avec 50$\leq$equivpt

tai2=99 : spécial.

2 Estimations

Cette section présente brièvement la méthode d'estimation appliquée depuis 2003 et les nouveautés de la révision 2007. La méthode appliquée pour la révision 2007 est documentée plus en détails dans les Sections 3 à 6.

2.1 Estimations depuis 2003

Le total de l'emploi est estimé depuis la révision 2003 à l'aide d'un estimateur par le quotient séparé dans des cellules très détaillées. La variable auxiliaire x est le nombre d'équivalents plein temps dans le cadre.

L'estimateur du total Y pour la variable y peut être écrit sous la forme $\widehat{Y} = \sum_r y_k w_k$, avec r l'ensemble des répondants et w_k le poids final ou poids d'extrapolation de l'établissement $k \in r$.

Le poids final w_k de l'établissement répondant k dans la cellule c est défini par :

$$w_k = \frac{N_c}{m_c} \frac{X_c}{\sum_{r_c} \frac{N_c}{m_c} x_k} = \frac{X_c}{\sum_{r_c} x_k}, \tag{1}$$

avec N_c le nombre total d'établissements du cadre dans la cellule c, m_c le nombre de répondants dans c, X_c le total des équivalents plein-temps du cadre dans c, r_c l'ensemble des répondants dans c et $\widehat{X}_c = \sum_{r_c} \frac{N_c}{m_c} x_k$ l'estimateur du total X_c sur la base des répondants r_c et du poids $b_k = N_c/m_c$.

On notera que le poids b_k correspond à un poids d'échantillonnage ajusté pour la non-réponse. Il permet l'extrapolation de r à U. Le poids d'échantillonnage d_k est relativement complexe dans la STATEM entre 2003 et 2007 car il dépend de la procédure en deux étapes : tirage national puis tirages régionaux dans l'ensemble des établissements restants. Il ne correspond donc pas à un rapport du type N_h/n_h comme c'est le cas dans un plan stratifié habituel. Connaissant le poids b_k ajusté par la non-réponse et le poids d_k d'échantillonnage, on détermine l'ajustement non-réponse $(N_c/m_c)/d_k$. Il s'agit donc d'un ajustement selon un modèle différent des cellules homogènes ou d'un modèle logistique. Il n'est pas explicité dans les estimations.

Les cellules c sont les croisements des deux types d'établissements ("fictif[3] + permanent" et "normal") avec les activités économiques `NOGA2reg`, les régions et des classes de taille déterminées à partir des classes du plan national et des plans régionaux (variable : `newclt`). Elles sont très nombreuses et souvent petites.

Les tailles N_c et les totaux de référence X_c sont déterminés sur le REE en cours dans l'idée de prendre en compte l'évolution de l'emploi dans le cadre. Ces valeurs ont cependant le désavantage de dépendre de variations dues à la maintenance en plus de celles de l'emploi.

Les données des multi-établissements "fictifs" sont assignées en bloc à la région et à l'activité économique de l'établissement de contact. Elles ne sont pas distribuées dans les établissements individuels. Le traitement en bloc fausse les répartitions régionales et économiques. On profite de la révision 2007 pour corriger la procédure et améliorer les estimations.

[3] Note : les multi-établissements sont appelés établissements fictifs avant la révision 2007.

2.2 Révision 2007

Pour la révision 2007, nous revoyons totalement la procédure d'estimation afin de tenir compte du nouveau plan d'échantillonnage et des difficultés observées dans la production entre 2003 et 2007.

Les données des multi-établissements sont réparties dans les établissements individuels de façon à avoir une meilleure distribution dans les régions et les activités économiques. De plus, des imputations sont mises en place pour traiter les valeurs manquantes du profiling et une partie des très grands établissements ; voir Section 4.

Des poids finaux sont définis au niveau de chaque établissement. Ils prennent en compte les poids d'échantillonnage, une robustification de valeurs très influentes et un ajustement pour tenir compte de la non-réponse et des informations auxiliaires. Ces poids sont utilisés pour estimer les totaux pour les différentes variables et les différents domaines d'intérêt ; voir Section 5.

La détection systématique des valeurs aberrantes et la robustification sont des processus nouveaux dans la STATEM. Ils permettent de détecter et corriger des erreurs manifestes dans les données puis de traiter les cas extrêmes restant sans modification des valeurs relevées. Seuls les poids sont potentiellement modifiés dans la robustification. Ces outils permettent de détecter des erreurs ayant échappé au traitement standard des données et surtout de réduire l'influence d'établissements atypiques. Dans le cas d'un établissement très petit lors du tirage et qui devient brusquement très grand, la procédure permet par exemple de réduire le poids sans modifier les valeurs relevées. On considère donc que cet établissement a fourni des données valables mais qu'il représente un nombre d'établissements dans la population inférieur à ce que son poids d'origine indiquait.

L'ajustement des poids pour la non-réponse et les 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 successive de la non-réponse puis d'un estimateur par le quotient comme fait entre 2003 et 2007. Cette approche calage, décrite dans Särndal et Lunström (2005), permet d'intégrer toutes les informations auxiliaires dans une unique procédure. On parlera ci-dessous de *calage direct*.

La robustification et le calage utilisent la variable auxiliaire `equivpt`, c'est-à-dire les équivalents plein temps au temps $T0$ (trimestre tirage).

Une estimation de la variance du total estimé est également proposée. Elle reprend les informations sur le plan d'échantillonnage et les corrections des poids dans une estimation du type GREG pour tenir compte du calage ; voir Section 6.

3 Cadre d'extrapolation

Dans le nouveau système révisé de la STATEM, un cadre de sondage est défini une fois par année ($T0$). Le cadre et l'échantillon tiré au trimestre $T0$ sont conservés pendant un an.

Au trimestre T, on définit un cadre d'extrapolation qui sert de base à la pondération de l'échantillon au temps T. Ce cadre contient les variables de base du cadre de sondage, l'identification de l'échantillon brut, l'identification des différents types d'établissements, et, pour chaque élément, la variable auxiliaire x (`equivpt`) utilisée dans la pondération.

Le cadre d'extrapolation T est défini à partir du cadre $T0$. L'idée est de garder au maximum les

éléments du cadre $T0$ et de faire des modifications uniquement pour les mutations ayant une influence sur le relevé. Le cadre T est donc très proche du cadre $T0$ tout en permettant le lien entre le cadre et les données au trimestre T. Les établissements qui ferment avec successeurs sont remplacés dans la liste par le successeur (le poids est repris également). Ceux qui ferment sans successeurs sont conservés avec un code indiquant qu'il n'est pas utile d'envoyer un questionnaire. Les établissements qui fusionnent sont conservés selon leur nouvelle structure (avec un poids adapté pour tenir compte de la fusion ; voir le partage des poids dans Renaud, 2008). On actualise les informations individuelles, notamment la variable auxiliaire `equivpt`, uniquement en cas de fusion.

Durant une année, il n'y a pas d'ajout de nouveaux établissements ni de retrait d'établissements disparus sans successeurs. On vise à rester le plus proche possible du cadre $T0$, et donc autant que possible indépendant des variations du REE durant une année. Il s'agit d'une optique conjoncturelle où l'on observe l'évolution de l'emploi à partir de l'échantillon, et non pas un mélange conjoncturel-structurel comme dans le cas d'un relevé combiné avec une adaptation continue du cadre de référence pour l'extrapolation. On peut également voir la procédure reposant sur une hypothèse de travail qui suppose que les disparitions et les apparitions se compensent durant une année[4].

L'application nommée *reporting* compare le REE aux temps $T0$ et T. Elle identifie les mutations suivantes : fermeture sans successeur, fermeture avec successeur, naissance sans prédécesseur, naissance avec prédécesseur(s) et changement d'entreprise. L'application permet également d'identifier les modifications du type d'établissements, notamment les entrées et sorties du profiling. Un fichier *Rahmen* T est créé à partir du cadre $T0$ et du reporting entre $T0$ et T. Il comporte la liste des établissements aux temps $T0$ et T, ainsi que les informations sur les mutations, les successeurs et les prédécesseurs.

Le cadre d'extrapolation T est un extrait de Rahmen T. On retire les nouveaux sans prédécesseurs, les inactifs avec successeurs vu que les successeurs sont dans la liste, et les inactifs du profiling. Cette exclusion du profiling inactif est due au traitement spécial du profiling. En effet, les établissements inactifs sont retirés en bloc et les établissements actifs sont repris en bloc car les mutations du profiling ne sont pas documentées dans le REE ; voir Section 8.

L'échantillon est également identifié dans le cadre d'extrapolation T. Il s'agit de l'échantillon au temps $T0$ avec quelques adaptations dues aux mutations telles les fermetures avec successeurs et les changements dans le profiling.

Le cadre d'extrapolation utilisé pour les estimations est défini pour la date de référence de l'enquête (dernier jour du trimestre de relevé). Un cadre intermédiaire peut cependant également être construit avant l'envoi des questionnaires pour le trimestre T. Ceci pour définir des listes d'envois actualisées (disparitions, successeurs et fusions).

Pour quelques informations complémentaires et plus techniques, voir Annexe B.

[4] Les disparitions et apparitions durant l'année sont suivies et documentées dans le nouveau système dénommé reporting. Une intégration d'une partie des disparitions et apparitions sera évaluée prochainement dans les estimations.

4 Préparation des données d'enquête

Trois traitements sont appliqués aux données d'enquête avant de passer aux estimations :

1. Répartition des emplois relevés pour les multi-établissements.
2. Imputation de données pour le profiling sans données.
3. Imputation d'une partie des très grands établissements sans données.

Ces opérations sont effectuées chaque trimestre T. Elles sont suivies par l'introduction des données dans le cadre d'extrapolation T. Ce dernier contient donc toutes les informations pour les estimations (relevé, imputations, variables auxiliaires).

Note à propos de la production : Au temps T, la liste des établissements de l'échantillon brut dans le système de production STATEM doit être identique à celle dans le cadre d'extrapolation. De plus, les répondants STATEM au trimestre T doivent tous se trouver dans l'échantillon identifié dans le cadre T. Les modifications de la structure des multi-établissements doivent également être suivies entre les trimestres. En cas de divergence, les cas sont étudiés individuellement.

4.1 Multi-établissements

Le relevé d'un multi-établissement correspond au relevé d'un agrégat. On a le total de l'emploi et l'information sur la liste des établissements contenus dans le multi-établissements mais pas le détail de l'emploi pour chaque établissement.

Chaque trimestre, les données des multi-établissements répondants sont redistribués dans les établissements individuels selon une clé correspondant à la distribution des équivalents plein temps dans le cadre (`equivpt`). La même clé est utilisée pour toutes les variables de relevé (VZM, VZF, TZ1M, TZ1F, TZ2M, TZ2F, TOTVZ, TOTTZ1, TOTTZ2, TOT). Cette procédure permet de répartir les emplois entre les régions et les activités économiques.

Seuls les établissements actifs sont considérés dans la répartition. Les établissements inactifs ne reçoivent pas d'emplois. Ils sont traités comme les établissements normaux non répondants, actifs ou inactifs, *i.e.* dans l'idée du cadre conservé pendant une année.

La répartition des multi-établissements est considérée comme une imputation au niveau individuel. Ces cas sont identifiés par la variable `imputMul`=1.

4.2 Profiling

Les données des établissements dans les groupes profiling sont relevées par la section Registre des entreprises et des établissements (BUR). Elles sont ensuite transmises à la section KE pour les besoins de la STATEM. L'idée du profiling est d'avoir un contact particulier avec un ensemble de grands groupes d'entreprises. Il est prévu que le relevé soit trimestriel mais il y a en pratique des décalages ou des valeurs manquantes. Les décalages signifient que les données disponibles ne correspondent pas au trimestre en cours mais à un trimestre antérieur. Les valeurs manquantes sont souvent observées pour des groupes entiers d'entreprises. Il y a aussi des groupes avec des informations partielles, notamment en cas de restructuration. De plus, certains groupes prévus pour le profiling au moment du plan d'échantillonnage n'ont pas encore intégrés la procédure.

Une imputation des valeurs manquantes est effectuée au niveau établissement afin de compléter tous les cas manquants au trimestre T. Cette procédure a été choisie pour éviter un traitement global non-réponse sur ces établissements avec un relevé particulier. L'actualité des données n'est pas toujours parfaite mais, sans informations complémentaires, on choisit de prendre les dernières valeurs connues, sans modification ou tentative de prendre en compte une possible évolution. Les données les plus anciennes datent du dernier recensement des entreprises et des établissements d'automne 2005.

Les données imputées dans le profiling sont identifiées par la variable `imputProf`=1.

4.3 Très grands établissements

La non-réponse de très grands établissements "very big" est un élément difficile à traiter dans les estimations. Les ajustements des poids dans des cellules homogènes corrigent en effet mal les effets de cette non-réponse.

Dans la STATEM, nous choisissons de faire une imputation sur la base du trimestre précédent. Cette imputation est appliquée uniquement pour une partie des très grands établissements hors profiling. La procédure est plus conservatrice que celle appliquée pour le profiling car cette population est moins suivie durant l'enquête.

Les valeurs de l'emploi sont imputées pour les établissements ayant l'ensemble des caractéristiques suivantes (`IDimp`=1) :

1. valeurs manquantes pour le trimestre en cours ;
2. "very big" et hors profiling pour le trimestre en cours ;
3. valeur de l'emploi disponible pour le trimestre précédent (relevé ou imputé) ;
4. pas de modification de structure touchant l'emploi selon le reporting[5] ;
5. actif dans le trimestre en cours.

L'imputation se base sur la valeur du trimestre précédent, avec une évolution moyenne déterminée sur la base d'un groupe d'établissements choisi comme suit dans chaque `NOGA2reg` (`IDsac`=1) :

1. very big pour le trimestre en cours ;
2. emploi disponible pour le trimestre en cours et le trimestre précédent (relevé pas imputé) ;
3. pas de modification de structure touchant l'emploi selon le reporting (voir plus haut) ;
4. actif dans le trimestre en cours.

L'évolution moyenne est calculée sur la base des équivalents plein temps EPT relevés. Elle est limitée à l'intervalle [-10%, +10%] dans chaque `NOGA2reg`. Le facteur d'évolution est appliqué aux établissements imputables (`IDimp`=1) pour toutes les valeurs de l'emploi.

L'imputation $\widehat{y}_{k,t}$ est effectuée à partir du relevé $y_{k,t-1}$ au temps $t-1$ et d'un facteur d'évolution $\rho_{t-1,t}$ défini au niveau de la `NOGA2reg` :

$$\widehat{y}_{k,t} = y_{k,t-1}\, \rho_{t-1,t}, \tag{2}$$

[5] Identification des mutations ayant un effet sur l'emploi : `code_2 in (13,14)` pour les cas devenus inactifs et `(code_2 in (5,6) and suc_anzvorg_2>1)` pour les établissements avec ≥ 2 prédécesseurs (fusions).

avec

$$\rho_{t-1,t} = \begin{cases} -0.1 & \text{si } \phi_{t-1,t} < -0.1 \\ \phi_{t-1,t} & \text{si } -0.1 \leq \phi_{t-1,t} \leq 0.1 \\ 0.1 & \text{si } \phi_{t-1,t} > 0.1. \end{cases}$$

Le facteur $\phi_{t-1,t}$ est estimé avec les établissements de référence (sac, `IDsac=1`) :

$$\phi_{t-1,t} = \frac{\sum_{k \in sac} y_{k,t}}{\sum_{k \in sac} y_{k,t-1}} - 1. \tag{3}$$

La méthode proposée est assez restrictive. Une imputation a en effet lieu seulement si l'on pense avoir de bonnes bases pour le faire. Les établissements restants sans réponse sont traités comme les plus petits (calage).

5 Pondération et estimation des totaux

Le total Y est estimé sur la base de l'ensemble r des répondants (valeurs relevées ou imputées). L'extrapolation a lieu à l'aide de poids finaux notés w_k pour $k \in r$.

L'estimateur $\widehat{Y}_d$ du total Y_d dans le domaine d est donné par la somme pondérée :

$$\widehat{Y}_d = \sum_{k \in r_d} w_k \, y_k \tag{4}$$

avec r_d l'ensemble des répondants dans d, y_k la valeur de la variable y et w_k le poids d'extrapolation de l'établissement k.

L'estimateur $\widehat{Y}_d$ peut également être exprimé par la somme sur tous les répondants :

$$\widehat{Y}_d = \sum_{k \in r} w_k \, y_k \, I_{kd},$$

avec

$$I_{kd} = \begin{cases} 1 & \text{si } k \in d \\ 0 & \text{sinon.} \end{cases} \tag{5}$$

Le poids final w_k (`poidsFin`) de l'établissement $k \in r$ a la forme :

$$w_k = d_k \, u_k \, v_k. \tag{6}$$

Il est construit sur la base du poids d'échantillonnage d_k (`poidsTir`), puis ajusté par le facteur u_k (`u_res_def`) durant la robustification des valeurs extrêmes, et finalement encore ajusté par le facteur v_k (`corr_cal`) durant le calage sur des totaux de référence de la variable auxiliaire x.

Chaque établissement de l'échantillon a un poids de sondage qui dépend du plan d'échantillonnage. Ainsi, l'établissement k de la strate h a le poids d'échantillonnage $d_k = N_h/n_h$ avec N_h le nombre d'établissements dans h selon le cadre de sondage $T0$ et n_h le nombre d'établissements tirés dans h. Le poids d_k (`poidsTir`) est disponible dans le cadre d'extrapolation pour tous les éléments de l'échantillon (`IDech=1`).

Aucun ajustement du type robustification ou calage n'est effectué sur les données du profiling ($u_k = v_k = 1$). Chaque établissement se représente lui-même[6] et les données sont considérées

[6] Quelques établissements du profiling ont des poids inférieurs ou supérieurs à 1. Il s'agit d'établissements tirés hors profiling et entré ensuite dans le profiling, ou de cas de fusions avec partage des poids.

de bonne qualité (contacts réguliers entre l'OFS et les groupes profiling). De plus, il n'y a pas de non-réponse (relevé ou imputation). On décrira donc ci-dessous les procédures de robustification et de calage uniquement pour les établissements hors profiling.

Les résultats numériques sont présentés dans la Section 7.3.

5.1 Robustification

La détection de valeurs aberrantes et la robustification sont appliquées sur les données de l'échantillon des répondants hors profiling.

La variable auxiliaire x (`equivpt`) est utilisée pour détecter d'éventuelles valeurs extrêmes ou *outliers* de la variable y, *i.e.* établissements avec une relation très atypique entre y et x. La variable principale de l'enquête y =TOT est choisie pour la détection des outliers et la robustification de la STATEM.

Seuls les établissements avec un poids d_k supérieur à 1 sont considérés dans la procédure. On ne modifie pas les poids des établissements avec un poids égal ou inférieur à 1, ni celui des "very big". Ces établissements se représentent en effet déjà eux-même dans la population. On leur assigne la valeur u_k=1 (pas de modification du poids). La détection est donc restreinte à de petits et moyens établissements d'après la classification du cadre (variable x).

La détection est appliquée de façon itérative dans chaque `NOGA2reg`. Les établissements détectés sont analysés par la section de production. En cas d'erreur, les valeurs y_k sont corrigées. Si aucune erreur n'est avérée, la valeur y_k est conservée telle quelle. La procédure est réappliquée tant que des nouvelles unités sont détectées. Une fois que les cas détectés ont tous été contrôlés, on applique une dernière fois la procédure afin de déterminer les ajustements $u_k \leq 1$ (`u_res_def`). La méthode consiste à réduire les poids des établissements considérés comme outliers sans modifier la valeur relevée.

Le niveau `NOGA2reg` est choisi après évaluation de différents sous-groupes tels que sections croisés avec les grandes régions. La corrélation entre y et x est en effet bonne à ce niveau.

La procédure suivante est appliquée dans chaque `NOGA2reg` (Hulliger, 1999) :

1. Méthode robuste pour estimer β dans le modèle $y = \beta x + \epsilon$ (PROC ROBUSTREG du logiciel SAS, sans ordonnée à l'origine, sans poids)
2. Détermination des résidus du modèle : $r_k = y_k - \beta x_k = y_k - \widehat{y}_k$.
3. Résidus normalisés : $\tilde{r}_k = |r_k|/\sqrt{x_k}$
4. Médiane des résidus normalisés : $\texttt{med}(\tilde{r}_k)$
5. Détermination du seuil : $\Omega = c \cdot \texttt{med}(\tilde{r}_k)/0.67$
6. Détermination des bornes de détection (pour les graphiques) :
   borne_sup $= \widehat{y}_k + \Omega\sqrt{x_k}$ et borne_inf $= \widehat{y}_k - \Omega\sqrt{x_k}$.
7. Ajustement intermédiaire :

$$u_{k,prov} = \begin{cases} \Omega/\tilde{r}_k & \text{si } \tilde{r}_k > \Omega \\ 1 & \text{sinon} \end{cases}$$

8. Ajustement final : $u_k = \max(\text{lim}, u_{k,prov})$.

Les constantes $c = 7$ et lim $= 0.2$ ont été choisies après divers essais, l'observation des graphiques et la comparaison des poids et totaux dans les différents domaines d'intérêts. On

vise un traitement plutôt conservateur, *i.e.* notamment une moyenne des u_k supérieure à 0.99. Les mêmes constantes sont appliquées dans toutes les `NOGA2reg`.

Notons que les variables x et y ont des valeurs strictement positives. Tous les établissements sont donc considérés dans la procédure, notamment dans le modèle $y = \beta x + \epsilon$.

Les graphiques permettent de visualiser les cas détectés et de procéder aux contrôles avant de robustifier les poids ; voir figure 1. Plus une unité est éloignée du domaine de détection plus elle sera robustifiée, c'est-à-dire plus son poids sera réduit. La réduction est cependant limitée au facteur lim $= 0.2$.

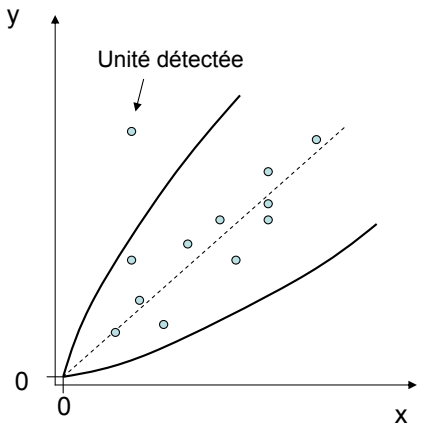


FIG. 1 Illustration de la méthode de détection de valeurs extrêmes.

5.2 Calage

Le calage direct intègre des informations auxiliaires dans les estimations afin d'ajuster pour la non-réponse et respecter les totaux de référence. Dans le cas de la STATEM, la variable auxiliaire est le nombre d'équivalents plein temps selon le cadre x (`equivpt`) et les valeurs de référence sont les totaux X_g^* de la variable x dans différents groupes $g = 1, .., G$ de la population U. Comme le calage est effectué à partir des établissements répondants hors profiling, on détermine les totaux de référence aussi hors profiling. Ces totaux sont calculés sur la base du cadre d'extrapolation.

Soit le vecteur $\gamma_k = (\gamma_{1k}, .., \gamma_{gk}, .., \gamma_{Gk})$, avec $\gamma_{gk} = 1$ si $k \in g$ et 0 sinon. Il identifie l'appartenance de k aux groupes $g = 1, .., G$.

On définit également le vecteur $\mathbf{x}_k^* = \gamma_k \, x_k$, qui comporte des 0 partout sauf dans les groupes qui contiennent k. Ces composantes non nulles valent alors x_k, le nombre d'équivalents plein temps de k. La notation $\mathbf{x}_k^*$ est reprise de Särndal et Lunström (2005). L'information auxiliaire x est connue pour toute la population (InfoU) et le calage se fait directement des répondants à la population, sans passer par le niveau de l'échantillon brut s.

Les valeurs de référence pour le calage sont définies par $\sum_U \mathbf{x}_k^* = (X_1^*, .., X_g^*, .., X_G^*) = \mathbf{X}^*$ avec X_g^* le total des équivalents plein temps dans le groupe g. Il s'agit des marges de calage.

Dans la STATEM, les groupes $g = 1, .., G$ sont définis à partir des objectifs de l'enquête et du comportement de non-réponse. Le premier ensemble de marges correspond aux `NOGA2reg` croisées avec deux classes de taille. Le deuxième ensemble de marges correspond aux sec-

tions croisées avec `region_new` et deux classes de taille. Les deux classes de taille sont définies comme suit : `taical`=1 si `tai2` in (1 2 3) et `taical`=2 si `tai2` in (4 99). Il s'agit donc des "petits normaux" (moins de 50 EPT) et des "grands ou spéciaux" (au moins 50 EPT ou partie de multi-établissement ou permanents STATEM).

On notera que les deux groupes de marges ne sont pas disjoints. Chaque élément se trouve dans deux marges différentes ; voir figure 2. Les groupes $g = 1, .., G$ ne forment donc pas une partition de U. Pour éviter une redondance dans l'information fournie par les marges, on retire les quatre dernières valeurs de la deuxième marge (`region_new`=7, *i.e.* Tessin). Les vecteurs γ_k et $\mathbf{x}_k^*$ contiennent donc deux valeurs non nulles si k est hors du Tessin et une seule valeur non nulle si k est au Tessin.

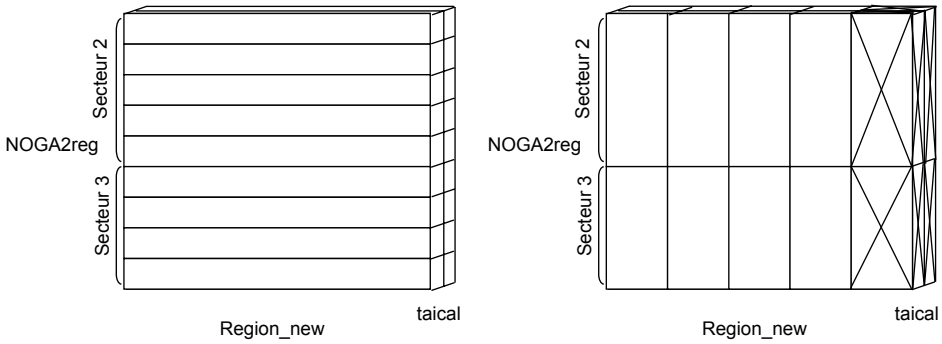


FIG. 2 Illustration des groupes de référence pour le calage : `NOGA2reg` x `taical` et `region_new` x `taical`. Les parties barrées représente les marges exclues pour éviter une redondance dans les informations auxiliaires.

Dans le cas du calage, l'estimateur du total Y est donné par $\widehat{Y}_w = \sum_r w_k y_k$ et on cherche un poids w_k qui permette de satisfaire l'équation de calage : $\sum_{k \in r} w_k \mathbf{x}_k^* = \mathbf{X}^*$ en minimisant la distance entre le poids initial et w_k. Les équations générales permettent de déterminer w_k pour $k \in r$ par une forme linéaire lorsque la distance a la forme dite du "chi-carré". Ces équations partent d'un poids initial qui est souvent le poids d'échantillonnage d_k et déterminent v_k de façon à ce que $w_k = d_k v_k$ satisfassent l'équation de calage. On obtient :

$$v_k = 1 + \lambda_r' \mathbf{x}_k^* \tag{7}$$

$$\lambda_r' = \left(\mathbf{X}^* - \sum_r d_k \mathbf{x}_k^{*'} \right)' \left(\sum_r d_k \mathbf{x}_k^* \mathbf{x}_k^{*'} \right)^{-1} \tag{8}$$

Il est également possible d'utiliser des logiciels tels la macro CALMAR de l'INSEE pour déterminer les poids finaux. Cette macro permet de déterminer l'estimateur linéaire donnant les même résultats que la formule avec le λ ou d'autres estimateurs (autres distances) tels que présentés par exemple dans Deville et Särndal (1992). On s'éloigne alors de l'estimateur linéaire basé sur le λ.

Dans le cas de la STATEM, le poids d'échantillonnage d_k est ajusté par la robustification des valeurs aberrantes avant d'effectuer le calage. Ainsi, on part du poids initial $\tilde{d}_k = d_k u_k$ et non pas de d_k dans les formules ; voir Särndal et Lunström (2005), chapitre (6.7). Nous choisissons de plus une forme non linéaire afin de limiter la correction des poids v_k dans un intervalle donné (méthode "logit" de CALMAR, limites : 0.25 et 4). Le poids initial donné dans la macro CALMAR est donc le produit $\tilde{d}_k = d_k u_k$ et le poids final prend la forme $w_k = \tilde{d}_k v_k = d_k u_k v_k$,

avec v_k l'ajustement pour le calage. CALMAR fournit le poids final w_k pour $k \in r$. L'ajustement v_k pour chaque répondant k est défini par $v_k = w_k/\tilde{d}_k$.

6 Estimation de la variance et du CV

L'estimation de la variance appliquée dans la révision de la STATEM est basée sur l'estimation de variance d'un estimateur GREG linéaire ("generalized regression estimator"). Sans non-réponse, la variance d'un estimateur de calage est en effet approximativement égale à celle d'un estimateur GREG, voir Deville et Särndal (1992). La formule est ensuite adaptée pour le cas avec non-réponse en faisant des hypothèses assez strictes sur le processus de réponse.

Pour rappel, le coefficient de variation de $\widehat{Y}$ est donné par le quotient :

$$\widehat{CV}(\widehat{Y}) = \frac{\sqrt{\widehat{var}(\widehat{Y})}}{\widehat{Y}}. \tag{9}$$

avec l'estimateur du total $\widehat{Y}$ et l'estimateur de sa variance $\widehat{var}(\widehat{Y})$. Il est souvent exprimé en %.

6.1 Estimateur GREG sans non-réponse

Soit un échantillon s tiré dans U selon un plan $p(s)$. La probabilité d'inclusion de k dans s est donnée par π_k et la probabilité d'inclusion d'ordre deux de k et ℓ dans s par $\pi_{k\ell}$. Le poids d'échantillonnage de k dans s est donné par $d_k = 1/\pi_k$ et on note $d_{k\ell} = 1/\pi_{k\ell}$.

Dans un environnement sans non-réponse avec un échantillon s tiré dans U et utilisation d'une information auxiliaire $\mathbf{x}$ par un modèle de régression, on obtient l'estimateur GREG :

$$\widehat{Y} = \sum_s d_k g_k y_k = \sum_s w_k y_k, \tag{10}$$

avec le poids final w_k et l'ajustement g_k défini par :

$$\begin{aligned} g_k &= 1 + c_k \lambda' \mathbf{x}_k, \text{ avec} \\ \lambda' &= \left(\sum_U \mathbf{x}_k - \sum_s d_k \mathbf{x}_k \right)' \left(\sum_s d_k \, c_k \, \mathbf{x}_k \mathbf{x}_k' \right)^{-1}, \end{aligned} \tag{11}$$

où c_k est le facteur d'échelle. Dans la plupart des applications $c_k = 1$.

Une estimation de la variance est donnée par Deville et Särndal (1992) :

$$\widehat{var}(\widehat{Y}) = \sum_s (d_k d_\ell - d_{k\ell}) \widehat{e}_k \widehat{e}_\ell \tag{12}$$

avec les résidus :

$$\begin{aligned} \widehat{e}_k &= y_k - \mathbf{x}_k' \widehat{\mathbf{b}}, \text{ avec} \\ \widehat{\mathbf{b}} &= \left(\sum_s d_k \, c_k \, \mathbf{x}_k \, \mathbf{x}_k' \right)^{-1} \left(\sum_s d_k \, c_k \, \mathbf{x}_k \, y_k \right). \end{aligned} \tag{13}$$

Il s'agit de la version sans pondération des résidus par g_k.

Dans le cas d'un plan stratifié avec partition $U = U_1 \cup .. \cup U_h \cup .. \cup U_H$ et $c_k = 1$, on obtient l'estimation de variance :

$$\widehat{var}(\widehat{Y}) = \sum_h N_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_{eh}^2}{n_h}, \text{ avec} \quad (14)$$

$$s_{eh}^2 = \frac{1}{n_h - 1} \sum_{k \in s_h} (\widehat{e}_k - \overline{\widehat{e}}_h)^2 \quad (15)$$

avec N_h la taille de U_h, s_h l'échantillon tiré dans U_h, n_h la taille de l'échantillon s_h et les résidus $\widehat{e}_k$ définis en (13). La moyenne des résidus $\widehat{e}_k$ dans h est notée $\overline{\widehat{e}}_h$.

La formule (14) a la forme classique de la variance dans un plan stratifié, avec remplacement de la variable d'intérêt y_k par le résidu $\widehat{e}_k$. L'utilisation de variables auxiliaires dans l'estimateur GREG permet de réduire la variance. En effet, on choisit des variables auxiliaires bien corrélées avec y et la variabilité du résidu devient inférieure à celle de y.

Plusieurs approches sont envisageables pour les estimations par domaine ; voir par exemple Särdnal et al. (1992). On choisira ici l'approche qui consiste à déterminer des résidus dans un modèle général et de les utiliser dans les estimations des différents domaines. On ne redéfinit pas le modèle et les résidus au niveau de chaque domaine. Ainsi, l'estimation de variance de $\widehat{Y}_d$ dans le domaine d reprend la formule (14) en remplaçant $\widehat{e}_k$ par $\widehat{e}_k I_{kd}$ avec $I_{kd} = 1$ si k dans domaine d et 0 sinon.

6.2 Intégration de la non-réponse dans l'estimation

Les estimations de variance se compliquent en cas de non-réponse car des biais d'une ampleur inconnue entrent en jeu. De plus, peu d'articles sont disponibles sur le sujet. Dans la plupart des application pratiques de la statistique publique des hypothèses relativement fortes sont utilisées de façon à définir des estimateurs pas trop complexes. Nous utilisons également cette approche dans les premières estimations de la variance dans la STATEM. Des travaux plus approfondis sont prévus de façon à tester des alternatives. Ils utiliseront notamment les estimations proposées par Särndal et Lundström (2005) car les poids STATEM ont été déterminés par l'approche calage.

L'estimateur de variance utilisé pour la révision 2007 de la STATEM est donné par :

$$\widehat{var}(\widehat{Y}) = \sum_h N_h^2 \left(1 - \frac{m_h}{N_h}\right) \frac{s_{eh}^2}{m_h} \quad (16)$$

$$s_{eh}^2 = \frac{1}{m_h - 1} \sum_{k \in r_h} (\widehat{e}_k - \overline{\widehat{e}}_h)^2 \quad (17)$$

avec N_h la taille de U_h, r_h l'échantillon des répondants dans U_h, m_h la taille de l'échantillon répondant r_h et les résidus $\widehat{e}_k$ définis plus bas.

On fait donc l'hypothèse que l'échantillon des répondants est un échantillon stratifié tiré directement dans la population, et donc que la probabilité de répondre de $k \in s$ est fixe au niveau de chaque strate d'échantillonnage. La réalité n'est probablement pas trop éloignée de cette hypothèse dans le cas STATEM car les strates sont très détaillées et représentent des groupes d'établissements relativement homogènes (taille, activité économique, région).

Le modèle permettant de déterminer les résidus $\widehat{e}_k$ pour $k \in r$ est donné par :

$$y = \mathbf{b}X + e \quad (18)$$

avec la variable d'intérêt y_k, la matrice des variables explicatives X et le vecteur des coefficients de régression $\mathbf{b}$. Le modèle n'a pas d'ordonnée à l'origine (no intercept). La matrice des variables explicatives X est formé des vecteurs $\mathbf{x}_k^* = \gamma_k \, x_k$ du calage ; voir Section 5.2.

Le vecteur des coefficients $\mathbf{b}$ est estimé par la méthode des moindres-carrés pondérée :

$$\widehat{\mathbf{b}} = (X' \, W \, X)^{-1} X' \, W \, y, \tag{19}$$

avec W la matrice diagonale des poids d'échantillonnage d_k (`poidsTir`).

Le résidu $\widehat{e}_k$ est donné par :

$$\widehat{e}_k = y_k - \widehat{\mathbf{b}} \mathbf{x}_k^*. \tag{20}$$

avec le vecteur de variable utilisé dans le modèle $\mathbf{x}_k^*$ et le vecteur de coefficients estimés $\widehat{\mathbf{b}}$.

L'estimation par domaine fonctionne de façon identique à celle décrite dans la section GREG sans non-réponse. On remplace $\widehat{e}_k$ par par $\widehat{e}_k I_{kd}$ avec $I_{kd} = 1$ si k dans domaine d et 0 sinon.

6.3 Quelques particularités de la STATEM

Les données de la STATEM présentent une situation plus complexe que celle présentée ci-dessus.

L'ensemble du profiling est sélectionné exhaustivement et les données sont complètes (relevé ou imputation). Sauf exception due à une mutation, les poids d'échantillonnage d_k et les poids finaux w_k sont donc tous égaux à 1. Nous pouvons alors retirer le profiling des estimations de variance. Il n'apporte aucune variabilité à l'estimateur $\widehat{Y}$. Le profiling intervient par contre dans le total général $\widehat{Y} = \sum_{k \in r} w_k \, y_k$ dans le calcul du coefficient de variation.

On a une correction u_k de robustification en plus du calage. Cette correction touche peu d'établissements et est négligée dans les estimations de variance. De façon générale, on note que le poids d_k est pris dans les estimations de $\mathbf{b}$. Les alternatives $d_k u_k$ ou $d_k u_k v_k$ ne sont pas retenues. Dans la STATEM, les estimations de variance dépendent peu du poids choisi.

Il y a des imputations dans les très grands établissements de la partie variable de l'échantillon. L'apport de ces imputations à la variance est négligée. La sous-estimation de variance doit cependant être petite car il y a peu de cas.

L'identification de la strate de tirage de chaque établissement est disponible dans le cadre de sondage au temps $T0$ (`straTir`). Au temps T la liste des établissements a légèrement changé pour cause de mutations. Ainsi, certains établissements n'ont plus de strate de tirage. C'est le cas par exemple du résultat de la fusion entre deux établissements provenant de deux strates différentes. Pour parer à cette difficulté, nous assignons une strate de tirage fictive aux établissements du cadre d'extrapolation sans strate de tirage au temps T. Cette strate `StraAll` est construite comme les strates de tirage, *i.e.* sur la base des variables `special`, `NOGA2reg`, `region_new` et `tai2`. Durant l'estimation de variance, on considère que le tirage a eu lieu dans les strates `StraAll`. Les probabilités d'inclusion sont conservées mais elles ne correspondent plus toujours au quotient n_h/N_h. L'effet de cette approximation doit être encore évaluée.

La stratification de la STATEM est très détaillée. Certaines strates sont très petites et quelques unes peuvent même n'avoir aucun répondant. Cette foultitude de strates ne pose pas de problème notable au moment de la détermination des poids car la méthode choisie travaille à des niveaux agrégés. Par contre, elle risque de mettre de l'instabilité dans les estimations de variances effectuées avec la méthode prévue. Des essais de regroupements de strates pourront être testés pour stabiliser les estimations.

Les diverses simplifications retenues dans la cas de la STATEM nous encouragent à considérer les estimations de variance avec un regard critique. Comme toujours en cas de non-réponse, les estimations peuvent contenir des biais et devraient être considérées comme des ordres de grandeur. Des travaux complémentaires sont en cours pour mieux comprendre le comportement des estimations proposées et les comparer à des alternatives telle l'approche calage de Särdnal et Lundström (2005). Quelques pistes sont données dans la Section 8.

7 Résultats

Cette section présente quelques résultats obtenus pour les 2ème, 3ème et 4ème trimestre 2007 comme illustration des méthodes mise en place dans la révision 2007. Les résultats numériques présentés dans ce rapport ne sont pas les résultats officiels de la STATEM. Pour les résultats officiels, voir les communiqués de presse et les informations disponibles sur le site de l'OFS http://www.statistique.admin.ch.

7.1 Préparation du cadre d'extrapolation

Le cadre de sondage de la STATEM défini pour le 2ème trimestre 2007 comportait 385'098 établissements pour un total de 3'718'117 emplois (betot). Le cadre d'extrapolation pour le même trimestre comportait un nombre d'établissements et un nombre d'emploi très similaire ; voir Table 3. Dans les trimestres suivants, le cadre subit peu de modifications. Le nombre d'établissements diminue légèrement et le nombre d'emploi augmente légèrement. Ces petites variations sont dues au remplacement du bloc profiling et aux fusions. La part des disparus conservés dans le cadre augmente dans le temps mais reste relativement petite par rapport au cadre.

TAB. 3 Information sur les cadres d'extrapolations des 2ème, 3ème et 4ème trimestres 2007. Nombre d'établissements et emploi betot selon le cadre (nombre absolus en milliers et répartition en %) pour la la partie fixe (base), les disparus (sans les modifications du profiling), les nouveaux profiling et les nouveaux apparus en tant que successeurs.

		Base	Disparition	Nouv. prof.	Nouv. succ	Total
2T07	N	379985	3961	655	57	384658
	%N	98.8	1	0.2	0	100
	betot	3688.1	21.1	8.8	0.7	3718.7
	%betot	99.1	0.6	0.2	0	100
3T07	N	377222	5773	1196	130	384321
	%N	98.2	1.5	0.3	0	100
	betot	3674.4	29.3	17.9	1.8	3723.4
	%betot	98.7	0.8	0.5	0	100
4T07	N	374604	8132	1286	171	384193
	%N	97.5	2.1	0.3	0	100
	betot	3663.9	42.4	20.1	2.0	3728.3
	%betot	98.3	1.1	0.5	0.1	100

7.2 Préparation des données d'enquête

Le taux de réponse des établissements de l'échantillon STATEM est élevé avec une valeur entre 84 et 85% ; voir Table 4. Après imputation, il reste une non-réponse de l'ordre de 9 à 11%. Une proportion non négligeable de 9-10% des emplois de l'échantillon final sont imputés.

TAB. 4 Établissements répondants, imputés et non-répondants aux 2ème, 3ème et 4ème trimestres 2007. Nombre d'établissements et emploi `betot` selon le cadre (nombres absolus en milliers et répartition en %)

		Non réponse	Réponse	Imputation	Total
2T07	N	5408	52760	4387	62555
	%N	8.6	84.3	7	100
	betot	108.4	1864.2	226.8	2199.5
	%betot	4.9	84.8	10.3	100
3T07	N	6527	53044	2780	62351
	%N	10.5	85.1	4.5	100
	betot	109.6	1903.4	191.8	2204.8
	%betot	5	86.3	8.7	100
4T07	N	7124	52301	2878	62303
	%N	11.4	83.9	4.6	100
	betot	126.2	1863.5	220.3	2209.9
	%betot	5.7	84.3	10	100

Les 294 multi-établissements sont répartis dans environ 2100 établissements individuels ; voir Table 5. Cette imputation technique de répartition des emplois relevés en agrégats correspond à env. 3.7% des emplois totaux (voir %TOTpond). Le reste des emplois imputés provient principalement du profiling pour le 2ème trimestre 2007, puis des très grands pour les trimestres suivants. L'imputation du profiling au 2ème trimestre est due au fait que certains groupes ne sont pas encore intégrés dans la procédure de relevé profiling (profiling futur). Une partie de ces établissements sont intégrés à partir du 3ème trimestre. Il y a très peu d'imputations de très grands établissements au 2ème trimestre 2007 car le taux de réponse est excellent (94.4% des normaux) et peu de cas sont utilisables dans le fichier du trimestre précédent. Ces imputations deviennent plus fréquentes ensuite car le taux de réponse a légèrement diminué (88-91% des normaux) et il y a plus de données disponibles pour l'imputation.

7.3 Pondération

Environ 73% des établissements répondants aux 2ème et 3ème trimestres 2007 ont un poids initial `PoidsTir` égal à 1. Il s'agit pour moitié de spéciaux et pour moitié de normaux tirés dans des strates exhaustives ; voir Table 6. On note que quelques spéciaux ont un poids initial différent de 1. Il s'agit d'établissements tirés comme normaux puis entrés dans les spéciaux entre le moment du tirage et le trimestre en cours. Quelques établissements ont aussi des poids initiaux inférieurs à 1. Il s'agit d'établissements fusionnés provenant en partie de l'échantillon et en partie d'établissements hors échantillon. Le poids déterminé par le partage des poids peut être inférieur à 1.

Les établissements avec poids initial `PoidsTir` égal à 1 sont exclus de la procédure de robustification. Le facteur `u_res_def` est donc égal à 1. Par contre, le calage a un effet sur les poids. Le facteur `corr_cal` prend des valeurs entre 0.9 et 2.9, resp. 3.6 dans les trimestres

TAB. 5 Établissements répondants et imputés aux 2ème, 3ème et 4ème trimestres 2007. Répondants à l'enquête réponse et imputations des multi-établissements imputmult, profiling imputprof et des très grands imputVB. Nombre d'établissements N (nombres absolus et répartition en %) et emploi TOT selon le relevé ou l'imputation (nombre absolus en milliers, répartition non pondérée et répartition pondérée en %). Le poids final d'extrapolation poidsFin est utilisé pour estimer %TOTpond.

		réponse	imputmult	imputprof	imputVB	Total
2T07	N	52760	2189	2188	10	57147
	%N	92.3	3.8	3.8	0	100
	TOT	1917.2	132.1	100.6	3.7	2153.5
	%TOT	89	6.1	4.7	0.2	100
	%TOTpond	93.5	3.8	2.7	0.1	100
3T07	N	53044	2181	377	222	55824
	%N	95	3.9	0.7	0.4	100
	TOT	1969.3	135.6	19.5	47.5	2171.8
	%TOT	90.7	6.2	0.9	2.2	100
	%TOTpond	94.5	3.7	0.5	1.3	100
4T07	N	52301	2092	426	360	55179
	%N	94.8	3.8	0.8	0.7	100
	TOT	1930.1	135.7	22.1	76.8	2164.7
	%TOT	89.2	6.3	1	3.5	100
	%TOTpond	93.6	3.7	0.6	2.1	100

considérés. Les poids finaux PoidsFin restent donc tous petits.

Les établissements avec poids initial PoidsTir différent de 1 sont en majorité des petits établissements (tai2=1, equivpt$<$5) et assez petits établissements (tai2=2, 5$\leq$equivpt$<$20). Les poids initiaux sont tous inférieurs à 42. La robustification touche peu de cas (189 au 2ème trimestre 2007, 188 au 3ème trimestre 2007 et 208 au 4ème trimestre 2007). L'ajustement de calage prend des valeurs entre 1 et 2.2, resp. 2.9 pour les trimestres considérés. Les poids finaux PoidsFin sont au maximum égaux à 48, resp. 55.

Les cas robustifiés sont présents dans toutes les region_new et presque toutes les NOGA2reg. La majorité des cas se trouvent dans tai2=1 avec TOT très supérieur à equivpt. Une partie des cas montre cependant un TOT inférieur à equivpt ; voir Figure 3. On notera que plusieurs établissements robustifiés sont saisonniers. Ils ont donc de grandes variations saisonnières.

Une analyse complémentaire sur les résultats des 2ème, 3ème et 4ème trimestres 2007 a montré que la robustification suivie du calage donne des résultats environ de 0.5 à 0.9% inférieur, pour le total des emploi, à un calage sans robustification. L'effet relativement réduit lors du 2ème trimestre augmente donc dans le temps. Il sera bon de réévaluer les paramètres lors du renouvellement au printemps 2008.

La détermination de v_k dans le cas de la STATEM utilise la macro CALMAR de l'Institut national de la statistique et des études économiques (INSEE). Pour des raisons pratiques liées au nombre important de marges, les différents objectifs sont traités par étapes successives : (1) calage sur les marges NOGA2reg x taical, (2) calage sur les marges region_new x secteur x taical, (3) calage sur les marges NOGA2reg x taical, et (4) calage sur les marges region_new x secteur x taical. Il s'agit donc d'une sorte de raking ratio manuel. A chaque étape, la méthode logit est appliquée de façon à restreindre les ajustements dans l'intervalle [0.25 ; 4] (m=3, lo=0.25 et up=4).

TAB. 6 Statistique des poids des répondants du 2ème et du 4ème trimestre 2007 avec les diverses étapes depuis le poids initial `PoidsTir` jusqu'au poids final `PoidsFin` en passant par la robustification `u_res_def`, le poids robustifié `PoidsCor` et l'ajustement provenant du calage `corr_cal`. Nombre de cas N, puis minimum Min, médiane Med et maximum Max. Valeurs pour les établissements avec poids initiaux égaux ou différent de 1 et les différentes classes de taille `tai2`.

2ème trimestre 2007

	PoidsTir différent de 1					PoidsTir =1					All
	tai2					tai2					All
	1	2	3	4	99	1	2	3	4	99	
Nb de cas N	9920	4811	969	26	14	3487	4148	6358	6479	20935	57147
PoidsTir Min	0.2	0.9	0.6	0.7	2.5	1	1	1	1	1	0.2
Med	31.3	7.6	4.4	1	27.7	1	1	1	1	1	1
Max	41.4	33.8	28.1	3.8	32	1	1	1	1	1	41.4
u_res_def Min	0.2	0.3	0.2	1	1	1	1	1	1	1	0.2
Med	1	1	1	1	1	1	1	1	1	1	1
Max	1	1	1	1	1	1	1	1	1	1	1
PoidsCor Min	0.2	0.8	0.5	0.7	2.5	1	1	1	1	1	0.2
Med	31.2	7.6	4.4	1	27.7	1	1	1	1	1	1
Max	41.4	33.8	28.1	3.8	32	1	1	1	1	1	41.4
corr_cal Min	1	1	1	1	1	1	1	0.9	0.9	0.9	0.9
Med	1	1.1	1.2	1	1	1	1.1	1.2	1	1	1
Max	1.1	1.5	2.2	1.2	1	1.2	1.8	2.9	1.7	1.6	2.9
PoidsFin Min	0.2	0.9	0.6	0.7	2.5	1	1	0.9	0.9	0.9	0.2
Med	31.5	8.1	5.8	1	27.7	1	1.1	1.2	1	1	1
Max	43.5	47.3	41.2	3.8	32	1.2	1.8	2.9	1.7	1.6	47.3

4ème trimestre 2007

	PoidsTir différent de 1					PoidsTir =1					All
	tai2					tai2					All
	1	2	3	4	99	1	2	3	4	99	
Nb de cas N	9398	4593	921	47	19	3207	3942	5949	6337	20766	55179
PoidsTir Min	0.2	0.3	0.6	0.6	0.5	1	1	1	1	1	0.2
Med	31.2	7.6	4.4	1	10.3	1	1	1	1	1	1
Max	41.4	33.8	28.1	5.3	32	1	1	1	1	1	41.4
u_res_def Min	0.2	0.2	0.2	1	0.4	1	1	1	1	1	0.2
Med	1	1	1	1	1	1	1	1	1	1	1
Max	1	1	1	1	1	1	1	1	1	1	1
PoidsCor Min	0.2	0.3	0.6	0.6	0.5	1	1	1	1	1	0.2
Med	31	7.6	4.4	1	10.3	1	1	1	1	1	1
Max	41.4	33.8	28.1	5.3	32	1	1	1	1	1	41.4
corr_cal Min	1	1	1.1	1	1	1	1	1	0.9	0.9	0.9
Med	1	1.1	1.3	1	1	1	1.1	1.4	1	1	1
Max	1.2	1.8	2.9	1.2	1.1	1.2	1.8	3.6	1.5	2	3.6
PoidsFin Min	0.2	0.4	0.7	0.6	0.5	1	1	1	0.9	0.9	0.2
Med	31.6	8.5	6.1	1	10.3	1	1.1	1.4	1	1	1
Max	44.3	54.5	42.5	5.3	32	1.2	1.8	3.6	1.5	2	54.5

La convergence de l'algorithme CALMAR est bonne dès qu'il y a suffisamment d'établissements dans l'échantillon des répondants. Des difficultés sont par contre observées dans le cas des calculs provisoires basés sur un ensemble réduit de répondants. Elles sont résolues si l'on relaxe les contraintes [0.25 ; 4]. Dans une application future, on pourra éventuellement revoir l'ordre des marges considérées dans CALMAR en commençant par les grandes marges `region_new` x `secteur` x `taical` et non pas `NOGA2reg` x `taical` de façon à pouvoir produire des résultats provisoires sans devoir modifier les bornes de la méthode logit.

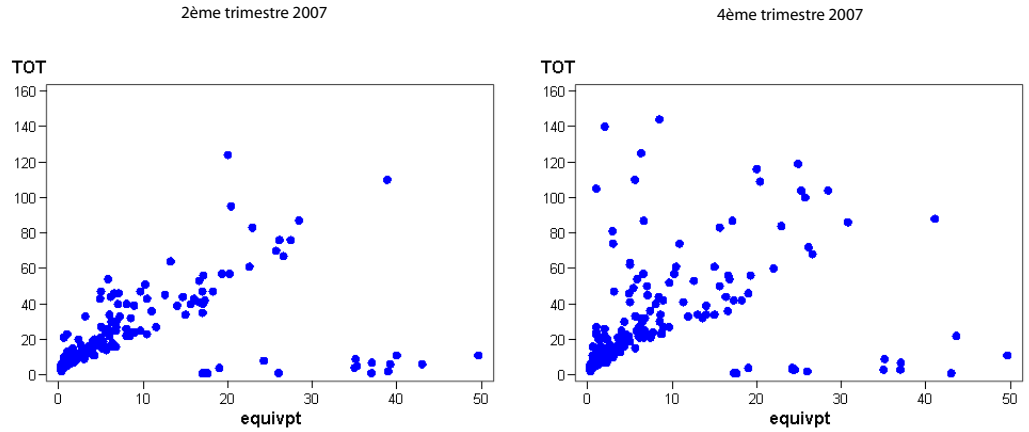


FIG. 3 Établissements avec poids robustifiés $0.2 \leq u_k < 1$ pour le 2ème et le 4ème trimestre 2007. Les établissements sans robustification se trouvent dans l'espace blanc entre les nuages.

7.4 Estimations

Les estimations des totaux des variables d'intérêt et de leur variabilité sont effectuées à l'aide de macros SAS. Ces macros utilisent notamment la procédure PROC SURVEYMEANS avec l'option DOMAIN appliquée sur les résidus du modèle correspondant au calage pour la variabilité.

Les objectifs de précision sont atteints dans tous les domaines objectifs pour la variable utilisée pour la construction du plan (équivalents plein temps, EPT). Les CV sont même en général inférieurs à ce qui était demandé pour les variables de l'emploi total et plein temps. Sans surprise, les précisions sont plus faibles pour les temps partiels et surtout les temps partiels par sexe ; voir les Tableaux dans l'Annexe C. Il faut cependant noter que l'estimation de la variance ne tient pas compte de la variabilité des imputations et qu'elle est très instable dans certains domaines. La variable "test" identifie quelques domaines avec des estimations de variance très peu fiables.

Les CV restent stables avec en moyenne une très légère augmentation entre les 2ème et 4ème trimestres car le taux de réponse diminue légèrement. Dans la figure 4, on visualise les intervalles de confiance pour l'emploi à temps plein global, hommes et femmes. Les CV correspondants sont de env. 0.2%, 0.25% et 0.45%.

Lors du renouvellement de l'échantillon au 2ème trimestre 2007, nous avons deux échantillons. Celui suivi depuis 2003 et celui nouvellement tiré dans le nouveau cadre. Les estimations avec le nouvel échantillon, le nouveau cadre et la nouvelle méthode sont globalement plus élevées que celles obtenues sur l'ancien échantillon avec l'ancienne méthode d'estimation et l'ancien cadre. Le total augmente de 2.5%, dont 1.9% dans le secteur 2 et 2.7% dans le secteur 3. Les différences varient entre -12% et 15% au niveau des activités économiques. Les plus grandes variations touchent les petites `NOGA2reg`. Ces différences sont lissées lors de l'adaptation des séries sur les recensements.

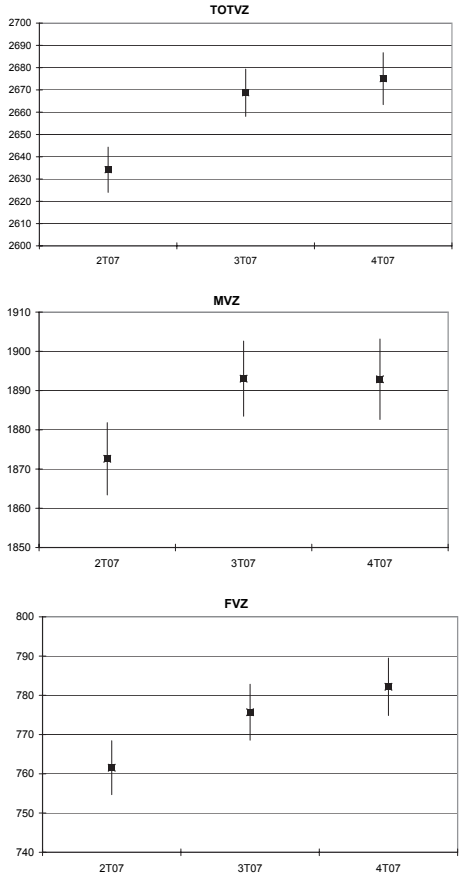


FIG. 4 Estimation du nombre d'emploi temps plein total (TOTVZ), pour les hommes (MVZ) et pour les femmes (FTZ). Total estimé et intervalles de confiance à 95% (en milliers d'emplois)

8 Un peu plus sur les estimations

Ce chapitre propose quelques réflexions sur les estimations ainsi que sur leur possible développement.

Cadre d'extrapolation et établissements du profiling

Le cadre d'extrapolation est construit sur la base du cadre de sondage et intègre les mutations par le biais du reporting. Le procédé marche bien dans la plupart des situations. Il reste cependant quelques simplifications qui peuvent affecter la qualité du cadre.

Premièrement, les établissements subissant plusieurs mutations successives ne sont pas toujours suivis de manière adéquate. Les comparaisons de cadres sont actuellement faites par rapport au cadre de sondage du dernier renouvellement, *i.e.* au maximum un an en arrière. Un établissement tiré au 2ème trimestre 2007, fermé avec successeurs au 3ème trimestre 2007 puis encore fermé avec successeur au 1er trimestre 2008 disparaîtra de l'échantillon comme s'il avait été fermé sans successeur. Ces cas, peu fréquents, doivent être analysés.

Deuxièmement, les établissements du profiling - regroupés en groupes profiling - ont un traitement très différent des autres établissements dans les estimations. Ils sont repris en bloc chaque trimestre, sans passer par un suivi du type reporting car les mutations telles les fermetures avec successeur ou les fusions ne sont pas documentées dans le REE. Les passages

entre normaux et profiling sont traités de manière adéquate par le reporting (par ex. reprise des poids d'origine). Par contre, les groupes profiling sans information de passage entre normaux et profiling sont considérés représenter toujours la même population. Il peut y avoir des réorganisations internes mais pas de vraies entrées ni de vraies sorties du groupe. Un suivi très détaillé et complexe est donc nécessaire sur les établissements du profiling afin de retirer les éléments hors population ou ajouter des parties perdues[7].

Des difficultés ont aussi apparu dans le profiling public car plusieurs types d'établissements coexistent dans le REE. Certains relevés sont en fait des agrégats du type multi-établissements ("structure L26"). Les agrégats ne sont pas à cheval sur deux régions mais ils peuvent toucher plusieurs activités économiques. Tous les emplois d'un département de l'instruction publique sont par exemple assignés à l'administration alors qu'une grande majorité des emplois sont en fait dans le domaine de l'enseignement. Le problème des "L26" devient encore plus complexe lorsqu'une partie des emplois ne fait pas partie de la population STATEM (jardiniers comptabilisés dans l'administration d'une ville alors qu'ils sont dans le secteur primaire). Ces cas ont été traités selon une procédure ad hoc durant les premiers trimestres suivant la révision. Ils recevront un traitement spécial du type redistribution des multi-établissements à partir du renouvellement 2008.

Pondération

La procédure de détection de valeurs aberrantes et la robustification introduite dans la révision sont des outils appréciés dans la production. Ils permettent d'avoir une procédure claire sur le traitement de ces cas perturbateurs. On notera cependant que le réglage des paramètres de la procédure est complexe. Les quelques résultats dans la Section 7.3 montrent que l'effet de la robustification est de plus en plus important dans les trois trimestres étudiés. Un suivi détaillé et une réévaluation des paramètres sont maintenant nécessaires pour parfaire la procédure.

Le poids final est basé sur la procédure de calage. A nouveau, il a fallu faire des choix. Nous avons choisi le nombre d'équivalents plein temps selon le cadre `equivpt` comme variable de calage et les totaux de référence dans les domaines objectifs croisés avec des classes de taille. Ce choix semble raisonnable pour les besoins de la STATEM. On pourrait cependant envisager ajouter un calage sur les tailles N_g^* au plus des totaux X_g^* de façon à ajouter des variables auxiliaires et conserver la propriété suivante : la somme des poids est égale à taille de la population. Ce complément pourrait être évalué de façon à voir si cet ajout améliore les estimations.

Variance

L'estimation de la variance proposée nous fournit des estimations qu'il faut considérer comme des ordres de grandeur, surtout pour des petits domaines. Nous avons vu dans la Section 6 que la situation de la STATEM est relativement complexe et qu'elle est donc partiellement prise en compte dans les estimations de variance proposées. Les travaux futurs permettront d'avoir une vue d'ensemble plus claire sur les forces et faiblesses de ces estimations, tout en sachant que des estimations en présence de non-réponse ne pourront jamais être considérées comme très précises et bien contrôlées. Maintenant que les résultats STATEM sont disponibles pour plusieurs trimestres, les travaux sur les estimations de variance peuvent reprendre. Ils continueront notamment les essais de regroupements de strates avec poids égaux, avec éventuels regroupements différents en fonction des domaines considérés. Ils avanceront également dans l'évaluation de l'approche de Särndal et Lundström (2005) dans le cadre général du ca-

[7] Lors du renouvellement de 2008, on décide de fixer pendant une année la liste des groupes profiling retenus dans la STATEM. La liste ne contient donc que des groupes avec relevé régulier et de bonne qualité afin de diminuer les incertitudes liées à la partie profiling du relevé.

lage. Cette approche a l'avantage de prendre en compte explicitement la non-réponse dans les estimations de variance.

Précision espérée et obtenue

La précision des estimateurs est en général plus élevée que celle exigée par les objectifs de l'enquête. Cette différence est due à plusieurs éléments. Tout d'abord, les objectifs de l'enquête ont aboutis à un plan d'échantillonnage très détaillé avec une taille d'échantillon importante pour des raisons techniques de taille minimale dans les strates ; voir discussion dans Renaud (2008). A la base, on pouvait donc espérer une bonne précision des estimateurs de totaux dans la plupart des domaines objectifs. De plus, les taux de réponse utilisés dans la construction du plan étaient un peu inférieurs à ceux observés dans les trimestres étudiés jusqu'à ce jour. On notera encore que l'imputation des très grands établissements, qui prend de l'importance dès le 3ème trimestre 2007, permet d'éviter les effets parfois catastrophique de la non-réponse des très grands. On rappellera cependant que les estimations de variance proposées actuellement restent des valeurs approximatives.

Renouvellements et adaptations de séries

Avant la révision 2007, les résultats de la STATEM étaient révisés périodiquement suite à la publication des résultats du recensement des entreprises et des établissements. On renouvelait l'échantillon et on adaptait les séries rétroactivement sur les recensements. Après la révision 2007, on renouvelle l'échantillon annuellement. La méthode de production des séries est cependant similaire. A chaque renouvellement, on effectue le relevé pour l'ancien et le nouvel échantillon. Ce double relevé est réalisé pour le 2ème trimestre de chaque année[8]. Dans le cas du renouvellement 2008, les résultats du 2ème trimestre 2008 sont publiés selon l'ancien échantillon puis, un trimestre plus tard, selon le nouvel échantillon. Cette deuxième publication est réalisée avec celle du troisième trimestres 2008 ainsi qu'une révision de la série du 2ème trimestre 2007 au 2ème trimestre 2008. A l'avenir, les données annuelles seront donc révisées une fois par an, suite au renouvellement du cadre.

9 Conclusion

La révision 2007 de la STATEM a permis de revoir non seulement le plan d'échantillonnage mais également la méthode d'estimation. Plusieurs nouveautés ont été intégrées. On peut citer le cadre d'extrapolation suivi pendant une année, la distribution des emplois des multi-établissements, le traitement spécial pour le profiling, les imputations, la détection des valeurs aberrantes, le calage direct sur des valeurs de référence au niveau population sans passer par un ajustement spécifique à la non-réponse, et l'estimation de variance.

La production des résultats de la STATEM est basée sur les nouvelles méthodes depuis le 2ème trimestre 2007. Les outils fonctionnent de manière tout à fait satisfaisante. Quelques besoins d'approfondissements sont cependant déjà identifiés. On notera ici le choix des paramètres de robustification, la procédure choisie dans l'application de CALMAR et l'estimation de la variance. Ils seront revus pour le renouvellement du 2ème trimestre 2008.

[8] Note : le recouvrement entre les deux échantillons est de l'ordre de 85% au 2ème trimestre 2008.

Annexes

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	Administration 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 :

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.

```
* 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 :

```
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		

Définition du cadre de sondage :

La liste des établissements est basée sur l'extrait du REE (fichier `week_loc`) du 4 avril 2007 avec restriction à la population STATEM comme suit :

```
if 10 <= noga2 <= 93;
if status in (1,4);
if betyp in ('L13','L20','L21','L22',
             'L23','L24','L25','L27');
if autokz in ('AG','AI','AR','BE','BL','BS','FR','GE','GL','GR',
              'JU','LU','NE','NW','OW','SG','SH','SO','SZ','TG',
              'TI','UR','VD','VS','ZG','ZH');
```

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
```

Définition des classes de taille

```
* 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;
tai2=tai;
if special in (1,2) then tai2=99;
```

B Construction du cadre d'extrapolation

La construction du cadre d'extrapolation au temps T passe par la création d'un fichier cadre nommé *Rahmen*. Ce dernier est construit sur la base du Rahmen au temps $T0$ (état référence échantillonnage), du reporting entre $T0$ et T, des informations sur les types d'établissements aux temps $T0$ et T, et des données relevées au temps T.

Dans cette section, on notera $T1$ pour le trimestre en cours.

Rahmen et spéciaux $T0$

Le cadre Rahmen au temps $T0$ correspond au cadre d'échantillonnage du renouvellement 2007, avec ajout de quelques variables pour la suite de la procédure. Le cadre d'échantillonnage permet également de définir le fichier *profiling* au temps $T0$. Ce fichier contient les données de tous les types spéciaux (multi, profiling, permanents) au temps $T0$.

Reporting entre $T0$ et $T1$

Comparaison entre les extraits du REE au temps $T0$ et au temps $T1$ (fichiers `week_loc`). Les travaux ont été réalisés par Mathieu Gunzinger (BUR) et Jann Potterat (METH).

Dates importantes pour les fichiers :
– `DatumT0` : date $T0$ création cadre d'échantillonnage
– `DatumT1` : date $T1$ pour reporting et création Rahmen $T1$

Fichiers :
– `week_loc_&DatumT0` et `week_ent_&DatumT0` : extraits REE pour $T0$.
– `week_loc_&DatumT1` et `week_ent_&DatumT1` : extraits REE pour $T1$.
– `profiling_&DatumT0` : spéciaux $T0$.
– `profiling_&DatumT1` : spéciaux $T1$
– `cadre_analyse_&DatumT0._&DatumT1` : résultat reporting entre $T0$ et $T1$.

Le fichier `profiling_&DatumT1` est créé sur la base des informations au temps $T1$ sur les multi, les permanents et le profiling. Les listes sont adaptées à partir de la base de données de production de la STATEM. La variable `eventbesta` de `cadre_analyse_&DatumT0._&DatumT1` identifie les événements détectés entre les deux états du REE, voir Table 7. Diverses tables de contrôles sont fournies pour suivre l'évolution du cadre.

TAB. 7 Définition de la variable `eventbesta`

`eventbesta`	Définition
0	Aucun changement
1	Fermeture sans successeur entre $T0$ et $T1$
2	Fermeture avec successeur entre $T0$ et $T1$
3	Fermeture avec successeur avant $T0$
4	Sortie de la population STATEM (unité devient support légal)
5	Sortie de la population STATEM (sortie secteur ou `betyp`)
6	Transfert dans une autre entreprise
7	Nouvelle unité STATEM
8	Nouvelle unité par réactivation
9	Nouvelle unité par entrée dans population STATEM

Note : on parle de fusion lorsque $\geq$2 établissements ont le même successeur ($a + b \rightarrow c$) ou si on a un successeur déjà dans le cadre ($a + b \rightarrow a$).

Rahmen $T1$

Utilisation de Rahmen $T0$ et du reporting entre $T0$ et $T1$ pour définir Rahmen $T1$.

Dates importantes pour les fichiers[9] :
– `DatumT0` : date création cadre d'échantillonnage $T0$
– `DatumT1` : date $T1$ pour reporting et création Rahmen $T1$

Fichiers :
– `cadre_analyse_&DatumT0._&DatumT1` : cadre analyse (voir plus haut).
– `rahmen&DatumT0` : cadre $T0$
– `rahmen&DatumT1` : cadre $T1$ résultat.

Procédure :
– Restriction de `cadre_analyse_&DatumT0._&DatumT1` aux unités dans `rahmen&DatumT0` ou dans la population STATEM au temps $T1$.
– Identification des successeurs.
– Création de la variable `code`, voir Table 8.
– Ajout des informations sur les têtes des multi-établissements.
– Adaptation des emplois `betot` et `equivpt`, notamment du profiling et des successeurs.
– Adaptation des poids si nécessaire : reprise du prédécesseur ou partage des poids dans le cas de fusions.
– Définition des types d'établissements et de diverse variables manquantes, notamment information sur le nombre de prédécesseurs en cas de fermetures avec successeur ou de fusions.
– Définition de `rahmen&DatumT1`.

TAB. 8 Définition de la variable `code`.

`code`	Définition
1	Pas de changements
2	Changement d'entreprise
3	Changement de type d'établissement (`sflgprof`)
4	Changement de type d'établissement (`sflgprof`) et d'entreprise
5	Successeur déjà dans STATEM en T0
6	Successeur déjà dans STATEM en T0 avec changement d'entreprise
7	Nouveau : spécial et successeur
8	Nouveau : spécial et pas successeur
9	Nouveau : normal et successeur
10	Nouveau : normal et pas successeur
11	Plus actif avec successeur
12	Plus actif : profiling
13	Plus actif : multi ou permanent
14	Plus actif : normal
15	Chefs multi

[9] Dans les programmes de la révision 2007, on utilise la notation `DatumT` au lieu de `DatumT0`, *i.e.* `cadre_analyse_&DatumT._&DatumT1` au lieu de `cadre_analyse_&DatumT0._&DatumT1` et `rahmen&DatumT` au lieu de `rahmen&DatumT0`.

On notera que tous les nouveaux spéciaux sans prédécesseurs sont traités de la même manière (code=8). Dans cette version de la définition du Rahmen $T1$, il n'y a donc pas de différence entre profiling BUR et les autres spéciaux tels permanents et multi-établissements si aucune mutation n'est documentée dans le reporting.

Cadre d'extrapolation $T1$

Dates importantes pour les fichiers[10] :
– `DatumT1` : date création cadre Rahmen $T1$.
– `DatumBesta` : date extrait relevé STATEM $T1$
– `DatumR1` : date création cadre Rahmen au trimestre précédent $T1$
– `DatumB1` : date extrait relevé STATEM au trimestre précédent $T1$

Fichiers :
– `rahmen&DatumT1` : cadre basé sur cadre initial et reporting,
– `normaux_permanents_&DatumBesta` : relevé STATEM
– `multi_representantes_&DatumBesta` : relevé STATEM
– `structure_multi_&DatumBesta` : relevé STATEM
– `profiling_&DatumBesta` : relevé STATEM
– `cadreextra_&DatumR1._&DatumB1` : cadre extrapolation au trimestre précédent $T1$
– `cadreextra_&DatumT1_&DatumBesta` : nouveau cadre d'extrapolation

Procédure :
– Reprise de `rahmen&DatumT1` et collage des données relevées disponibles.
– Distribution des relevés des multi-établissements dans les établissements individuels.
– Imputation de tous le profiling sans réponses.
– Imputation d'une partie des very bigs non répondants.
– Exclusion des codes 10,11 et 12.
– Export du fichier `cadreextra_&DatumT1_&DatumBesta` sans les têtes multi (doubles entrées).

Des tables de contrôle fournissent une information sur les emplois et leur évolution entre cadres successifs ainsi que des informations sur les very bigs.

[10] Dans les programmes de la révision 2007, on utilise la notation `DatumRahmen` au lieu de `DatumT1`, *i.e.* `rahmen&DatumRahmen` au lieu de `rahmen&DatumT1`, `cadreextra_&DatumRahmen_&DatumBesta` au lieu de `cadreextra_&DatumT1_&DatumBesta` et `cadreextra_&DatumRahmen_&DatumBesta` au lieu de `cadreextra_&DatumT1_&DatumBesta`.

C Estimations

Les estimations de totaux sont disponibles ci-dessous pour les 2ème et 4ème trimestre 2008 afin d'illustrer les méthodes. Il ne s'agit pas des résultats officiels de la STATEM. Ces derniers sont disponibles sur http://www.statistique.admin.ch.

Légende

Information sur la population, le plan d'échantillonnage, l'échantillon des répondants et les domaines :
- glo : résultat global Suisse si glo=1
- secteur : secteur économique
- noga2reg : divisions économiques `noga2reg`
- secgr=100*secteur + gr, *i.e.* croisements entre secteurs et grandes régions
- secreg=100*secteur + region_new, *i.e.* croisements entre secteurs et `region_new` (pour régions avec compléments).
- Npop : nombre d'unités dans domaine (population)
- nBrut : nombre d'unités tirées dans l'échantillon brut (dans domaine)
- nNet : nombre d'unités dans ech net
- nNetVar : nombre d'unités dans ech net avec variabilité
- nstraPopVar : nombre total strates avec variabilité dans domaine (population)
- nstraNetVar : nombre de strates dans le domaine pour l'estimation de la variance

Estimations des totaux (en milliers) et des CV (en %) pour les variables principales :
- sum_EPT
- cv_EPT
- sum_TOT
- cv_TOT
- sum_TOTVZ
- cv_TOTVZ
- sum_TOTTZ1
- cv_TOTTZ1
- sum_TOTTZ2
- cv_TOTTZ2

Estimation des CV (en %) pour les variables détaillées par sexe :
- cv_MVZ
- cv_MTZ1
- cv_MTZ2
- cv_FVZ
- cv_FTZ1
- cv_FTZ2

Information sur la qualité de l'estimation de la variance :
- test : "ok" si nNetVar / nstraNetVar > 3 et "-" sinon.

2ème trimestre 2007

(totaux en milliers, CV en %)

											EPT		TOT		TOTVZ		TOTTZ1		TOTTZ2		MVZ	MTZ1	MTZ2	FVZ	FTZ1	FTZ2	
glo	secteur	noga2reg	secgr	secreg	Npop	nBrut	nNet	nNetVar	nstraPopVar	nstraNetVar	sum	cv	sum	cv	sum	cv	sum	cv	sum	cv	cv	cv	cv	cv	cv	cv	test
1	.	.	.	.	384658	62555	57147	38618	2595	2558	3213	0.15	3830	0.17	2634	0.19	661	0.56	535	0.91	0.25	1.14	1.62	0.45	0.63	0.98	ok
.	2	.	.	.	78705	15523	13536	13268	1214	1195	964	0.27	1029	0.27	900	0.29	78	1.20	51	1.94	0.33	2.60	3.74	0.78	1.16	2.10	ok
.	3	.	.	.	305953	47032	43611	25350	1384	1363	2249	0.18	2801	0.21	1734	0.25	583	0.62	484	0.99	0.35	1.26	1.75	0.52	0.69	1.06	ok
.	.	10.14	.	.	393	393	347	347	43	42	5	0.48	5	0.45	5	0.53	0	2.18	0	2.73	0.55	3.33	3.48	1.86	2.23	3.33	ok
.	.	15	.	.	2589	898	722	687	70	68	54	0.60	60	0.74	48	0.80	7	4.24	5	6.12	1.14	8.16	7.78	1.84	3.80	7.66	ok
.	.	16	.	.	15	15	12	12	11	10	3	0.08	3	0.09	2	0.01	0	2.43	0	0.42	0.37	12.12	0.00	0.96	1.35	0.62	-
.	.	17	.	.	621	456	394	394	43	42	10	0.47	11	0.59	9	0.54	2	2.61	1	4.66	0.87	8.52	10.59	1.26	2.43	4.63	ok
.	.	18	.	.	804	400	333	333	38	38	5	0.75	5	0.87	4	0.79	1	4.52	1	5.47	2.44	17.61	12.00	1.14	4.37	6.19	ok
.	.	19	.	.	208	208	179	179	35	35	2	0.24	2	0.26	1	0.40	0	2.00	0	3.37	0.56	4.21	7.17	0.62	2.46	3.66	ok
.	.	20	.	.	6450	886	748	747	59	59	36	1.00	38	1.06	34	1.10	2	7.49	2	8.73	1.12	11.82	21.03	6.09	7.58	8.64	ok
.	.	21	.	.	223	223	198	198	48	46	12	0.19	13	0.22	11	0.20	1	1.45	0	2.26	0.44	3.06	5.46	1.21	1.37	2.04	ok
.	.	22	.	.	4118	872	769	766	56	56	39	0.60	46	0.80	33	0.72	8	3.20	5	4.55	0.96	5.27	6.95	1.87	3.32	5.09	ok
.	.	23	.	.	14	14	14	14	11	11	1	0.00	1	0.00	1	0.00	0	0.00	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-
.	.	24	.	.	992	689	604	585	56	56	65	0.17	68	0.18	60	0.19	6	0.93	2	3.65	0.27	2.57	3.84	0.63	0.94	4.24	ok
.	.	25	.	.	851	564	485	485	47	45	24	0.45	25	0.48	22	0.48	2	2.64	1	4.39	0.76	6.31	8.67	1.86	2.95	4.93	ok
.	.	26	.	.	1450	613	522	520	57	56	18	0.62	19	0.71	17	0.65	1	6.42	1	7.58	0.78	12.86	13.36	5.15	4.88	8.59	ok
.	.	27	.	.	281	281	253	253	46	45	15	0.36	16	0.36	15	0.36	1	1.38	0	1.74	0.41	2.21	4.25	1.34	1.48	1.84	ok
.	.	28	.	.	7799	1009	896	895	61	60	82	0.67	88	0.83	77	0.65	7	5.61	5	7.17	0.88	9.22	14.41	4.04	6.07	7.29	ok
.	.	29	.	.	3745	1042	935	929	59	58	100	0.81	105	0.81	95	0.82	6	2.90	3	4.63	0.56	4.13	9.99	5.72	3.64	4.84	ok
.	.	30	.	.	136	136	111	111	34	34	2	0.71	2	0.72	2	0.74	0	2.93	0	4.54	0.81	3.69	9.34	2.25	3.47	4.46	ok
.	.	31	.	.	1088	630	549	536	52	52	31	0.41	33	0.44	29	0.45	3	2.85	1	4.12	0.82	5.09	9.56	1.53	3.20	4.14	ok
.	.	32	.	.	692	496	430	422	51	51	21	0.27	22	0.28	19	0.29	2	1.45	1	3.39	0.44	3.19	6.90	0.80	1.77	3.23	ok
.	.	33	.	.	3441	1036	943	941	56	56	85	0.58	90	0.60	78	0.64	9	1.98	3	8.42	0.88	4.20	11.10	1.22	2.21	9.54	ok
.	.	34	.	.	187	187	151	151	36	34	5	0.22	5	0.28	5	0.22	0	2.39	0	4.83	0.22	4.71	5.97	0.76	1.73	5.06	ok
.	.	35	.	.	451	451	392	392	51	51	15	0.06	15	0.07	14	0.06	1	0.71	0	1.68	0.06	0.97	3.21	0.34	0.99	1.73	ok
.	.	36.37	.	.	3820	917	797	796	56	55	27	0.84	29	0.92	24	0.94	3	4.09	2	6.03	1.18	7.80	9.99	3.88	4.64	7.07	ok
.	.	40.41	.	.	937	649	613	585	62	61	24	0.34	26	0.36	23	0.35	2	1.42	1	2.83	0.33	2.48	5.35	2.10	1.61	2.82	ok
.	.	45	.	.	37400	2458	2139	1990	77	75	286	0.78	303	0.78	273	0.82	14	4.80	15	4.76	0.87	7.99	9.07	4.86	5.40	5.09	ok
.	.	50	.	.	15685	1277	1122	1070	58	57	80	0.81	88	0.97	74	0.97	7	5.99	8	8.18	1.31	10.47	13.43	5.84	6.92	8.49	ok
.	.	51	.	.	21961	1704	1512	1415	76	75	176	0.94	197	0.90	155	1.12	25	3.56	17	5.39	1.39	6.82	9.00	2.89	4.09	5.71	ok
.	.	52	.	.	48676	4768	4298	2116	64	63	240	0.61	312	0.69	179	0.95	69	2.11	64	2.68	2.04	5.54	5.66	1.50	2.22	2.96	ok
.	.	55	.	.	29066	1802	1552	1201	60	59	182	1.02	226	1.11	148	1.37	36	4.51	41	4.21	1.95	8.85	7.81	2.13	4.61	4.37	ok
.	.	60	.	.	8511	2061	1904	996	62	62	89	0.60	100	0.74	80	0.77	10	4.00	10	7.06	0.85	5.40	10.04	2.61	4.38	8.77	ok
.	.	61	.	.	118	118	105	105	37	35	2	0.46	3	0.68	2	0.51	0	2.31	0	8.00	0.91	2.69	8.51	1.78	2.50	8.25	-
.	.	62	.	.	229	229	204	204	39	38	8	0.16	9	0.13	6	0.27	2	0.59	1	0.41	0.55	0.51	1.01	0.38	0.76	0.35	ok
.	.	63	.	.	4281	926	811	746	63	63	53	0.77	61	0.82	47	0.92	8	3.01	6	4.35	1.20	5.00	6.12	2.09	3.51	4.99	ok
.	.	64	.	.	5558	5557	5413	570	45	45	61	0.19	79	0.55	47	0.25	13	0.22	19	2.69	0.25	0.54	4.17	0.42	0.18	2.08	ok
.	.	65	.	.	4534	2277	2175	977	67	67	118	0.40	129	0.41	107	0.43	16	1.19	6	2.51	0.50	2.42	3.59	0.63	1.26	2.99	ok
.	.	66	.	.	2185	1966	1825	840	56	54	48	0.27	52	0.27	42	0.28	8	0.74	2	2.72	0.33	1.47	6.16	0.57	0.80	2.68	ok
.	.	67	.	.	5019	1239	1150	681	59	58	24	1.02	28	1.05	21	1.31	4	4.36	3	9.85	1.46	9.65	17.80	2.69	4.68	8.11	ok
.	.	70	.	.	5362	850	762	669	55	52	23	0.95	31	1.10	18	1.30	5	3.98	8	3.77	1.77	7.90	5.55	2.54	4.48	4.08	ok
.	.	71	.	.	949	597	504	504	39	39	4	0.76	5	0.89	3	0.97	1	4.66	1	4.78	1.23	8.13	6.53	3.04	5.16	6.24	ok
.	.	72	.	.	11872	1144	1029	976	53	53	62	1.11	69	1.16	55	1.24	8	5.29	6	6.06	1.35	7.79	11.04	3.38	5.83	6.94	ok
.	.	73	.	.	662	662	605	575	57	50	16	0.32	18	0.25	14	0.52	3	1.38	1	1.53	1.75	1.79	2.20	2.69	2.12	1.55	ok
.	.	74	.	.	62292	3777	3359	3162	70	70	278	0.57	357	0.66	224	0.77	54	2.39	79	2.55	1.04	4.56	3.46	1.86	2.64	2.78	ok
.	.	75	.	.	8388	4494	4452	1346	62	62	147	0.34	169	0.38	122	0.37	33	1.43	14	2.35	0.54	2.58	6.94	0.93	1.62	2.32	ok
.	.	80	.	.	14997	4209	3984	1577	67	66	173	0.62	249	0.87	98	1.26	77	1.37	74	2.98	1.65	1.97	5.12	2.44	1.82	2.95	ok
.	.	85	.	.	26374	3345	3191	2309	70	69	337	0.38	453	0.54	199	0.74	169	1.08	86	2.34	1.12	3.15	5.26	1.01	1.19	2.53	ok
.	.	90	.	.	2008	969	912	829	59	59	18	0.55	19	0.66	17	0.64	1	3.90	1	7.44	0.69	6.73	10.37	2.75	4.46	11.94	ok
.	.	91	.	.	5776	915	837	832	52	52	33	1.30	48	1.57	21	1.96	13	2.73	14	4.86	2.06	4.45	6.69	3.53	3.01	5.11	ok
.	.	92	.	.	7159	1191	1066	845	63	61	41	0.90	55	1.30	29	1.32	12	2.88	14	5.03	1.71	4.37	6.58	2.66	3.54	6.16	ok
.	.	93	.	.	14291	955	839	805	55	54	34	1.34	44	1.66	26	1.94	9	5.42	8	7.82	4.87	12.83	18.41	2.78	5.72	7.95	ok
.	.	.	201	.	12441	3017	2530	2476	259	252	131	0.67	138	0.70	124	0.69	9	3.13	5	7.11	0.80	7.67	11.28	1.31	2.64	7.40	ok
.	.	.	202	.	18683	3658	3230	3181	277	271	230	0.47	246	0.49	214	0.51	19	2.70	13	4.08	0.62	5.33	7.55	1.52	2.79	4.45	ok
.	.	.	203	.	10298	1652	1455	1404	106	105	157	0.57	168	0.59	146	0.63	14	2.76	8	4.67	0.73	7.28	9.89	1.22	2.54	5.22	ok
.	.	.	204	.	11835	2146	1882	1857	176	175	134	0.87	145	0.86	123	0.95	13	3.21	9	4.54	1.10	6.43	8.43	1.93	3.04	4.85	ok
.	.	.	205	.	13613	2854	2507	2464	200	197	169	0.77	180	0.75	158	0.82	12	2.41	10	4.41	0.92	5.49	8.32	2.87	2.41	4.98	ok
.	.	.	206	.	8283	1317	1161	1145	100	99	100	0.75	108	0.80	93	0.78	8	3.82	6	5.36	0.92	6.82	12.48	2.92	3.74	5.26	ok
.	.	.	207	.	3552	879	771	741	96	96	44	1.14	46	1.12	42	1.18	2	5.10	1	10.93	1.58	12.12	18.56	2.21	5.38	10.33	ok
.	.	.	301	.	57261	10056	9208	5435	323	316	451	0.35	539	0.38	358	0.48	114	1.16	67	2.08	0.68	2.70	3.02	1.00	1.24	2.32	ok
.	.	.	302	.	60897	10743	10066	5435	302	299	449	0.37	577	0.46	330	0.56	132	1.25	114	1.99	0.75	2.83	4.02	1.17	1.45	2.14	ok
.	.	.	303	.	40077	5249	4860	3080	118	117	292	0.59	367	0.68	222	0.85	79	2.00	66	2.93	1.08	3.56	5.01	1.82	2.27	3.18	ok
.	.	.	304	.	57774	8794	8170	4700	218	215	496	0.33	612	0.41	390	0.44	121	1.25	101	2.00	0.62	2.17	3.26	0.98	1.43	2.14	ok
.	.	.	305	.	43017	6236	5800	3530	209	206	260	0.64	331	0.70	198	0.91	67	1.94	66	2.66	1.17	4.00	5.29	1.52	2.10	2.76	ok
.	.	.	306	.	30792	3488	3204	2021	112	111	196	0.67	252	0.84	150	1.00	47	2.71	55	3.75	1.39	5.91	6.84	2.20	2.82	3.87	ok
.	.	.	307	.	16135	2466	2303	1149	102	99	104	0.92	124	0.98	85	1.17	23	3.09	15	5.22	1.96	6.91	8.12	2.35	3.35	5.70	ok
.	.	.	.	211	3327	980	785	782	83	79	38	0.94	40	0.93	37	0.98	3	3.87	1	13.82	1.17	6.34	29.20	1.29	4.31	12.54	ok
.	.	.	.	212	5974	1417	1216	1189	95	94	60	1.02	63	1.10	56	1.05	5	5.03	3	10.98	1.26	12.24	16.35	2.10	3.94	10.77	ok
.	.	.	.	221	2509	840	738	735	84	84	30	0.74	32	0.77	28	0.82	2	4.63	1	5.90	0.90	9.64	12.12	1.63	4.72	6.60	ok
.	.	.	.	222	2030	854	757	754	85	80	30	0.73	32	0.71	29	0.81	3	2.81	1	6.38	1.05	10.81	12.72	1.19	2.58	7.39	ok
.	.	.	.	241	2550	803	700	687	78	77	31	1.61	34	1.59	29	1.69	4	2.99	2	7.68	1.87	4.47	13.08	2.21	3.55	7.02	ok
.	.	.	.	251	5725	1600	1404	1378	100	99	79	0.64	84	0.64	74	0.69	6	2.81	4	4.48	0.94	7.48	9.65	2.10	2.47	4.96	ok
.	.	.	.	311	19648	3547	3155	2237	112	111	185	0.55	214	0.60	151	0.71	45	1.79	19	4.08	1.03	4.80	5.00	1.50	1.83	4.70	ok
.	.	.	.	312	25477	4704	4380	2246	115	112	195	0.47	238	0.54	152	0.70	52	1.63	34	2.82	1.02	3.57	4.23	1.51	1.85	3.16	ok
.	.	.	.	321	8550	2567	2409	1295	94	93	54	0.85	71	1.01	39	1.23	18	2.30	14	3.86	1.70	6.47	7.84	2.49	2.40	4.02	ok
.	.	.	.	322	6316	1579	1435	1156	94	92	40	0.72	52	1.05	29	1.09	13	2.26	10	5.29	1.48	5.41	8.49	2.53	2.38	5.23	ok
.	.	.	.	341	22985	4430	4170	2201	104	103	255	0.37	310	0.49	201	0.51	62	1.53	47	2.71	0.66	2.54	4.12	1.11	1.75	2.59	ok
.	.	.	.	351	17718	3072	2835	1891	104	102	116	0.66	147	0.79	89	0.96	30	2.41	29	3.18	1.26	4.41	5.72	1.69	2.65	3.26	ok

4ème trimestre 2007

(totaux en milliers, CV en %)

											EPT		TOT		TOTVZ		TOTTZ1		TOTTZ2		MVZ	MTZ1	MTZ2	FVZ	FTZ1	FTZ2	
glo	secteur	noga2reg	secgr	secreg	Npop	nBrut	nNet	nNetVar	nstraPopVar	nstraNetVar	sum	cv	sum	cv	sum	cv	sum	cv	sum	cv	cv	cv	cv	cv	cv	cv	test
1	.	.	.	.	384193	62303	55179	36798	2601	2561	3261	0.18	3883	0.20	2675	0.22	670	0.57	538	0.98	0.27	1.15	1.68	0.47	0.63	1.04	ok
.	2	.	.	.	78638	15490	12925	12657	1216	1189	975	0.29	1040	0.29	910	0.31	80	1.16	50	1.95	0.37	2.52	3.74	0.69	1.14	2.09	ok
.	3	.	.	.	305555	46813	42254	24141	1390	1374	2287	0.22	2843	0.25	1765	0.29	590	0.63	488	1.07	0.38	1.28	1.82	0.55	0.69	1.12	ok
.	.	10.14	.	.	392	392	321	321	43	43	5	0.56	5	0.61	4	0.57	0	3.08	0	4.16	0.59	4.58	5.24	2.78	3.47	4.67	ok
.	.	15	.	.	2586	895	756	721	70	69	54	0.53	60	0.65	48	0.59	7	2.49	5	5.70	0.80	4.83	6.57	1.75	2.64	7.25	ok
.	.	16	.	.	15	15	12	12	11	10	3	0.04	3	0.02	3	0.11	0	1.89	0	0.45	0.42	9.11	0.01	0.65	0.99	0.68	-
.	.	17	.	.	621	456	374	374	43	42	10	0.60	11	0.69	9	0.65	2	3.08	1	4.25	1.02	13.06	7.61	1.44	2.51	4.63	ok
.	.	18	.	.	804	400	305	305	38	38	4	0.73	5	0.82	4	0.89	1	4.04	1	4.99	2.43	8.50	9.87	1.34	4.38	5.57	ok
.	.	19	.	.	208	208	156	156	35	33	2	1.86	2	1.66	2	2.16	0	2.39	0	4.62	0.55	6.11	8.47	4.58	2.60	5.10	ok
.	.	20	.	.	6449	885	713	712	60	59	36	1.09	39	1.13	34	1.18	2	7.80	2	8.17	1.21	12.29	17.06	5.75	7.68	8.47	ok
.	.	21	.	.	223	223	184	184	48	45	12	0.18	13	0.20	11	0.18	1	1.57	0	1.86	0.45	2.79	3.19	1.25	1.58	1.95	ok
.	.	22	.	.	4111	867	714	711	56	56	40	0.66	46	0.87	33	0.77	8	2.84	5	4.92	1.02	5.17	8.02	1.85	3.26	5.01	ok
.	.	23	.	.	14	14	11	11	11	9	1	0.01	1	0.07	1	0.02	0	0.20	0	4.56	0.06	0.23	14.19	0.29	0.19	0.14	-
.	.	24	.	.	987	686	581	562	56	56	65	0.25	69	0.26	61	0.25	6	1.11	2	3.96	0.37	2.30	4.43	0.91	1.24	4.65	ok
.	.	25	.	.	847	560	468	468	47	47	25	0.54	26	0.55	23	0.56	2	2.55	1	4.08	0.84	5.46	7.54	1.84	2.71	4.57	ok
.	.	26	.	.	1448	612	495	493	57	55	18	0.60	19	0.69	17	0.64	1	6.86	1	6.97	0.85	12.92	12.47	5.60	5.34	8.00	ok
.	.	27	.	.	281	281	241	241	46	46	16	0.48	16	0.48	15	0.48	1	1.75	0	2.89	0.54	2.51	5.58	1.81	2.08	2.93	ok
.	.	28	.	.	7799	1011	866	864	61	60	84	0.73	90	0.83	78	0.76	7	5.42	5	7.43	1.00	8.76	14.36	3.75	5.77	7.65	ok
.	.	29	.	.	3739	1039	893	887	59	59	101	0.48	106	0.50	96	0.51	7	3.10	3	4.51	0.55	5.25	9.44	2.09	3.62	4.69	ok
.	.	30	.	.	135	135	100	100	34	33	2	1.16	2	1.18	2	1.15	0	3.43	0	4.10	1.09	3.45	9.94	2.81	4.68	4.31	ok
.	.	31	.	.	1084	628	512	499	53	53	32	0.62	34	0.64	30	0.67	3	2.51	1	4.71	0.98	5.29	10.67	1.31	2.66	4.73	ok
.	.	32	.	.	689	493	405	397	51	51	21	0.60	22	0.60	19	0.63	2	2.21	1	4.06	0.78	3.60	8.15	1.50	2.91	4.10	ok
.	.	33	.	.	3439	1034	899	897	56	56	88	0.47	93	0.52	81	0.51	9	2.06	3	8.89	0.80	4.27	11.28	1.17	2.33	10.15	ok
.	.	34	.	.	187	187	145	145	36	33	5	0.34	5	0.40	5	0.33	0	2.61	0	5.30	0.35	4.60	7.72	1.17	2.23	4.80	ok
.	.	35	.	.	450	450	361	361	51	49	15	0.11	15	0.11	14	0.11	1	1.08	0	2.62	0.10	1.46	4.14	0.63	1.57	2.51	ok
.	.	36.37	.	.	3818	916	757	756	59	55	27	0.93	30	0.98	25	1.04	3	3.85	2	5.82	1.16	7.67	10.30	4.29	4.35	6.76	ok
.	.	40.41	.	.	937	649	588	560	62	60	24	0.44	26	0.45	23	0.44	2	1.54	1	3.12	0.47	2.52	6.11	2.58	1.85	3.10	ok
.	.	45	.	.	37375	2454	2068	1920	78	74	287	0.91	303	0.90	273	0.95	14	4.83	15	4.84	0.99	8.02	9.42	5.10	5.37	5.11	ok
.	.	50	.	.	15679	1276	1061	1010	60	59	81	1.05	89	1.12	74	1.23	7	6.05	7	9.02	1.51	10.54	13.24	5.44	6.89	9.65	ok
.	.	51	.	.	21948	1709	1441	1343	78	75	179	1.09	201	1.03	159	1.27	26	3.92	17	5.47	1.49	7.52	9.77	3.18	4.16	5.59	ok
.	.	52	.	.	48670	4776	4193	2008	69	65	247	0.75	320	0.81	185	1.06	70	1.98	64	3.00	2.17	5.76	5.57	1.66	2.12	3.36	ok
.	.	55	.	.	29050	1797	1485	1134	62	60	183	1.43	228	1.53	150	1.72	34	4.79	44	4.97	2.29	9.66	8.49	2.33	4.93	4.92	ok
.	.	60	.	.	8492	2047	1859	963	63	63	93	0.81	103	0.84	84	0.96	10	3.42	10	6.15	0.94	4.72	7.57	3.09	3.90	8.70	ok
.	.	61	.	.	118	118	95	95	37	35	2	1.09	2	1.10	2	1.19	0	2.54	0	9.58	1.29	2.34	11.11	2.13	3.12	9.24	-
.	.	62	.	.	229	229	187	187	39	38	8	0.32	9	0.28	6	0.47	2	0.85	1	0.68	0.83	0.68	1.62	0.54	1.09	0.52	ok
.	.	63	.	.	4281	928	774	703	63	63	55	0.79	62	0.89	48	0.90	8	3.24	6	4.97	1.24	5.17	6.36	2.22	3.60	5.56	ok
.	.	64	.	.	5725	5725	5513	502	45	45	60	0.31	79	0.79	46	0.32	13	0.44	20	3.48	0.32	0.84	4.89	0.47	0.42	2.84	ok
.	.	65	.	.	4513	2256	2154	980	70	69	120	0.33	131	0.35	108	0.34	17	1.18	6	2.52	0.41	2.49	3.09	0.58	1.25	3.03	ok
.	.	66	.	.	2157	1939	1771	800	55	54	48	0.29	53	0.28	42	0.31	8	0.84	2	2.90	0.48	1.91	5.50	0.70	0.81	2.93	ok
.	.	67	.	.	5013	1237	1120	654	59	58	26	1.04	29	1.07	23	1.19	4	3.94	3	5.68	1.46	8.36	9.53	2.36	4.39	6.63	ok
.	.	70	.	.	5356	850	753	660	55	52	24	1.00	32	1.12	18	1.34	5	3.89	8	3.68	1.76	7.91	5.47	2.75	4.33	3.85	ok
.	.	71	.	.	945	594	465	465	39	39	4	0.91	5	0.98	4	1.17	1	4.24	1	4.94	1.39	7.11	6.45	3.13	5.35	5.72	ok
.	.	72	.	.	11865	1143	965	910	53	53	63	1.26	70	1.27	57	1.35	8	5.06	6	5.95	1.53	7.63	10.68	3.58	5.43	7.07	ok
.	.	73	.	.	656	656	558	529	56	53	15	0.97	17	0.96	13	1.00	3	1.68	1	2.28	0.51	2.14	3.22	3.49	2.09	2.19	ok
.	.	74	.	.	62211	3768	3203	3012	72	72	285	0.65	363	0.71	230	0.83	55	2.37	78	2.60	1.10	4.45	3.75	1.87	2.64	2.80	ok
.	.	75	.	.	8132	4239	4125	1274	64	63	150	0.59	172	0.63	123	0.67	35	1.70	14	2.81	0.90	2.89	7.55	1.09	1.81	2.73	ok
.	.	80	.	.	14966	4183	3961	1539	67	66	176	0.60	253	0.98	99	1.22	79	1.49	75	3.37	1.71	2.06	5.30	2.31	1.79	3.31	ok
.	.	85	.	.	26330	3317	3091	2235	70	69	342	0.40	460	0.54	202	0.74	171	1.10	87	2.29	1.19	3.35	5.62	1.02	1.21	2.43	ok
.	.	90	.	.	2005	967	884	803	59	59	18	0.65	20	0.71	17	0.77	1	4.12	1	7.65	0.79	6.84	10.76	4.07	4.11	12.03	ok
.	.	91	.	.	5776	915	811	804	53	52	33	1.15	48	1.60	20	1.57	14	2.65	14	4.85	2.02	4.12	7.12	2.75	2.98	5.04	ok
.	.	92	.	.	7155	1190	987	766	63	62	41	1.40	55	1.77	29	1.61	12	3.52	14	5.92	1.82	5.18	7.61	2.94	3.76	6.55	ok
.	.	93	.	.	14283	954	798	765	55	55	35	1.40	44	1.74	27	2.04	9	5.46	9	7.69	4.90	12.37	17.54	2.93	5.88	8.12	ok
.	.	.	201	.	12427	3007	2396	2342	260	251	132	0.79	138	0.79	124	0.83	9	3.23	5	6.19	0.94	7.70	9.59	1.30	2.59	6.71	ok
.	.	.	202	.	18665	3646	3078	3031	278	270	233	0.59	249	0.59	217	0.65	20	2.51	12	4.26	0.75	5.05	8.15	1.45	2.68	4.70	ok
.	.	.	203	.	10287	1647	1382	1331	107	104	159	0.58	170	0.58	148	0.64	15	2.76	7	4.60	0.79	7.29	9.61	1.43	2.44	4.98	ok
.	.	.	204	.	11828	2147	1833	1808	176	173	136	0.89	147	0.88	125	0.93	13	2.77	9	4.26	1.06	5.98	7.93	1.98	2.80	4.61	ok
.	.	.	205	.	13605	2850	2420	2377	200	199	169	0.76	181	0.74	158	0.81	13	2.43	9	4.53	1.02	5.19	8.77	1.73	2.59	5.04	ok
.	.	.	206	.	8277	1316	1098	1080	100	98	101	0.89	108	0.93	94	0.93	9	4.12	6	5.53	1.05	6.90	12.02	3.29	4.00	5.37	ok
.	.	.	207	.	3549	877	718	688	96	94	45	1.37	47	1.35	44	1.42	2	5.39	1	13.98	1.84	13.47	18.63	2.41	5.69	13.70	ok
.	.	.	301	.	57181	10039	8843	5041	323	317	462	0.45	553	0.47	367	0.57	117	1.24	69	2.02	0.79	2.78	2.90	1.02	1.31	2.22	ok
.	.	.	302	.	60760	10633	9717	5190	305	302	451	0.47	577	0.54	332	0.63	135	1.30	110	2.07	0.81	3.00	4.22	1.21	1.47	2.18	ok
.	.	.	303	.	40022	5216	4698	2944	117	117	296	0.68	372	0.80	225	0.95	80	1.91	68	3.43	1.19	3.66	5.41	1.83	2.13	3.72	ok
.	.	.	304	.	57765	8813	7993	4507	221	219	504	0.39	620	0.48	397	0.49	121	1.35	102	2.25	0.68	2.21	3.65	1.07	1.53	2.33	ok
.	.	.	305	.	42957	6193	5661	3422	209	209	268	0.70	340	0.76	205	0.97	67	1.84	68	3.00	1.24	3.79	5.68	1.64	2.08	2.95	ok
.	.	.	306	.	30759	3470	3112	1943	114	112	202	0.88	258	1.02	154	1.18	48	2.65	56	3.85	1.53	5.85	6.47	2.38	2.72	3.97	ok
.	.	.	307	.	16111	2449	2230	1094	102	99	103	1.07	122	1.03	85	1.35	22	2.95	15	4.73	2.06	5.10	8.16	2.71	3.19	5.08	ok
.	.	.	.	211	3321	977	729	726	83	81	39	0.98	41	0.98	37	1.01	3	3.95	1	10.72	1.19	7.79	18.03	1.43	4.17	13.08	ok
.	.	.	.	212	5965	1411	1162	1135	95	94	60	0.97	63	1.01	56	1.02	5	5.05	2	8.23	1.17	11.78	13.71	1.83	3.97	8.35	ok
.	.	.	.	221	2507	838	704	701	84	84	31	0.93	33	0.93	29	0.99	2	4.87	2	8.20	1.13	11.89	17.54	2.22	4.18	9.86	ok
.	.	.	.	222	2026	851	695	692	85	79	31	0.77	33	0.76	29	0.82	3	2.87	1	7.27	1.03	7.35	18.71	1.64	2.77	8.06	ok
.	.	.	.	241	2548	804	684	671	78	76	32	1.74	34	1.73	29	1.81	3	3.35	2	7.23	2.07	4.42	14.14	2.19	3.68	6.55	ok
.	.	.	.	251	5721	1596	1346	1320	100	100	81	0.69	86	0.67	76	0.74	6	2.92	4	4.46	1.09	6.85	8.77	2.39	3.38	4.94	ok
.	.	.	.	311	19596	3528	2934	1991	111	109	188	0.68	218	0.74	152	0.83	46	2.01	19	3.80	1.22	4.28	4.83	1.52	2.06	4.29	ok
.	.	.	.	312	25447	4697	4261	2131	115	113	198	0.57	241	0.62	154	0.80	53	1.74	35	2.81	1.09	3.61	4.05	1.56	1.98	3.08	ok
.	.	.	.	321	8540	2558	2328	1218	96	95	55	0.94	72	1.04	40	1.30	18	2.35	14	3.85	1.81	8.09	8.43	2.66	2.33	3.98	ok
.	.	.	.	322	6308	1574	1379	1103	94	92	41	0.78	53	1.10	29	1.08	14	2.19	10	5.30	1.61	5.29	9.29	2.38	2.33	5.32	ok
.	.	.	.	341	22968	4426	4079	2118	107	105	259	0.47	314	0.58	206	0.60	62	1.58	47	3.09	0.75	2.60	4.74	1.17	1.84	2.85	ok
.	.	.	.	351	17690	3054	2781	1849	104	104	118	0.73	149	0.80	90	1.01	30	2.28	29	3.01	1.29	4.38	5.36	1.81	2.44	3.22	ok

Références

Deville, J.-C., and Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*(418), 376-382.

Hulliger, B. (1999). Simple and robust estimators for sampling. In *ASA : Proceedings of the section on Survey Research.*

Renaud, A. (2008). *Statistique de l'emploi. Révision 2007 : cadre de sondage et échantillonnage.* (Office fédéral de la statistique. Rapport de méthodes 338-0052).

Särndal, C. E. and Lundström, S. (2005). *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., Panchard, C. et Potterat, J. (2008). Statistique de l'emploi. Révision 2007 : méthodes d'estimation. Numéro de commande : 338-0055

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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. Les estimateurs des totaux ont également été revus. Ils utilisent une procédure de détection de valeurs aberrantes et un calage direct sur des valeurs de référence connues au niveau du cadre sans passer par des modèles de non-réponse. Une estimation de la variance est également proposée afin de donner des informations sur la précision des estimateurs. Le rapport présente finalement une évaluation des nouvelles méthodes sur les trois premiers trimestres d'application, c'est-à-dire les 2ème, 3ème et 4ème trimestres 2007.

N° de commande: 338-0055
Commandes: 032 713 60 60 Fax: 032 713 60 61 E-Mail: order@bfs.admin.ch
Prix: gratuit
ISBN 978-3-303-00413-5